

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC DÂN LẬP HẢI PHÒNG
-----o0o-----

TÌM HIỂU PHƯƠNG PHÁP TRA CỨU ẢNH Y TẾ

ĐỒ ÁN TỐT NGHIỆP ĐẠI HỌC HỆ CHÍNH QUY

Ngành: Công nghệ Thông tin

HẢI PHÒNG - 2012

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC DÂN LẬP HẢI PHÒNG
-----o0o-----

TÌM HIỂU PHƯƠNG PHÁP TRA CỨU ẢNH Y TẾ

ĐỒ ÁN TỐT NGHIỆP ĐẠI HỌC HỆ CHÍNH QUY

Ngành: Công nghệ Thông tin

Sinh viên thực hiện: VŨ THỊ HỒNG LIÊN

Giáo viên hướng dẫn: PGS TS NGÔ QUỐC TẠO

Mã số sinh viên: 1013101002

HẢI PHÒNG - 2012

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: Vũ Thị Hồng Liên

Mã số: 1013101002

Lớp: CTL401

Ngành: Công nghệ Thông tin

Tên đề tài: Tìm hiểu phương pháp tra cứu ảnh y tế

NHIỆM VỤ ĐỀ TÀI

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

a. Nội dung:

- + Trình bày tổng quan về tra cứu ảnh.
- + Một số phương pháp tìm kiếm ảnh theo nội dung.
- + Tra cứu ảnh y tế và giới thiệu một ứng dụng của phương pháp tra cứu ảnh theo nội dung trong y tế; những hạn chế và khả năng mở rộng của chương trình ứng dụng đó.

b. Các yêu cầu cần giải quyết

- + Giới thiệu chi tiết phương pháp tra cứu ảnh theo nội dung.
- + Sơ lược về cách đánh giá hiệu năng của một hệ thống tra cứu ảnh.
- + Xây dựng được chương trình thử nghiệm theo hai đặc điểm là : Hình dạng và màu sắc.

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

.....

.....

.....

.....

.....

.....

3. Địa điểm thực tập

Viện khoa học và công nghệ Việt Nam – Số 18 đ- ờng Hoàng Quốc Việt, quận Cầu Giấy, Hà Nội.

CÁN BỘ HƯỚNG DẪN ĐỀ TÀI TỐT NGHIỆP

Người hướng dẫn thứ nhất:

Họ và tên: Ngô Quốc Tạo

Học hàm, học vị: Phó giáo sư – Tiến sĩ

Cơ quan công tác: Viện khoa học và công nghệ Việt Nam

Nội dung hướng dẫn:

.....

.....

.....

.....

Người hướng dẫn thứ hai:

Họ và tên:

Học hàm, học vị:

Cơ quan công tác:

Nội dung hướng dẫn:

.....

.....

.....

.....

Đề tài tốt nghiệp được giao ngày tháng năm 2012

Yêu cầu phải hoàn thành trước ngày tháng năm 2012

Đã nhận nhiệm vụ: Đ.T.T.N
Sinh viên

Đã nhận nhiệm vụ: Đ.T.T.N
Cán bộ hướng dẫn Đ.T.T.N

Hải Phòng, ngàytháng.....năm 2012
Hiệu trưởng

GS.TS.NGUYỄN Trần Hữu Nghị

PHÂN NHẬN XÉT TÓM TẮT CỦA CÁN BỘ HƯỚNG DẪN

1. Tinh thần thái độ của sinh viên trong quá trình làm đề tài tốt nghiệp:

.....
.....
.....
.....
.....
.....

2. Đánh giá chất lượng của đề tài tốt nghiệp (so với nội dung yêu cầu đã đề ra trong nhiệm vụ đề tài tốt nghiệp)

.....
.....
.....
.....
.....
.....
.....

3. Cho điểm của cán bộ hướng dẫn:

(Điểm ghi bằng số và chữ)

.....
.....

Ngày.....tháng.....năm 20...
Cán bộ hướng dẫn chính
(Ký, ghi rõ họ tên)

LỜI CẢM ƠN

Em xin chân thành cảm ơn PGS - TS. Ngô Quốc Tạo, người đã trực tiếp hướng dẫn và tận tình giúp đỡ em trong quá trình thực hiện đề án này, những kiến thức, và phương pháp nghiên cứu em học từ Thầy thực sự rất quý giá, không những giúp ích cho em ở hiện tại mà còn là tiền đề để em có thể tiếp thu kiến thức mới một cách tốt hơn, một lần nữa em xin cảm ơn Thầy rất nhiều. Em xin cảm ơn Thạc sỹ Ngô Trường Giang vì thông qua môn học Đồ họa máy tính và Xử lý ảnh đã giúp em có niềm đam mê với lĩnh vực đồ họa máy tính, những kiến thức từ hai môn học đã góp phần giúp em hoàn thành đề án này.

Em xin gửi lời cảm ơn tới khoa CNTT trường ĐHDL Hải Phòng, vì trong thời gian học tập ở trường em đã học hỏi được những kiến thức, và tư duy, giúp em phát triển ý tưởng trong đề án này.

Cuối cùng em xin gửi lời cảm ơn tới Gia đình và bạn bè đã bên cạnh giúp đỡ đồng thời ủng hộ em trong quá trình thực hiện đề án này.

Hải Phòng, tháng 11 năm 2012

Sinh viên thực hiện

Vũ Thị Hồng Liên

MỤC LỤC

LỜI CẢM ƠN	1
LỜI MỞ ĐẦU	4
Chương 1 : TỔNG QUAN VỀ TRA CỨU ẢNH	6
1.1. Giới thiệu về tra cứu ảnh	6
1.2. Mô hình hệ thống tra cứu ảnh theo nội dung	6
1.3. Ứng dụng của tra cứu ảnh theo nội dung.....	7
Chương 2 :MỘT SỐ PHƯƠNG PHÁP TÌM KIẾM ẢNH THEO NỘI DUNG	9
2.1. Phương pháp trích chọn theo màu sắc tổng thể và cục bộ	10
2.1.1. Không gian màu.....	10
2.1.1.1 Không gian màu RGB	11
2.1.1.2. Không gian màu CMY.....	11
2.1.1.3 Không gian màu HSx.....	12
2.1.1.4. Không gian màu YUV và YIQ.....	13
2.1.1.5. Không gian màu CIE XYZ và LUV	13
2.1.2. Lượng tử hoá màu	14
2.1.3. Các moment màu	14
2.1.4. Biểu đồ màu (Color Histogram)	15
2.1.4.1. Biểu đồ màu toàn cục (Global Color Histogram)	16
2.1.4.2. Biểu đồ màu cục bộ (Local Color Histogram).....	18
2.2. Phương pháp trích chọn đặc trưng theo kết cấu	19
2.2.1. Các đặc trưng Tamura.....	20
2.2.2. Các đặc trưng Wold.....	22
2.2.3. Mô hình tự thoái lui đồng thời (mô hình SAR).....	23
2.2.4. Ma trận đồng khả năng (Co-occurrence matrix)	24
2.2.5. Lọc Gabor	26
2.2.6. Biến đổi dạng sóng (wavelet transform)	27
2.3. Phương pháp trích chọn đặc trưng theo hình dạng	28
2.3.1. Biên và các phương pháp phát hiện biên	29
2.3.2. Xử lý ảnh trong miền tần số và biến đổi Fourier.....	32

2.4. Độ đo khoảng cách và độ đo tương tự	37
2.4.1. Độ đo khoảng cách	37
2.4.1.1. Độ đo khoảng cách Minkowski	37
2.4.1.2 Độ đo khoảng cách Quadratic	38
2.4.1.3. Độ đo khoảng cách Non-histogram	39
2.4.1.4. Khoảng cách Mahalanobis	40
2.4.2. Độ đo tương tự	40
2.4.2.1. Độ phân kỳ Kullback-Leibler	40
2.4.2.2. Độ phân kỳ Jeffrey	40
2.4.2.3. Độ đo tương tự hình dạng toàn cục	41
2.5. Đánh giá hiệu năng của hệ thống tra cứu ảnh	44
Chương 3 : TRA CỨU ẢNH Y TẾ	47
3.1. Tìm hiểu đặc trưng của ảnh y tế	47
3.2 Tác dụng của tra cứu ảnh y tế.....	49
3.3. Chương trình mô phỏng	50
3.4. Khả năng mở rộng của chương trình	51
3.4.1. Những hạn chế của chương trình.....	51
3.4.2. Khả năng mở rộng	52
KẾT LUẬN	53
TÀI LIỆU THAM KHẢO	55

LỜI MỞ ĐẦU

Cùng với sự bùng nổ thông tin trên web và sự phát triển của công nghệ kỹ thuật số, lượng ảnh lưu trữ trên Web cũng tăng một cách nhanh chóng. Vì vậy, việc xây dựng các hệ thống tìm kiếm là rất cần thiết. Các công cụ tìm kiếm ảnh thường dựa vào hai đặc trưng chính là văn bản đi kèm ảnh hoặc nội dung ảnh. Một số công cụ tìm kiếm ảnh theo văn bản đi kèm như Google Image Search, Yahoo!, MSN,... Một số công cụ tìm kiếm ảnh dựa vào nội dung ảnh như Google Image Swirl, Bing, Tiltomo, Tineye,...

Tìm kiếm ảnh theo nội dung đã nhận được nhiều sự quan tâm của các nhà khoa học. 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.

Trong thực tế, bài toán tra cứu ảnh có nhiều ứng dụng quan trọng. Đặc biệt trong lĩnh vực y học hiện nay việc ứng dụng tra cứu ảnh y tế cũng hết sức quan trọng. Việc tìm ra một bức ảnh giống hoặc tương tự với ảnh mẫu hay không. Trong trường hợp này nếu sử dụng bằng mắt thường để duyệt thì sẽ tốn rất nhiều thời gian và độ chính xác không cao, ngược lại nếu có các phần mềm cho phép tìm kiếm trong cơ sở dữ liệu ảnh có sẵn những hình ảnh tương tự với hình ảnh mẫu thì việc đánh giá và phát hiện bệnh lý sẽ dễ dàng hơn rất nhiều.

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 hay một đặc điểm nào đó của ảnh, việc tra cứu ảnh được thực hiện dựa trên những lời chú thích này. Phương pháp này khá đơn giản. Tuy nhiên, việc tìm kiếm chỉ dựa vào văn bản đi kèm còn có nhiều nhập nhằng giữa nội dung hiển thị ảnh và nội dung văn bản đi kèm ảnh trong quá trình tìm kiếm. Ví dụ, với truy vấn “Apple”, máy tìm kiếm khó phân biệt được người dùng muốn tìm hình ảnh quả táo hay logo của hãng Apple. Bên cạnh đó phương pháp tra cứu ảnh dựa theo văn bản không thể áp dụng để tra cứu các cơ sở dữ liệu ảnh có số lượng ảnh lớn.

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 tưởng 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. Một số hệ thống tra cứu ảnh nổi tiếng như QBIC (IBM), Virage (Virage Inc.), Photobook (MIT), VisualSEEK (Columbia University)... đã áp dụng khá thành công phương pháp tra cứu này.

Nội dung của đề tài này là 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 ứng dụng tra cứu ảnh y tế. 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 hai đặc điểm là hình dạng và màu sắc của ảnh.

Báo cáo được chia làm ba chương:

Chương 1: Trình bày tổng quan về tra cứu ảnh

Chương 2: Một số phương pháp tìm kiếm ảnh theo nội dung

Chương 3: Tra cứu ảnh y tế và giới thiệu một ứng dụng của phương pháp tra cứu ảnh theo nội dung trong y tế; những hạn chế và khả năng mở rộng của chương trình ứng dụng đó.

Chương 1 : TỔNG QUAN VỀ TRA CỨU ẢNH

1.1. Giới thiệu về tra cứu ảnh

Ngày nay, lĩnh vực tra cứu ảnh nhận được sự quan tâm ngày càng lớn. Lý do một phần là sự phát triển của công nghệ chế tạo thiết bị thu nhận và lưu trữ ảnh cũng như sự phát triển mạnh mẽ của mạng Internet. Người ta sử dụng trong nhiều lĩnh vực khác nhau có cơ hội để truy cập và sử dụng các kho lưu trữ ảnh thuộc đủ loại chủ đề và với nhiều kiểu định dạng ảnh khác nhau. Tuy nhiên người ta cũng nhận thấy rằng việc tìm được một bức ảnh mong muốn trong bộ sưu tập ảnh đa dạng có kích thước lớn là rất khó khăn.

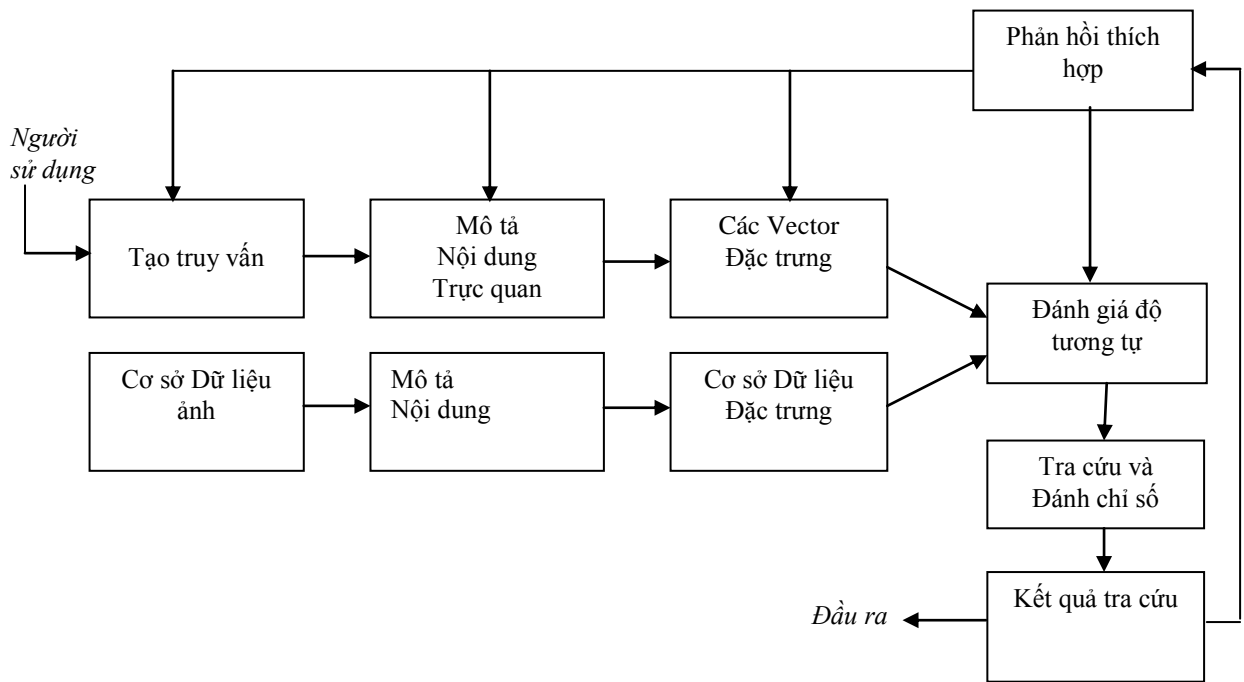
Tra cứu ảnh là một quá trình tìm kiếm trong một cơ sở dữ liệu ảnh - những ảnh thoả mãn một yêu cầu nào đó. Ví dụ một người muốn tìm tất cả các ảnh tương tự với một bức ảnh mẫu nào đó trong một cơ sở dữ liệu ảnh.

Vấn đề tra cứu ảnh đã được nhìn nhận rộng rãi và việc tìm kiếm các giải pháp cho vấn đề này trở thành một lĩnh vực rất sôi động, thu hút sự quan tâm của nhiều nhà nghiên cứu.

Những kỹ thuật tra cứu ảnh số đã được nghiên cứu từ cuối những năm 70 của thế kỷ 20. Năm 1979 một cuộc hội thảo chuyên đề về "Các kỹ thuật tổ chức cơ sở dữ liệu cho các ứng dụng đồ hoạ" được tổ chức ở thành phố Florence, Italia. Từ đó đến nay, khả năng ứng dụng cao của các kỹ thuật quản lý cơ sở dữ liệu ảnh đã thu hút sự quan tâm của nhiều nhà nghiên cứu.

1.2. Mô hình hệ thống tra cứu ảnh theo nội dung

Phương pháp tra cứu ảnh theo nội dung (Content-Based Image Retrieval) sử dụng các nội dung trực quan của ảnh như màu sắc, hình dạng, kết cấu (texture) và phân bố không gian để thể hiện và đánh chỉ số các ảnh. Trong một hệ thống tra cứu ảnh theo nội dung điển hình (hình vẽ 1.1) các nội dung trực quan của ảnh được trích chọn và mô tả bằng những vectơ đặc trưng nhiều chiều. Tập hợp các vectơ đặc trưng của các ảnh trong một cơ sở dữ liệu ảnh tạo thành cơ sở dữ liệu đặc trưng. Quá trình tra cứu ảnh được tiến hành như sau: Người sử dụng cung cấp cho hệ thống tra cứu một ảnh mẫu cụ thể. Sau đó hệ thống sẽ chuyển những mẫu này thành các véc tơ đặc trưng và tính toán sự giống nhau (hay độ tương tự) giữa véc tơ đặc trưng của ảnh mẫu và véc tơ đặc trưng của các ảnh trong cơ sở dữ liệu. Sau cùng việc tra cứu được tiến hành với sự trợ giúp của các sơ đồ đánh chỉ số. Sử dụng sơ đồ đánh chỉ số là cách hiệu quả để tìm kiếm trong các cơ sở dữ liệu ảnh. Một số hệ thống tra cứu ảnh mới phát triển gần đây còn tích hợp cả chức năng xử lý phản hồi của người sử dụng để cải tiến các qui trình tra cứu và ra những kết quả tra cứu tốt hơn.



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

Trong mô hình này, người sử dụng sẽ tạo truy vấn bằng cách chọn một ảnh mẫu trong một cơ sở dữ liệu ảnh cho trước hoặc phác thảo một hình vẽ mô tả đối tượng ảnh cần tìm bằng cách sử dụng một giao diện đồ họa của hệ thống.

Ảnh mẫu đó được đưa qua khối mô tả nội dung trực quan, trong đó người ta sử dụng một phương pháp mô tả nội dung trực quan nào đó để trích chọn một đặc điểm nội dung trực quan để xây dựng thành một véc tơ đặc trưng.

Véc tơ đặc trưng của ảnh mẫu sẽ được so sánh với véc tơ đặc trưng tương ứng của các ảnh trong cơ sở dữ liệu ảnh. Kết quả của phép so sánh là một chỉ số đánh giá độ tương tự giữa ảnh mẫu và ảnh lấy ra để so sánh.

Dựa vào chỉ số độ tương tự tính toán được ở trên, hệ thống sẽ sắp xếp các ảnh tìm được trong cơ sở dữ liệu ảnh theo một sơ đồ đánh chỉ số nào đó. Danh sách các ảnh tìm được (đã được sắp xếp) được đưa ra đầu ra của hệ thống.

1.3. Ứng dụng của tra cứu ảnh theo nội dung

Ứng dụng của tra cứu ảnh có rất nhiều trong đời sống xã hội, phục vụ cho nhiều mục đích khác nhau, nhằm xác nhận, tra cứu thông tin. Giảm bớt công việc của con người nhằm tăng hiệu suất làm việc: Album ảnh số của người dùng, ảnh y khoa, bảo tàng ảnh, tìm kiếm nhãn hiệu, mô tả nội dung MPEG-7, ảnh tội phạm, hệ thống tự động nhận biết điều khiển giao thông , ...

Sau đây là một vài hệ thống lớn đại diện cho các lĩnh vực đặc trưng:

+ Hệ thống truy vấn ảnh theo nội dung (QBIC-query by image content) được nghiên cứu và phát triển bởi nhóm nghiên cứu Visual Media Management thuộc tập đoàn IBM, đây là một hệ thống tra cứu ảnh thương mại được phát triển từ rất sớm. Hiện nay, hệ thống này hỗ trợ một vài độ đo tương tự cho ảnh như: trung bình màu sắc, lược đồ màu sắc, và kết cấu. Công nghệ sử dụng trong hệ thống bao gồm 2 phần chính là: đánh chỉ số và tìm kiếm. Hơn nữa, hệ thống này còn cung cấp vài cách tiếp cận truy vấn theo đơn đặc trưng, đa đặc trưng và đa giai đoạn.

+ Hệ thống VisualSEEK tại trường đại học Columbia. Hệ thống cho phép người dùng nhập vào truy vấn, sử dụng các đặc trưng mức thấp của hình ảnh như: màu sắc, bố cục không gian, và kết cấu. Các đặc trưng đó được mô tả theo tập các màu sắc và biến đổi Wavelet dựa trên đặc trưng kết cấu.

+ Hệ thống NeTra sử dụng các đặc trưng của ảnh: Màu sắc, hình dạng, kết cấu, không gian.

+ Ngoài ra còn một vài hệ thống khác như: Virage system, Stanford SIMPLICity system, NEC PicHunter system, ...

1.4. Kết luận chương

Tra cứu ảnh theo nội dung (CBIR) là một lĩnh vực khoa học được phát triển dựa trên cơ sở lý thuyết và ứng dụng của xử lý ảnh. Hệ thống cho phép người dùng tra cứu các ảnh tương tự trong một cơ sở dữ liệu hình ảnh. Các hình ảnh này có thể được thu thập thông qua các thiết bị chụp hình, cảm biến, và thiết bị quét hình ảnh, cũng có thể được chia sẻ thông qua hệ thống mạng máy tính toàn cầu.

Tra cứu ảnh theo nội dung là việc tính độ tương tự giữa hai bức ảnh được biểu diễn bởi một trong số các đặc trưng của ảnh như: Màu sắc, hình dạng, kết cấu... Kết quả là tập các bức ảnh tương tự với ảnh truy vấn được sắp xếp theo thứ tự giảm dần độ tương tự.

Chương 2 :MỘT SỐ PHƯƠNG PHÁP TÌM KIẾM ẢNH THEO NỘI DUNG

Nếu nhìn một cách tổng quát thì nội dung của một bức ảnh có thể bao gồm cả nội dung trực quan và nội dung ngữ nghĩa.

Nội dung trực quan của ảnh lại được phân làm hai loại là nội dung tổng quan và nội dung đặc tả. Nội dung tổng quan bao gồm màu sắc, kết cấu, hình dạng và các quan hệ không gian giữa các đối tượng ảnh hoặc giữa các vùng ảnh. Nội dung đặc tả thì tùy vào từng ứng dụng cụ thể, ví dụ với các ứng dụng tra cứu ảnh người thì mặt người hoặc con mắt là các nội dung đặc tả; với các ứng dụng tra cứu ảnh ô tô thì bánh xe là một nội dung đặc tả.

Nội dung ngữ nghĩa có thể phát hiện thông qua các chú thích hoặc sử dụng các phương pháp suy diễn từ nội dung trực quan.

Trong khuôn khổ của đề án này tôi chỉ tập trung vào việc mô tả nội dung trực quan tổng quan của ảnh.

Một phương pháp mô tả nội dung trực quan được thiết kế tốt phải có tính bất biến đối với các biến đổi bất thường sinh ra trong quá trình xử lý ảnh (ví dụ như những biến đổi bất thường của độ sáng của cảnh vật). Tuy nhiên cũng cần phải chú ý tới sự cân bằng giữa tính bất biến và khả năng đáp ứng những thay đổi tùy ý của các đặc trưng trực quan của ảnh, bởi vì một hệ thống có tính bất biến lớn thì thường là không có tính nhạy cảm, mất khả năng phản ánh những thay đổi nhỏ nhưng rất quan trọng.

Một phương pháp mô tả nội dung trực quan có thể là phương pháp toàn cục hoặc phương pháp cục bộ. Phương pháp mô tả nội dung toàn cục sử dụng các đặc trưng trực quan của toàn bộ bức ảnh còn phương pháp mô tả nội dung cục bộ lại sử dụng những đặc trưng trực quan của các vùng ảnh hoặc các đối tượng ảnh để mô tả nội dung của ảnh.

Để mô tả được nội dung cục bộ trước hết người ta phải chia ảnh thành các phần riêng biệt. Cách đơn giản nhất để phân chia ảnh là sử dụng một bộ phân hoạch chia ảnh thành các ô có kích thước và hình dạng giống nhau. Cách phân chia đơn giản như vậy không tạo ra được những vùng ảnh có ý nghĩa thực sự nhưng nó là cách đơn giản để biểu diễn nội dung toàn cục của ảnh với độ chính xác cao hơn. Một phương pháp phân chia tốt hơn là phân chia ảnh thành các vùng đồng nhất dựa

vào các tiêu chí sử dụng các thuật toán phân vùng ảnh đã được nghiên cứu và áp dụng trong ngành thị giác máy tính.

Một cách phức tạp hơn để phân chia ảnh là thực hiện phân chia theo các đối tượng ảnh để tách ra các đối tượng ảnh có nghĩa thực sự (như quả bóng, cái ô tô hay con ngựa). Phần tiếp theo giới thiệu một số kỹ thuật đang được sử dụng rộng rãi để trích chọn các đặc điểm màu sắc, kết cấu, hình dạng và các quan hệ không gian của các đối tượng ảnh.

2.1. Phương pháp trích chọn theo màu sắc tổng thể và cục bộ

Tra cứu ảnh dựa trên màu sắc hầu hết là biến đổi dựa trên ý tưởng giống nhau của các biểu đồ màu. Mỗi ảnh khi đưa vào tập hợp ảnh đều được phân tích, tính toán một biểu đồ màu. 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. Quá trình tra cứu sẽ đối sánh biểu đồ màu này với biểu đồ màu trong cơ sở dữ liệu để tìm ra kết quả tương tự nhất. Kỹ thuật đối sánh được sử dụng phổ biến nhất là biểu đồ màu giao nhau được phát triển đầu tiên bởi Swain. 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. Kết quả của các hệ thống này đã tạo những ấn tượng khá sâu sắc.

2.1.1. Không gian màu

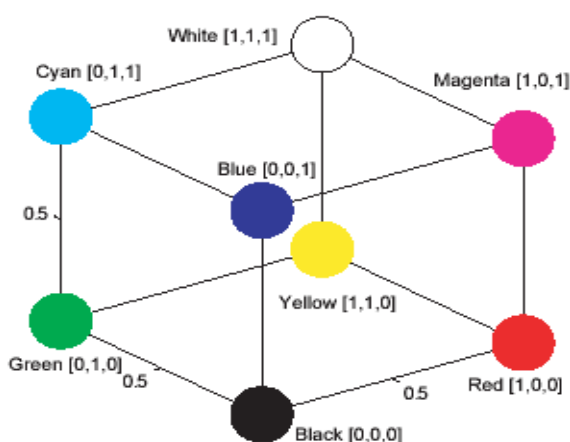
Mỗi điểm ảnh trên một bức ảnh có thể được biểu diễn bằng một điểm trong một không gian màu 3 chiều. Những không gian màu được sử dụng nhiều nhất trong các hệ thống tra cứu ảnh là RGB, CIE L^*a^*b , CIE L^*u^*v , HSV, YUV, YIQ và không gian màu đối lập.

Người ta cũng chưa chỉ ra được rằng không gian màu nào được sử dụng tốt hơn cho việc tra cứu ảnh. Tuy nhiên tính chất quan trọng nhất để một không gian màu phù hợp để sử dụng trong một hệ thống tra cứu ảnh là tính đồng nhất. Một không gian màu được gọi là có tính đồng nhất nếu các cặp 2 màu tương tự nhau trong không gian màu thì cũng được con người cảm nhận như nhau. Nói một cách khác, khoảng cách đo được giữa hai màu bất kỳ phải có liên quan trực tiếp với độ tương tự sinh học giữa hai màu đó.

2.1.1.1 Không gian màu RGB

RGB là không gian màu được sử dụng phổ biến nhất để hiển thị ảnh. Không gian RGB bao gồm 3 thành phần màu là Đỏ (Red), Xanh lá cây (Green) và Xanh lam (Blue). Các thành phần này gọi là màu cộng bởi vì các màu sắc trong không gian RGB đều có thể thu được bằng cách cộng 3 thành phần màu này lại với nhau.

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.

2.1.1.2. Không gian màu CMY

Hệ thống màu CMY theo mô hình in trên giấy trắng và theo khuôn mẫu trừ từ màu trắng thay vì thêm vào từ màu đen như hệ thống màu RGB. 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. Cyan hấp thụ sự chiếu sáng của màu đỏ, Magenta hấp thụ màu xanh lục, Yellow hấp thụ màu xanh dương. Do đó, tạo ra sự phản ánh tương ứng như khi in ảnh được chiếu sáng với ánh sáng trắng. Hệ thống dưới dạng âm tính vì mã hóa theo dạng hấp thụ màu. Có một số mã hóa như sau: trắng (0,0,0) vì không có ánh sáng trắng được hấp thụ, đen (255,255,255) vì tất cả các thành phần của màu trắng đều được hấp thụ.

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

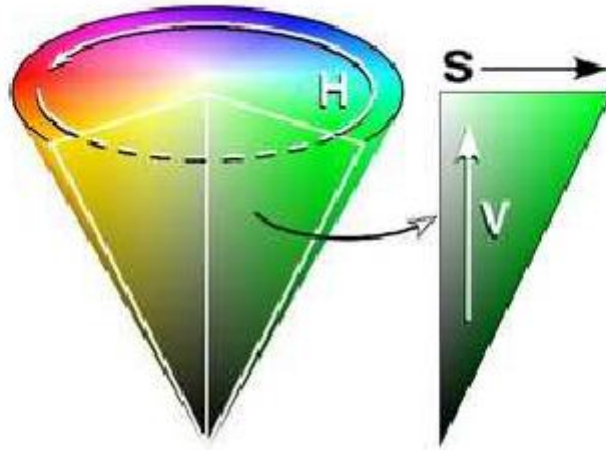
khuyết điểm của nó cũng tương tự như không gian màu RGB, tức là cách mã hóa khác với cách mà con người cảm nhận về màu sắc. Không thích hợp cho bài toán tìm kiếm ảnh dựa vào nội dung.

2.1.1.3 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 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.

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 2.2. Vùng giá trị này từ 0^0 đến 360^0 . Theo uỷ ban quốc tế về màu sắc CIE (Commission International d'Eclairage) 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 mức với H nhận 18 mức, S và V nhận 3 mức.



Hình 2.2. Mô tả không gian màu HSV.

2.1.1.4. 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.

2.1.1.5. Không gian màu CIE XYZ 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.

Các không gian màu RGB và CIE có thể chuyển đổi với nhau, tức là chúng ta có thể sử dụng các công thức để chuyển đổi một giá trị màu từ không gian màu này sang không gian màu khác.

2.1.2. Lượng tử hoá 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 nhỏ 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:

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

$$R' = \frac{n * R}{2^8} \quad G' = \frac{n * G}{2^8} \quad B' = \frac{n * B}{2^8} \quad (2.1)$$

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

2.1.3. Các moment màu

Các moment màu đã được sử dụng rất thành công trong nhiều hệ thống tra cứu ảnh đặc biệt là khi ảnh chỉ chứa một đối tượng ảnh. Các thành phần moment bậc nhất (trung vị), bậc hai (phương sai) và bậc ba (độ lệch) đã được chứng minh là có thể được sử dụng rất hiệu quả để biểu diễn sự phân bố màu sắc của ảnh.

Công thức toán học để biểu diễn 3 moment này như sau:

$$\mu_i = \frac{1}{N} \sum_{j=1}^N f_{ij} \quad (2.2)$$

$$\sigma_i = \sqrt{\frac{1}{N} \sum_{j=1}^N (f_{ij} - \mu_i)^2} \quad (2.3)$$

$$s_i = \sqrt[3]{\frac{1}{N} \sum_{j=1}^N (f_{ij} - \mu_i)^3} \quad (2.4)$$

Trong đó f_{ij} là giá trị của thành phần màu thứ i của điểm ảnh thứ j và N là số lượng điểm ảnh của ảnh đó.

Sử dụng cả thành phần moment bậc ba s_i sẽ giúp tăng cường hiệu năng tra cứu so với khi chỉ sử dụng các moment bậc nhất μ_i và bậc hai σ_i . Tuy nhiên đôi khi việc sử dụng moment bậc 3 làm cho việc biểu diễn đặc trưng của ảnh nhạy cảm hơn đối với những thay đổi của cảnh nền và do đó làm giảm khả năng của hệ thống.

Do chỉ sử dụng 9 giá trị (3 moment cho mỗi màu trong bộ ba màu) để biểu diễn nội dung màu của mỗi ảnh nên việc sử dụng moment màu để biểu diễn véc tơ đặc trưng màu là rất đơn giản nếu so sánh với các phương pháp biểu diễn khác. Và cũng chính do sự đơn giản đó nên phương pháp này cho kết quả khá hạn chế.

Thông thường, phương pháp biểu diễn bằng moment màu được sử dụng trong những bước đầu tiên của quá trình tra cứu ảnh với mục đích làm giảm kích thước không gian tìm kiếm trước khi áp dụng các phương pháp phức tạp hơn để tra cứu.

2.1.4. Biểu đồ màu (Color Histogram)

Biểu đồ màu để biểu diễn nội dung màu của một bức ảnh. Biểu đồ màu dễ tính toán và rất hiệu quả để biểu diễn cả sự phân bố màu tổng quan và sự phân bố màu cục bộ của ảnh. Ngoài ra, biểu đồ màu không bị ảnh hưởng bởi sự dịch chuyển hay sự quay của ảnh và rất ít bị ảnh hưởng của tỉ lệ và góc nhìn ảnh.

Biểu đồ màu liên kết có bổ sung thêm các thông tin khác mà không làm giảm sự đơn giản của biểu đồ màu. Để đạt được điều này, người ta phải lựa chọn rất cẩn thận các đặc trưng cục bộ sẽ bổ sung vào biểu đồ màu liên kết. Mỗi phần tử trong biểu đồ màu liên kết chứa số lượng điểm ảnh trong ảnh được mô tả bởi một bộ các giá trị đặc trưng. Như vậy biểu đồ màu liên kết là lược đồ biểu đồ màu đa chiều.

Mặt khác, do biểu đồ màu không phản ánh được các thông tin mang tính không gian của các điểm ảnh, vì vậy về mặt lý thuyết, các ảnh rất khác nhau có thể có sự phân bố màu tương tự nhau. Vấn đề này rất dễ xảy ra đối với các cơ sở dữ liệu

ảnh lớn. Người ta đã đề xuất một vài cách khác nhau để khắc phục vấn đề này, một cách đơn giản nhất là chia các bức ảnh thành những phân vùng nhỏ hơn và tính biểu đồ màu của từng phân vùng. Có thể sử dụng phương pháp phân vùng đơn giản là phân thành các hình chữ nhật đồng đều hoặc phức tạp hơn là phân theo các vùng ảnh hoặc các đối tượng ảnh. Ảnh được phân vùng càng nhỏ thì độ chính xác càng cao nhưng khối lượng tính toán cũng nhiều hơn.

Ta có thể sử dụng tập hợp của các mức mà mỗi mức chỉ ra số pixel của một màu riêng biệt trong ảnh. Biểu đồ màu H cho ảnh được định nghĩa như là một vector: $H = \{H[0], H[1], H[2], \dots, H[i], \dots, H[N]\}$. Ở đây 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, và N là số mức 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 mức 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:

$$H' = \{H'[0], H'[1], H'[2], \dots, H'[i], \dots, H'[N]\} \quad (2.5)$$

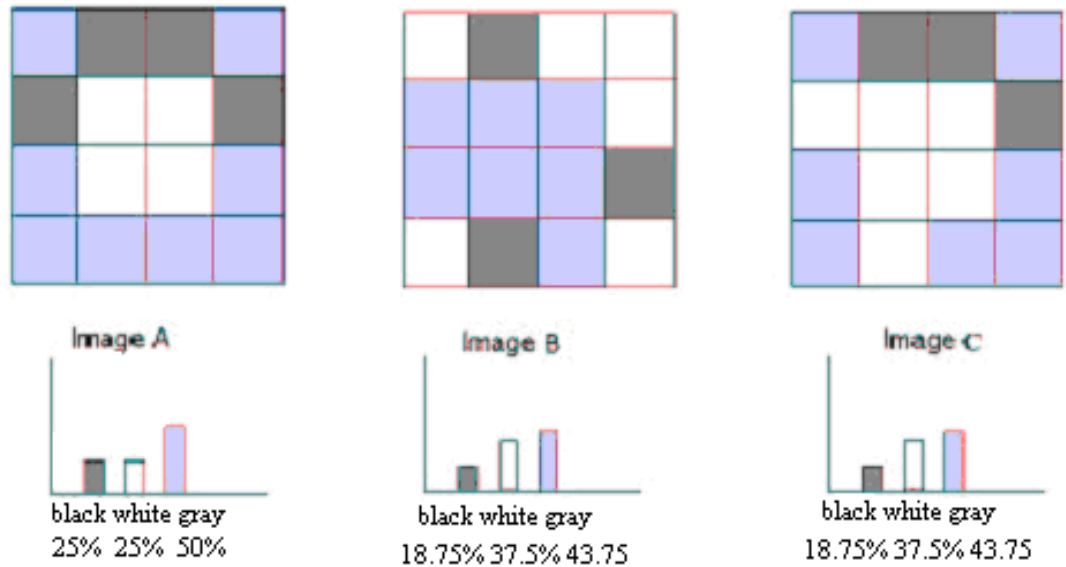
Với $H'[i] = \frac{H[i]}{P}$ và P là tổng số các điểm trong ảnh.

Một lượng tử hóa không gian màu lý tưởng sao cho những màu khác biệt không 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 mức khác nhau nhưng nó làm tăng khả năng những màu phân biệt được gán cho những mức 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 mức sẽ chứa nhiều thông tin về nội dung ảnh hơn, nhưng nó làm giảm khả năng các màu riêng biệt sẽ được gán cho các mức 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 mức nên được sử dụng trong biểu đồ màu.

2.1.4.1. Biểu đồ màu toàn cục (Global Color Histogram)

Biểu đồ màu loại này mô tả phân bố màu sử dụng tập các mức. Việc sử dụng biểu đồ màu toàn cục (gọi tắt là GCH trong luận văn 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 độ 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:



Hình 2.3. Ba ảnh và biểu đồ màu của chúng.

Trong biểu đồ màu ví dụ có 3 mức: 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 độ 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 cục là:

$$d_{GCH}(A,B) = \sqrt{(0.25-0.1785)^2 + (0.25-0.375)^2 + (0.5-0.4375)^2} = 0.153$$

$$\text{và } d_{GCH}(A,C) = d_{GCH}(A,B), \quad d_{GCH}(B,C)=0.$$

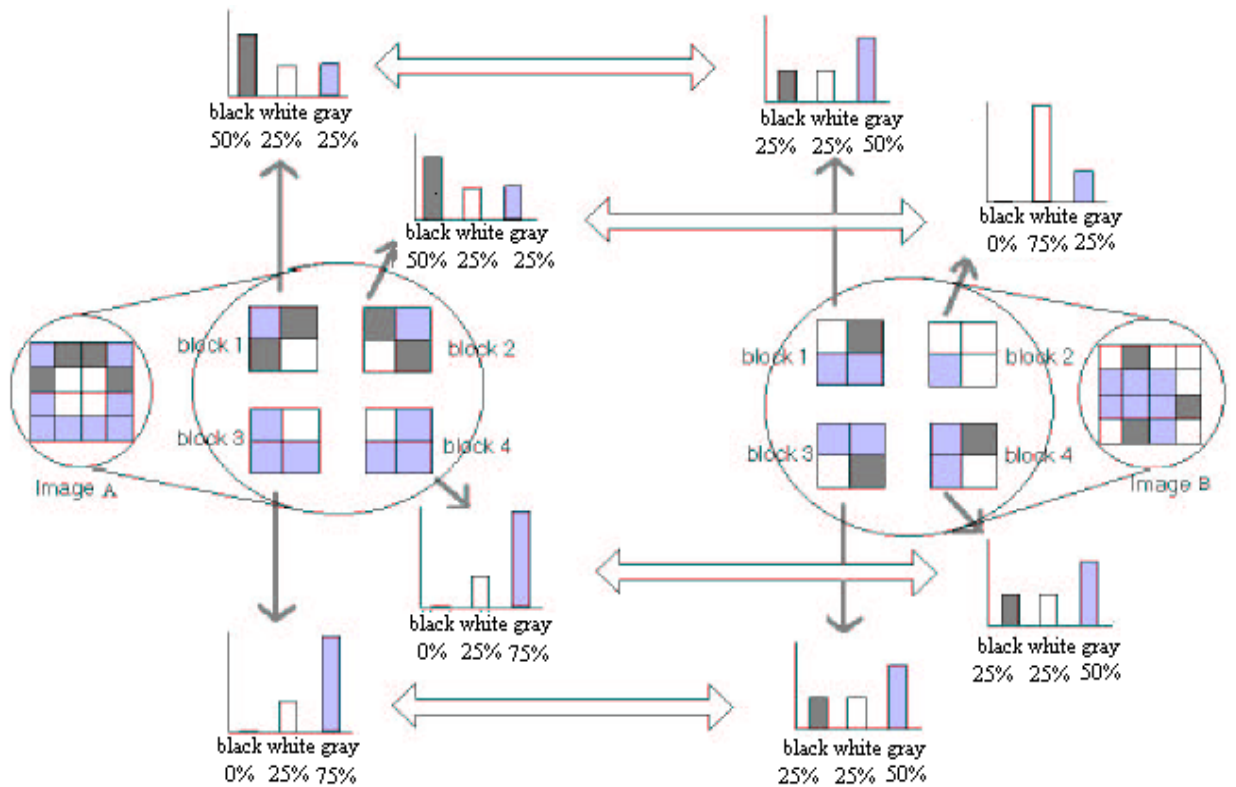
GCH là một phương pháp truyền thống cho việc tra cứu ảnh dựa trên màu sắc. Mặc dù vậy, nó không chứa các thông tin liên quan đến sự phân bố màu của các vùng. Do đó, khoảng cách giữa các ảnh đôi khi không thể chỉ ra được sự khác nhau thực sự giữa chúng. Ví dụ, khoảng cách giữa ảnh A và C khác so với khoảng 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.

2.1.4.2. Biểu đồ màu cục bộ (Local Color Histogram)

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:

$$d(Q, I) = \sum_{k=1}^M \sqrt{\sum_{i=1}^N (H_Q^k[i] - H_I^k[i])^2} \quad (2.6)$$

Ở đây M là số vùng được phân đoạn trong ảnh, N là số mức trong biểu đồ màu và $H[i]$ là giá trị của mức 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 2.4 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 2.4. Sử dụng LCH để tính toán khoảng cách giữa ảnh A và B

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

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

2.2. Phương pháp trích chọn đặc trưng theo kết cấu

Kết cấu (texture) là một tính chất quan trọng khác của ảnh. Kết cấu là một thành phần có ảnh hưởng rất quan trọng đối với sự nhận thức trực quan của con người. Tất cả mọi người đều có thể nhận ra kết cấu nhưng lại rất khó có thể định nghĩa chính xác nó là gì.

Không giống như màu sắc, kết cấu “xảy ra” trên một vùng chứ không phải tại một điểm ảnh và thường được định nghĩa bằng các mức xám.

Rất nhiều cách thể hiện kết cấu đã được nghiên cứu trong lĩnh vực nhận dạng và thị giác máy tính. Xét một cách cơ bản, các phương pháp biểu diễn kết cấu có thể được chia thành hai loại: Phương pháp cấu trúc và phương pháp thống kê.

Các phương pháp cấu trúc bao gồm các toán tử hình thái và đồ thị liên kề, mô tả kết cấu bằng cách định nghĩa các nguyên thủy cấu trúc và luật sắp đặt của chúng. Các phương pháp này tỏ ra có hiệu quả khi áp dụng trong trường hợp kết cấu thông thường.

Các phương pháp thống kê bao gồm: Phương pháp phổ năng lượng Fourier, ma trận đồng khả năng, Tamura, Phân tích Wold, trường ngẫu nhiên Markov, mô hình fractal, các bộ lọc đa phân giải như biến đổi Gabor và biến đổi dạng sóng... thể hiện kết cấu bằng sự phân bố thống kê của độ sáng của các điểm ảnh.

Phần tiếp theo sẽ trình bày một số phương pháp biểu diễn kết cấu được sử dụng rộng rãi và có hiệu quả trong các hệ thống tra cứu ảnh theo nội dung.

2.2.1. Các đặc trưng Tamura

Các đặc trưng Tamura bao gồm độ thô, độ tương phản, độ định hướng, độ tuyến tính, độ đồng đều và độ gồ ghề, được thiết kế phù hợp với sự cảm nhận của thị giác con người đối với kết cấu.

Ba thành phần đầu tiên của các đặc trưng Tamura là độ thô, độ tương phản và độ định hướng được dùng nhiều nhất trong các hệ thống tra cứu ảnh nổi tiếng như QBIC, Photobook.

Cách tính toán các đại lượng đặc trưng này như sau:

Độ thô

Độ thô được dùng để đo tính chất hạt của kết cấu. Để tính toán độ thô, tại mỗi điểm ảnh (x,y) ta tính toán một trung bình động (average moving) $A_k(x,y)$ sử dụng một cửa sổ kích thước $2^k \times 2^k$ ($k = 0, 1, \dots, 5$):

$$A_k(x, y) = \frac{1}{2^{2k}} \sum_{i=x-2^{k-1}}^{x+2^{k-1}-1} \sum_{j=y-2^{k-1}}^{y+2^{k-1}-1} g(i, j) \quad (2.7)$$

Trong đó $g(i,j)$ là độ sáng của điểm ảnh ở vị trí (i,j) .

Bước tiếp theo là tính toán sự khác nhau giữa các cặp trung bình động không chồng lấn lên nhau theo chiều dọc và chiều ngang của các điểm ảnh:

$$E_{k,h}(x,y) = |A_k(x+2^{k-1},y) - A_k(x-2^{k-1},y)| \quad (2.8)$$

$$E_{k,v}(x,y) = |A_k(x,y+2^{k-1}) - A_k(x,y-2^{k-1})| \quad (2.9)$$

Sau đó, giá trị k nào làm cho E lớn nhất theo bất kỳ chiều nào sẽ được sử dụng để đặt cho kích thước tốt nhất cho mỗi điểm ảnh, nghĩa là:

$$S_{best}(x,y) = 2^k \quad (2.10)$$

Độ thô được tính toán bằng cách lấy giá trị trung bình của S_{best} trên phạm vi toàn bức ảnh, nghĩa là:

$$F_{crs} = \frac{1}{m.n} \sum_{i=1}^m \sum_{j=1}^n S_{best}(i,j) \quad (2.11)$$

Ngoài ra người ta cũng có thể tính toán độ thô bằng cách sử dụng histogram để tính toán sự phân bố của S_{best} . Nếu so sánh với việc chỉ dùng một giá trị để thể hiện độ thô thì các tính toán độ thô kiểu histogram có thể giúp cải thiện đáng kể hiệu năng của hệ thống tra cứu ảnh vì cách này có khả năng phù hợp với các ảnh hoặc vùng ảnh có nhiều loại kết cấu.

Độ tương phản

Công thức tính độ tương phản như sau:

$$F_{con} = \frac{\sigma}{\sqrt[4]{\alpha_4}} \quad (2.12)$$

Trong đó $\alpha_4 = \frac{\mu_4}{\sigma^4}$, μ_4 là moment thứ tư (trung vị) và σ^2 là độ biến đổi. Công thức này có thể sử dụng trên toàn bức ảnh hoặc trong một vùng ảnh nào đó.

Độ định hướng

Để tính toán độ định hướng, ta nhân chập ảnh với 2 ma trận kích thước 3×3

$$\begin{matrix} -1 & 0 & 1 & & & & 1 & 1 & 1 \\ -1 & 0 & 1 & & & & 0 & 0 & 0 \\ -1 & 0 & 1 & & & & -1 & -1 & -1 \end{matrix}$$

là $\begin{matrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{matrix}$ và $\begin{matrix} -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \end{matrix}$ sau đó tính toán véc tơ gradient cho mỗi điểm ảnh.

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

$$|\Delta G| = \frac{|\Delta_H| + |\Delta_V|}{2} \quad (2.13)$$

$$\theta = \tan^{-1}\left(\frac{\Delta_V}{\Delta_H}\right) + \frac{\pi}{2} \quad (2.14)$$

trong đó ΔH và ΔV là các sai khác theo chiều ngang và chiều dọc của tích chập.

Sau đó bằng cách lượng tử hoá θ và đếm số điểm ảnh có trọng số $|\Delta G|$ lớn hơn một mức ngưỡng nào đó ta có thể xây dựng được histogram của θ và ký hiệu là HD. Histogram này sẽ có nhiều đỉnh cực đại nếu ảnh có độ định hướng cao, còn với những ảnh không có tính định hướng thì histogram này sẽ bằng phẳng hơn.

Cuối cùng, tính toán histogram tổng hợp của cả ảnh để xác định độ định hướng tổng thể dựa trên độ sắc của các đỉnh:

$$F_{dir} = \sum_p^{n_p} \sum_{\Phi \in w_p} (\Phi - \Phi_p)^2 H_D(\Phi) \quad (2.15)$$

Trong công thức này, p chạy trên tất cả các đỉnh n_p , và đối với mỗi đỉnh p , w_p là tập các bin phân bố trên đỉnh đó còn Φ_p là bin đạt giá trị tại đỉnh đó.

2.2.2. Các đặc trưng Wold

Một cách tiếp cận khác để biểu diễn kết cấu là sử dụng phân tích Wold. Phân tích Wold có 3 thành phần *độ hài hoà*, *độ phai mờ* và *độ bất định* tương ứng với các đặc điểm *tính chu kỳ*, *tính định hướng* và *tính ngẫu nhiên* của kết cấu.

Các kết cấu có tính chu kỳ có thành phần *độ hài hoà* cao, các kết cấu có tính định hướng cao có thành phần *độ phai mờ* lớn còn các kết cấu có tính cấu trúc ít hơn thì thành phần *độ bất định* lớn hơn.

Đối với trường ngẫu nhiên đồng đều thuần nhất $\{y(m,n), (m,n) \in Z^2\}$ thì phép phân tích Wold 2D sẽ cho 3 thành phần trực giao từng đôi một:

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

trong đó $u(m,n)$ là thành phần bất định còn $d(m,n)$ là thành phần tất định. Thành phần tất định lại được phân tích thành các thành phần độ hài hoà $h(m,n)$ và độ phai mờ $e(m,n)$.

Trong miền tần số ta cũng xây dựng được các công thức tương ứng:

$$F_y(\xi, \eta) = F_u(\xi, \eta) + F_d(\xi, \eta) = F_u(\xi, \eta) + F_h(\xi, \eta) + F_e(\xi, \eta) \quad (2.16)$$

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

Trong miền không gian, 3 thành phần trực giao có thể tính toán được bằng phép ước lượng khả năng tối đa (MLE) liên quan đến việc điều chỉnh qui trình tự thoái lui (AR) bậc cao, tối thiểu hàm định giá và giải hệ các phương trình tuyến tính.

Trong miền tần số, có thể tính toán được các thành phần Wold bằng cách đặt ngưỡng tổng thể cho các biên độ phổ Fourier của ảnh.

2.2.3. Mô hình tự thoái lui đồng thời (mô hình SAR)

Mô hình SAR là một biến thể của trường ngẫu nhiên Markov (MRF), được sử dụng rất hiệu quả để mô hình hoá kết cấu ảnh trong những năm gần đây. So với những mô hình MRF khác, SAR sử dụng ít tham số hơn. Trong mô hình SAR, độ chói của các điểm ảnh được coi như các biến ngẫu nhiên. Độ chói $g(x,y)$ của điểm ảnh (x,y) được coi như là một tổ hợp tuyến tính của đội chói của điểm ảnh liền kề $g(x',y')$ và mẫu nhiễu dương $\varepsilon(x,y)$ tức là:

$$g(x, y) = \mu + \sum_{(x',y') \in D} \theta(x', y') g(x', y') + \varepsilon(x, y) \quad (2.17)$$

trong đó μ là giá trị sai lệch xác định bởi giá trị trung bình trên toàn ảnh; D là tập các điểm ảnh lân cận của (x,y) ; $\theta(x',y')$ là tập các trọng số của các điểm ảnh lân cận; $\varepsilon(x,y)$ là biến ngẫu nhiên Gauss độc lập với trung bình θ và biến thiên σ^2 .

Các tham số θ và σ được sử dụng để đo kết cấu. Ví dụ, giá trị σ cao hơn thể hiện kết cấu mịn hơn, đỡ thô hơn; các giá trị $\theta(x,y+1)$ và $\theta(x,y-1)$ lớn hơn cho biết

ảnh có tính định hướng theo chiều dọc. Kỹ thuật sai số bình phương tối thiểu (LSE) hoặc phương pháp ước lượng khả năng tối đa (MLE) thường được sử dụng để ước lượng các tham số của mô hình SAR.

2.2.4. Ma trận đồng khả năng (Co-occurrence matrix)

Cách biểu diễn kết cấu bằng ma trận đồng khả năng thể hiện sự liên quan về mặt không gian của các mức xám. Định nghĩa toán học của ma trận đồng khả năng như sau:

Giả sử có một toán tử vị trí $P(i,j)$

A là một ma trận kích thước $n \times n$, phần tử $A[i][j]$ biểu thị số lần mà các điểm có mức xám (độ chói) là $g[i]$ và $g[j]$ thoả mãn toán tử P .

Đặt C là ma trận kích thước $n \times n$ tính được bằng cách chia ma trận A cho tổng số cặp điểm thoả mãn toán tử P . $C[i][j]$ là xác suất để một cặp điểm thoả mãn toán tử P có cặp giá trị $g[i], g[j]$.

C được gọi là ma trận đồng khả năng định nghĩa bởi toán tử P .

Một ví dụ về toán tử P : “ i nằm phía trên j ”, hoặc “ i nằm ở cách j một vị trí về phía phải và hai vị trí về phía dưới”.

Cũng có thể diễn tả về ma trận đồng khả năng theo cách sau đây: giả sử t là một dịch chuyển, khi đó ma trận đồng khả năng C_t của một vùng được định nghĩa cho mỗi cặp mức xám (a,b) theo công thức:

$$C_t(a,b) = \text{card}\{(s, s+t) \in R^2 \mid A[s] = a, A[s+t] = b\}$$

Ở đây, $C_t(a, b)$ là số cặp điểm (ký hiệu là $(s, s+t)$) được xác định bởi véc tơ dịch chuyển t mà a là độ xám của s và b là độ xám của $s+t$.

Ví dụ: với một ảnh có 8 mức xám và véc tơ t là một dịch chuyển một vị trí thì:

Với ảnh mẫu:

```

1  2  1  3  4
2  3  1  2  4
3  3  2  1  1

```

Ta thu được một ma trận đồng khả năng như sau:

	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0
1	0	1	2	0	0	0	0	0
2	0	1	0	2	0	0	0	0
3	0	0	1	1	0	0	0	0
4	0	1	0	0	1	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0

Trước hết người ta xây dựng những ma trận đồng khả năng dựa trên phương hướng và khoảng cách giữa các điểm ảnh. Sau đó từ những ma trận đó có thể rút ra được các con số thống kê có nghĩa về kết cấu.

Một số đặc trưng của kết cấu có thể tính được dựa vào phương pháp ma trận đồng khả năng là:

$$\text{Năng lượng: } \sum_i \sum_j C^2(i, j)$$

$$\text{Entropy: } \sum_i \sum_j C(i, j) \log C(i, j)$$

$$\text{Độ tương phản: } \sum_i \sum_j (i - j)^2 C(i, j)$$

$$\text{Tính đồng nhất: } \sum_i \sum_j \frac{C(i, j)}{1 + |i - j|}$$

Ngoài ra còn có thể xây dựng được nhiều đặc trưng khác như độ tương quan, phương sai, tổng trung bình, tổng phương sai, tổng entropy, trung vị cục bộ...

Như vậy với mỗi đặc trưng kết cấu chúng ta thu được một ma trận đồng khả năng. Những ma trận đồng khả năng này thể hiện sự phân bố không gian và sự phụ thuộc của các mức xám trong một vùng cục bộ nào đó. Mỗi phần tử (i, j) của ma trận biểu diễn xác suất xuất hiện một điểm có mức xám i và một điểm có mức xám j ở những vị trí có khoảng cách và tạo thành một góc đã được qui định trước. Dựa vào những ma trận này có thể tính toán được các con số thống kê về ảnh hay chính là các véc tơ đặc trưng cho kết cấu của ảnh đó.

2.2.5. Lọc Gabor

Bộ lọc Gabor được sử dụng khá rộng rãi để trích chọn đặc điểm của ảnh số, đặc biệt là các đặc điểm kết cấu. Lọc Gabor được coi là tối ưu xét về khả năng tối thiểu hoá những sự không chắc chắn liên kết trong không gian và tần số và thường được sử dụng như là bộ phát hiện hướng và phát hiện biên điều hướng được. Có nhiều cách tiếp cận để phát hiện các đặc điểm kết cấu dựa vào bộ lọc Gabor. Ý tưởng chính của việc sử dụng bộ lọc Gabor để trích chọn các đặc điểm kết cấu như sau:

Một hàm lọc Gabor hai chiều $g(x,y)$ được định nghĩa là:

$$g(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right) + 2\pi jWx\right] \quad (2.18)$$

Trong đó σ_x, σ_y là độ lệch tiêu chuẩn của đường bao Gauss dọc theo hướng x và y .

Từ đó có thể thu được các bộ lọc Gabor bằng cách kéo dẫn và quay hàm lọc Gabor $g(x,y)$:

$$g_{mn}(x,y) = a^{-m}g(x',y') \quad (2.19)$$

$$x' = a^{-m}(x\cos\theta + y\sin\theta) \quad (2.20)$$

$$y' = a^{-m}(-x\sin\theta + y\cos\theta) \quad (2.21)$$

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 hướng và số tỷ lệ co giãn. Hệ số co giãn a^{-m} được đưa vào để đảm bảo năng lượng độc lập với m .

Cho trước một ảnh $I(x,y)$ thì biến đổi Gabor của nó được định nghĩa như sau:

$$W_{mn}(x, y) = \int I(x, y) g_{mn}^*(x - x_1, y - y_1) dx_1 dy_1 \quad (2.22)$$

Ở đây dấu $*$ thể hiện liên hợp phức. Sau đó trung vị μ_{mn} và độ lệch tiêu chuẩn σ_{mn} của biên độ của $W_{mn}(x,y)$ là

$$f = [\mu_{00}, \sigma_{00}, \dots, \mu_{mn}, \sigma_{mn}, \Lambda, \mu_{S-1 K-1}, \sigma_{S-1 K-1}] \quad (2.23)$$

Có thể được sử dụng để biểu diễn các đặc điểm của một kết cấu thuần nhất.

2.2.6. Biến đổi dạng sóng (wavelet transform)

Cũng giống như phương pháp lọc Gabor, phương pháp biến đổi dạng sóng là cách phân tích và phân loại áp dụng cho các kết cấu nhiều chiều.

Biến đổi dạng sóng phân tích một tín hiệu thành một họ các hàm cơ sở $\psi_{mn}(x)$ bằng cách dịch chuyển và co giãn một hàm gốc $\psi(x)$, tức là:

$$\psi_{mn}(x) = 2^{-m/2} \psi(2^{-m}x - n) \quad (2.24)$$

Trong đó m và n là tham số co giãn và tham số dịch chuyển. Một tín hiệu $f(x)$ có thể được biểu diễn dưới dạng:

$$f(x) = \sum_{m,n} c_{mn} \psi_{mn}(x) \quad (2.25)$$

Việc tính toán các biến đổi dạng sóng của các tín hiệu hai chiều có liên quan tới quá trình lọc đệ qui và lấy mẫu lại. Ở mỗi mức thì tín hiệu được phân tích thành 4 dải tần số con là LL, LH, HL và HH, trong đó L ký hiệu cho tần số thấp và H ký hiệu cho tần số cao.

Hai dạng biến đổi dạng sóng chủ yếu được dùng trong phân tích kết cấu ảnh là *biến đổi dạng sóng theo kiểu hình hình chóp* (PWT) và *biến đổi dạng sóng theo kiểu hình cây* (TWT).

PWT phân tích một cách đệ qui dải tần số LL, tuy nhiên đối với một số loại kết cấu thì những thông tin quan trọng nhất thường xuất hiện ở các kênh tần số trung bình. Để khắc phục nhược điểm này của PWT thì TWT còn có thể phân tích ở các dải tần số khác như LH, HL hoặc HH nếu cần.

Sau quá trình phân tích, có thể xây dựng các véc tơ đặc trưng bằng cách sử dụng trung vị và độ lệch chuẩn của phân bố năng lượng của mỗi dải tần con (sub-band) tại mỗi mức đệ qui.

Khi thực hiện phân tích mức thì PWT cho kết quả là một véc tơ đặc trưng có $3 \times 4 \times 2$ thành phần. Đối với TWT, véc tơ đặc trưng phụ thuộc vào thứ tự phân tích các dải tần số con. Có thể xây dựng được một cây phân tích cố định bằng cách phân tích tuần tự các dải tần LL, LH và HH, kết quả cho ra sẽ là một véc tơ đặc trưng có 52×2 thành phần.

Lưu ý là trong ví dụ này thì véc tơ đặc trưng kết quả của phân tích PWT chỉ là tập con của véc tơ do phân tích TWT sinh ra. Ngoài ra qua so sánh sự khác nhau

của véc tơ đặc trưng thu được khi sử dụng các phương pháp biến đổi dạng sóng khác nhau, người ta thấy rằng việc lựa chọn bộ lọc dạng sóng không ảnh hưởng lớn lắm đến các phân tích kết cấu ảnh.

2.3. Phương pháp trích chọn đặc trưng theo hình dạng

Các đặc điểm phát hiện biên của các vùng ảnh và các đối tượng ảnh được sử dụng trong rất nhiều hệ thống tra cứu ảnh. So với các đặc điểm về màu sắc và các đặc điểm về kết cấu thì các đặc điểm về hình dạng thường chỉ được sử dụng sau khi ảnh đã phân thành các vùng hoặc các đối tượng ảnh. Nhưng do việc phân vùng và tách đối tượng ảnh khó thu được kết quả tốt nên việc sử dụng các đặc điểm hình dạng để tra cứu ảnh thường bị bó hẹp trong một số ứng dụng mà ở đó các vùng ảnh hoặc đối tượng ảnh đã được tách biệt rõ ràng.

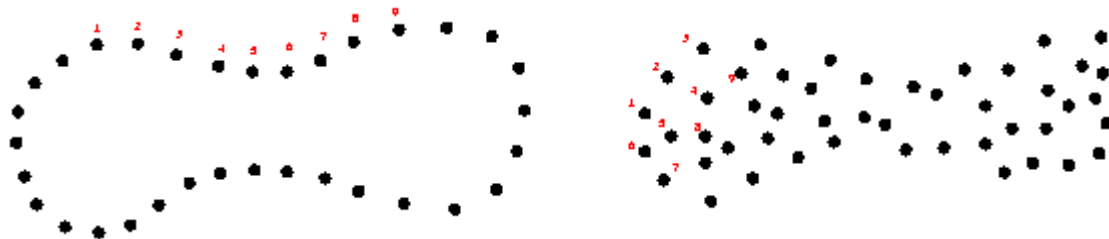
Các phương pháp trích chọn đặc điểm hình dạng thường được chia thành hai loại là trích chọn dựa theo đường biên (xấp xỉ đa giác, mô hình phần tử hữu hạn, mô tả hình dạng theo Fourier) và trích chọn dựa theo vùng ảnh (mô hình thống kê).

Một phương pháp trích chọn đặc điểm hình dạng tốt phải đảm bảo yêu cầu là phải không phụ thuộc vào vị trí, góc quay hay sự co giãn của đối tượng ảnh.

Trước khi áp dụng các phương pháp trích chọn đặc điểm hình dạng, các đối tượng ảnh cần phải được tách ra khỏi ảnh. Giả sử là trong mỗi ảnh chỉ có một đối tượng ảnh duy nhất, nhiệm vụ của hệ thống trước hết là phải tách được đối tượng ảnh ra khỏi nền ảnh.

Cách biểu diễn hình dạng của đối tượng ảnh có thể chia thành hai kiểu: Theo đường bao quanh (biên) và theo vùng

Cách biểu diễn theo đường viền bao quanh chỉ sử dụng đường biên bên ngoài của hình dạng, điều này có thể thực hiện được bằng cách mô tả vùng đang quan tâm bằng cách đặc tính bên ngoài của nó tức là các điểm ảnh dọc theo đường viền bao quanh đối tượng ảnh. Cách biểu diễn theo vùng sử dụng cả vùng ảnh bằng cách mô tả vùng đang quan tâm bằng các đặc tính bên trong tức là các điểm ảnh ở bên trong vùng đó.



Hình 2.5: Biểu diễn hình dạng theo đường biên và theo vùng

2.3.1. Biên và các phương pháp phát hiện biên

Nếu đã biết là một đối tượng có một biên rời rạc bao quanh và có thể tìm được một điểm nằm trên biên đó thì từ điểm đó có thể đi theo đường biên bao quanh đối tượng và quay trở lại điểm xuất phát. Dò biên là một thao tác rất quan trọng, đặc biệt là khi cần xác định xem một điểm ảnh có nằm trong một vùng ảnh nào đó hay không.

Một điểm ảnh được gọi là biên nếu ở đó có sự thay đổi đột ngột về mức xám. Tập hợp các điểm biên tạo thành biên của ảnh.

Phương pháp phát hiện biên trực tiếp

Phương pháp này làm nổi biên dựa vào sự biến thiên độ xám của ảnh. Kỹ thuật chủ yếu dùng để phát hiện biên là kỹ thuật đạo hàm.

Nếu lấy đạo hàm bậc nhất của ảnh ta có phương pháp Gradient. Vì ảnh số là các tín hiệu rời rạc nên không tồn tại đạo hàm nên thực chất phương pháp này chỉ là mô phỏng và xấp xỉ đạo hàm bằng kỹ thuật nhân chập.

Một số dạng xấp xỉ đạo hàm bậc nhất:

$$A = \begin{pmatrix} -1 & -1 \\ 1 & 1 \end{pmatrix} : \text{xấp xỉ đạo hàm theo } y$$

$$B = \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix} : \text{xấp xỉ đạo hàm theo } x$$

$$H_x = \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix}; H_y = \begin{pmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

$$H_x = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}; H_y = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix}$$

Các kỹ thuật Gradient làm việc khá tốt khi độ xám thay đổi rõ nét.

Nếu lấy đạo hàm bậc hai của ảnh ta có kỹ thuật Laplace.

Toán tử Laplace được định nghĩa như sau:

$$\begin{aligned} \text{Ta có: } \quad \nabla^2 f &= \frac{\partial^2 f}{dx^2} + \frac{\partial^2 f}{dy^2} \\ \frac{\partial^2 f}{\partial x^2} &= 2f(x, y) - f(x-1, y) - f(x+1, y) \\ \frac{\partial^2 f}{\partial y^2} &= 2f(x, y) - f(x, y-1) - f(x, y+1) \end{aligned}$$

$$\text{Vậy: } \nabla^2 f = -f(x-1, y) - f(x, y-1) + 4f(x, y) - f(x, y+1) - f(x+1, y)$$

Một số dạng xấp xỉ đạo hàm bậc hai của ảnh:

$$H = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix};$$

$$H_1 = \begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix}; H_2 = \begin{pmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{pmatrix}; H_3 = \begin{pmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{pmatrix}$$

Các kỹ thuật Laplace làm việc hiệu quả với ảnh có mức xám thay đổi chậm, miền chuyển tiếp trải rộng.

Phương pháp phát hiện biên gián tiếp

Nếu bằng một cách nào đó ta phân được ảnh thành các vùng thì ranh giới giữa các vùng là đó chính là biên. Kỹ thuật dò biên và kỹ thuật phân vùng ảnh là hai bài toán đối ngẫu nhau bởi vì dò biên để thực hiện phân lớp đối tượng mà khi đã phân lớp xong thì có nghĩa là đã phân vùng được ảnh và ngược lại khi đã phân vùng

được ảnh tức là đã phân lớp được thành các đối tượng do đó ta có thể phát hiện được biên.

Kỹ thuật dò biên gián tiếp đơn giản

Giả sử đã tìm được một vị trí (x, y) nằm trên biên của một vùng ảnh hoặc đối tượng ảnh nào đó.

Đánh dấu điểm đó là "đã sử dụng" (để điểm đó không bị sử dụng lại) và đánh giá tất cả giá trị gradient Sobel 3×3 (hoặc lớn hơn) có trung tâm lần lượt là các điểm trong 8 điểm lân cận với (x, y) .

Chọn ra ba điểm có biên độ gradient tuyệt đối lớn nhất. Đẩy vị trí của ba điểm đó vào một mảng có 3 cột, mỗi cột tương ứng với vị trí của một điểm, sắp xếp thành từng hàng theo độ lớn của biên độ gradient. Chọn điểm có biên độ gradient lớn nhất.

Bây giờ điểm này sẽ là một trong 8 hướng từ 0 đến 7 xung quanh điểm (x, y) sắp xếp theo mô hình sau (trong đó * là vị trí điểm (x, y)):

0	1	2
7	*	3
6	5	4

Ví dụ, nếu biên độ gradient cực đại đã tìm được bằng toán tử Sobel với trung tâm là điểm $(x+1, y)$ thì hướng sẽ là 3. Gọi hướng của dịch chuyển là d .

Đây là một kỹ thuật dò biên đơn giản, tuy nhiên vẫn đề có thể xảy ra là thời gian tiêu tốn khá lớn.

Kỹ thuật dò biên gián tiếp bằng cách xác định chu tuyến của đối tượng ảnh.

Kỹ thuật này chỉ xét với ảnh nhị phân vì mọi ảnh đều có thể đưa về ảnh nhị phân bằng kỹ thuật phân ngưỡng.

Ký hiệu F là tập các điểm vùng (điểm đen), F' là tập các điểm nền.

- Định nghĩa chu tuyến:

Chu tuyến của một đối tượng ảnh là dãy các điểm ảnh của đối tượng p_0, p_1, \dots, p_n sao cho:

+ $\forall i, \exists Q$ không thuộc đối tượng ảnh là 4-láng giềng của p_i .

+ p_i và p_{i+1} là các 8-láng giềng của nhau

+ p_0 trùng với p_n

- Định nghĩa chu tuyến đối ngẫu

Chu tuyến $c = \langle p_1, p_2, \dots, p_n \rangle$, $c_{\perp} = \langle Q_1, Q_2, \dots, Q_m \rangle$ được gọi là đối ngẫu của nhau nếu:

+ $\forall i, \exists j, k$ sao cho:

1. Q_j là 4-láng giềng của p_i

2. Q_k là 8-láng giềng của p_{i+1}

3. Q_j và Q_k là 8-láng giềng của nhau.

+ Nếu p_i là nền thì Q_j là vùng và ngược lại.

+ Các điểm Q_j nằm về một phía với p_i .

- Thuật toán dò biên tổng quát như sau:

Bước 1: Xác định cặp nền-vùng xuất phát: cặp nền-vùng xuất phát được xác định bằng cách duyệt ảnh lần lượt từ trên xuống dưới, từ trái sang phải và kiểm tra theo định nghĩa cặp nền-vùng.

Bước 2: Xác định cặp nền-vùng tiếp theo.

Bước 3: Lựa chọn điểm biên

Bước 4: Nếu gặp lại cặp xuất phát thì dừng, nếu không quay lại bước 2.

2.3.2. Xử lý ảnh trong miền tần số và biến đổi Fourier

Nhiều quá trình xử lý tín hiệu có thể được thực hiện trong một không gian toán học như miền tần số. Để biểu diễn dữ liệu trong miền tần số cần phải thực hiện một số biến đổi thông qua các phép biến đổi. Phép biến đổi thường được sử dụng nhiều nhất là biến đổi Fourier.

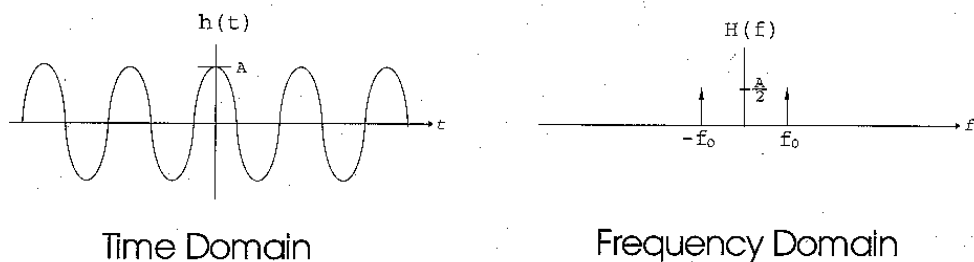
Theo nghiên cứu của Fourier thì tất cả các tín hiệu tuần hoàn đều có thể được tổng hợp từ một loạt các tín hiệu sin có tần số và biên độ khác nhau.

Có nghĩa là bất kỳ tín hiệu nào cũng được tạo thành từ những thành phần tần số khác nhau, điều này áp dụng được cho cả các tín hiệu một chiều như tín hiệu âm tần đưa ra loa hoặc tín hiệu hai chiều như ảnh số chẳng hạn.

Tần số không gian của một ảnh biểu thị tốc độ thay đổi độ chói của các điểm ảnh.

Cách dễ nhất để xác định các thành phần tần số của tín hiệu là khảo sát tín hiệu đó trong miền tần số. Miền tần số biểu diễn độ lớn (cường độ) của các thành phần tần số khác nhau của một tín hiệu.

Xét một ví dụ đơn giản: tín hiệu đầu vào có dạng hình cosin được biểu diễn trong miền thời gian và miền tần số như hình vẽ dưới đây :



Hình 2.6: Miền thời gian và miền tần số

Trong hình vẽ trên chỉ có một thành phần hình sin nên chỉ có một giá trị tần số biểu diễn trong miền tần số.

Biến đổi Fourier cho tín hiệu 2 chiều được biểu diễn qua công thức toán học:

$$h(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H(u, v) e^{j2\pi(ux+vy)} dudv \quad (2.26)$$

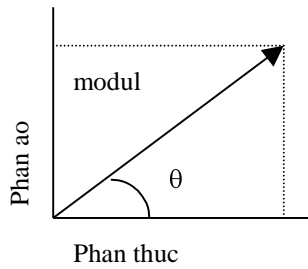
trong đó $j = \sqrt{-1}$, $e^{\pm jx} = \cos(x) \pm j \sin(x)$

Cũng có thể biến đổi dữ liệu ảnh từ miền tần số về miền không gian thông qua phép biến đổi Fourier ngược.

$$h(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H(u, v) e^{-j2\pi(ux+vy)} dudv \quad (2.27)$$

Trong miền tần số, u thể hiện tần số không gian dọc theo trục x và v là thể hiện tần số không gian dọc theo trục y. Trung tâm của ảnh là gốc của trục tọa độ u, v.

Biến đổi Fourier yêu cầu đầu vào là dạng số phức. Có thể biểu diễn số phức thông qua cường độ và góc pha như sau:

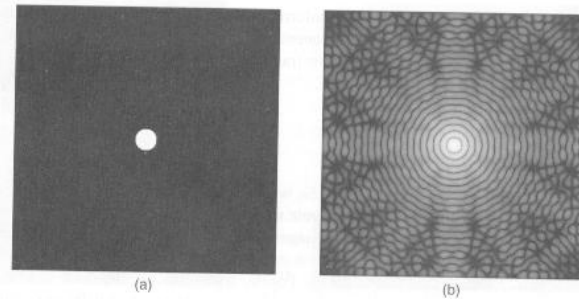


$$|H(u, v)| = \sqrt{R^2(u, v) + I^2(u, v)} \quad (2.28)$$

$$\theta(u, v) = \tan^{-1} \left[\frac{I(u, v)}{R(u, v)} \right] \quad (2.29)$$

trong đó $R(u, v)$ là phần thực còn $I(u, v)$ là phần ảo. Biên độ của số phức chính là biên độ của thành phần hình sin trong công thức của biến đổi Fourier.

Hình vẽ sau mô tả ảnh của một điểm sáng và ảnh thu được sau phép biến đổi Fourier:



Hình 2.7: Ảnh thực (a) và ảnh thu được sau biến đổi Fourier (b)

Mỗi điểm ảnh trong phổ tần số biểu thị sự thay đổi của tần số không gian của một chu kỳ theo chiều rộng của ảnh gốc. Góc tọa độ (ở giữa ảnh) biểu diễn thành phần cố định (hay còn gọi là thành phần 1 chiều của ảnh). Nếu tất cả điểm ảnh đều xám thì sẽ chỉ có một giá trị trong phổ tần số và điểm đó chính là góc tọa độ.

Điểm ảnh tiếp theo bên phải góc tọa độ biểu thị một chu kỳ theo chiều rộng của ảnh, điểm ảnh cạnh đó biểu diễn 2 chu kỳ theo chiều rộng của ảnh...các điểm ảnh càng xa góc tọa độ biểu diễn cho thành phần tần số không gian càng cao.

Biến đổi Fourier rời rạc (DFT)

Khi làm việc với ảnh số, chúng ta không thể có được các tín hiệu liên tục mà phải làm việc với một số hữu hạn các mẫu rời rạc, các mẫu này là các điểm ảnh tạo nên bức ảnh. Vì vậy để phân tích ảnh số phải sử dụng biến đổi Fourier rời rạc.

Công thức biến đổi Fourier rời rạc cho ảnh kích thước $M \times N$ là:

$$H(u, v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} h(x, y) e^{-2\pi j \left(\frac{ux}{M} + \frac{vy}{N} \right)} \quad (2.30)$$

Công thức để chuyển về miền không gian sẽ là:

$$h(x, y) = \frac{1}{MN} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} H(u, v) e^{2\pi j \left(\frac{ux}{M} + \frac{vy}{N} \right)} \quad (2.31)$$

Biến đổi Fourier nhanh (FFT)

Phép biến đổi Fourier rời rạc có độ phức tạp rất cao, cần phải có N^2 phép nhân số phức khi tính chuỗi Fourier cho N phần tử.

Năm 1942 người ta phát hiện ra rằng có thể chia dãy biến đổi Fourier rời rạc có độ dài N thành 2 dãy có độ dài mỗi dãy là $N/2$. Sau đó mỗi dãy còn lại có thể chia đôi tiếp đến khi chỉ còn dãy chỉ còn 2 phần tử.

Kỹ thuật chia để trị được sử dụng để xây dựng phép biến đổi Fourier nhanh (FFT), thuật toán này giúp làm giảm độ phức tạp của thuật toán từ cấp N^2 xuống cấp $N \log N$.

Để thực hiện trước khi thực hiện FFT, dãy tín hiệu vào phải có số phần tử là 2^N .

Trong xử lý ảnh, tính chất chia đôi được của dãy Fourier cho phép chúng ta tách phép biến đổi một phép biến đổi 2 chiều thành 2 phép biến đổi 1 chiều, có nghĩa là ta sẽ tính biến đổi Fourier cho từng hàng trước sau đó biến đổi Fourier cho các cột.

Hai thuật toán cơ bản để thực hiện biến đổi Fourier nhanh là thuật toán con bướm (Butterfly) để tính toán cho 2 phần tử và thuật toán Đảo bit (Bit reversal) để sắp xếp dãy đầu vào theo thứ tự hợp lý để thực hiện biến đổi.

Ngoài biến đổi Fourier, người ta còn sử dụng rất nhiều phép biến đổi khác như biến đổi Harley, biến đổi Hough ...

Mô tả Fourier

Bộ mô tả Fourier mô tả hình dạng của đối tượng ảnh bằng một biến đổi Fourier của đường biên của đối tượng. Một lần nữa ta lại coi biên của một đối tượng

ảnh 2D là dãy liên tiếp các điểm ảnh biên nằm cạnh nhau (x_s, y_s) , trong đó $0 \leq s \leq N-1$ và N là tổng số điểm ảnh nằm ở biên của đối tượng. Thế thì có thể định nghĩa ba loại biểu diễn đường biên là biểu diễn bằng *độ cong*, bằng *khoảng cách trọng tâm* và bằng *hàm tọa độ phức*.

Độ cong $K(s)$ tại một điểm s nằm trên biên có thể được định nghĩa là tốc độ thay đổi hướng của tiếp tuyến của đường biên tại điểm đó, tức là:

$$K(s) = \frac{d\theta(s)}{ds} \quad (2.32)$$

trong đó $\theta(s)$ là hàm xoay của đường biên (xem công thức 2.41).

Khoảng cách trọng tâm được định nghĩa là hàm khoảng cách giữa giữa một điểm ảnh nằm trên biên và trọng tâm (x_c, y_c) của đối tượng ảnh:

$$R(s) = \sqrt{(x_s - x_c)^2 + (y_s - y_c)^2} \quad (2.33)$$

Hàm tọa độ phức đơn giản là biểu diễn tọa độ của các điểm ảnh biên bằng một số phức:

$$Z(s) = (x_s - x_c) + j(y_s - y_c) \quad (2.34)$$

Biến đổi Fourier của ba kiểu biểu diễn này của đường biên tạo thành một tập hợp các hệ số phức thể hiện hình dạng của một đối tượng ảnh trong miền tần số. Các hệ số tần số thấp hơn biểu diễn các thuộc tính tổng quan của hình dạng trong khi các hệ số tần số cao hơn biểu diễn các chi tiết của hình dạng.

Để không bị phụ thuộc vào góc quay của đối tượng (tức là việc mã hoá đường biên không bị ảnh hưởng bởi cách chọn điểm tham chiếu) thì ta chỉ sử dụng biên độ của các hệ số phức và bỏ qua thành phần pha.

Để không bị phụ thuộc vào tỉ lệ thì ta phải chia biên độ của các hệ số phức cho biên độ của thành phần một chiều DC hay là cho hệ số khác 0 đầu tiên.

Bản thân các cách biểu diễn đường biên này đã không phụ thuộc vào sự tịnh tiến của đối tượng ảnh.

Mô tả Fourier của độ cong là:

$$f_K = \left[|F_1|, |F_2|, \dots, |F_{M/2}| \right] \quad (2.35)$$

Mô tả Fourier của khoảng cách trọng tâm là:

$$f_R = \left[\frac{|F_1|}{|F_0|}, \frac{|F_2|}{|F_0|}, \dots, \frac{|F_{M/2}|}{|F_0|} \right] \quad (2.36)$$

trong đó F_i là thành phần hệ số thứ i của biến đổi Fourier. Ở đây ta chỉ quan tâm đến các trục tần số dương bởi vì các hàm độ cong và hàm khoảng cách trọng tâm là các hàm thực do đó biến đổi Fourier của chúng đối xứng nhau.

Mô tả Fourier của hàm tọa độ phức là:

$$f_Z = \left[\frac{|F_{-(M/2-1)}|}{F_1}, \dots, \frac{|F_{-1}|}{|F_1|}, \frac{|F_2|}{|F_1|}, \dots, \frac{|F_{M/2}|}{|F_1|} \right] \quad (2.37)$$

Trong đó F_1 là thành phần khác 0 đầu tiên sử dụng để chuẩn hoá các hệ số biến đổi. Ở đây cả thành phần tần số âm và dương đều được xét đến. Thành phần một chiều DC phụ thuộc vào vị trí của đối tượng và do đó bị bỏ qua.

Để đảm bảo là các đặc trưng hình dạng thu được của tất cả các đối tượng ảnh trong cơ sở dữ liệu có cùng độ dài thì đường biên của mỗi đối tượng sẽ phải được lấy mẫu lại thành M mẫu trước khi thực hiện biến đổi Fourier. Ví dụ M có thể đặt bằng $2^m = 64$ để có thể thực hiện biến đổi Fourier bằng phương pháp biến đổi Fourier nhanh.

2.4. Độ đo khoảng cách và độ đo tương tự

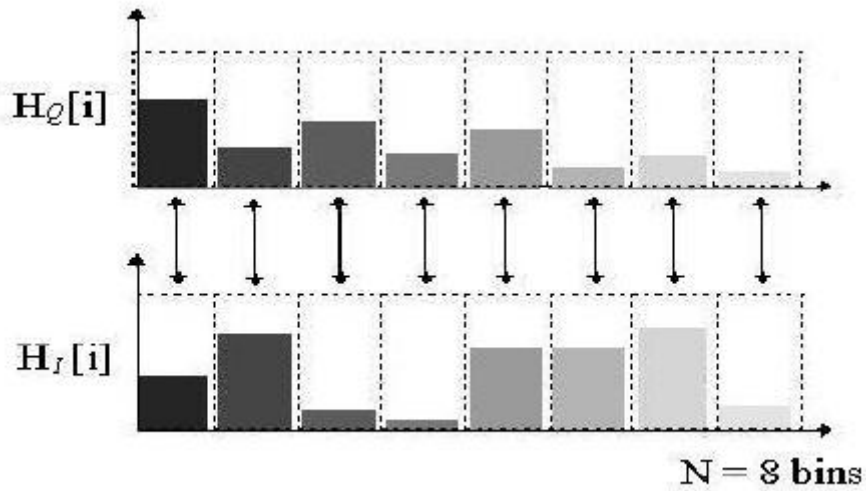
2.4.1. Độ đo khoảng cách

2.4.1.1. Độ đo khoảng cách Minkowski

Trong độ đo khoảng cách dạng Minkowski chỉ so sánh những mức giống nhau giữa các biểu đồ màu (như hình 2.5) và nó được định nghĩa như sau:

$$d(Q, I) = \sum_{i=1}^N |H_Q[i] - H_I[i]|^r \quad (2.38)$$

Với Q và I là 2 ảnh, N là số mức 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 mức). $H_Q[i]$ là giá trị của mức i trong biểu đồ màu, H_Q đại diện cho ảnh Q và $H_I[i]$ là giá trị của mức i trong biểu đồ màu H_I đại diện cho ảnh I .



Hình 2.8: 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 luận văn này căn bậc hai của bình phương khoảng cách Euclidean

$$d(Q, I) = \sqrt{\sum_{i=1}^N (H_Q[i] - H_I[i])^2} \quad (2.39)$$

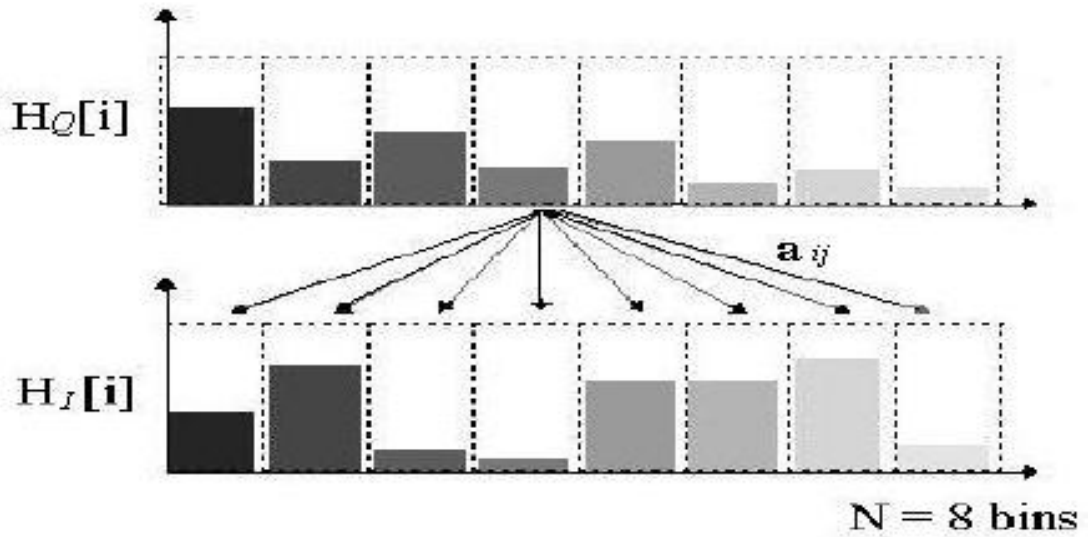
Sẽ được sử dụng để tính toán khoảng cách giữa hai biểu đồ màu

2.4.1.2 Độ đo khoảng cách Quadratic

Dự án QBIC sử dụng độ đo khoảng cách màu Quadratic so sánh không chỉ các mức giống nhau mà còn so sánh nhiều mức khác nhau giữa các biểu đồ màu (như hình 2.6) và nó được định nghĩa như sau:

$$d(Q, I) = (H_Q - H_I)^T A (H_Q - H_I) \quad (2.40)$$

Với Q và I là 2 ảnh, H_Q là biểu đồ màu của ảnh Q và H_I nó là biểu đồ màu của ảnh I , $A = [a_{i,j}]$ là một ma trận cỡ $N \times N$ với N là số mức 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 độ đo khoảng cách dạng Minkowski đó là các mức trong biểu đồ màu hoàn toàn không liên quan đến nhau.



Hình 2.9: Độ đo khoảng cách Quadretic

2.4.1.3. Độ đo khoảng cách Non-histogram

Stricker và Orengo 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 của mỗi kênh màu, và chúng được định nghĩa như sau:

$$E_i = \frac{1}{F} \sum_{j=1}^F P_{ij} ; \quad (2.41)$$

$$\delta_i = \left(\frac{1}{F} \sum_{j=1}^F (P_{ij} - E_i)^2 \right)^{\frac{1}{2}} ; \quad (2.42)$$

$$S_i = \left(\frac{1}{F} \sum_{j=1}^F (P_{ij} - E_i)^3 \right)^{\frac{1}{3}} \quad (2.43)$$

Ở đây \$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\$. \$\delta_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\$ và \$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:

$$d(Q, I) = \sum_{i=1}^r \left[W_{i1} |E_i^Q - E_i^I| + W_{i2} |\delta_i^Q - \delta_i^I| + W_{i3} |S_i^Q - S_i^I| \right] \quad (2.44)$$

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

2.4.1.4. Khoảng cách Mahalanobis

Hệ khoảng cách Mahalanobis phù hợp khi mà mỗi kích thước của véc tơ đặc trưng hình ảnh phụ thuộc vào mỗi kích thước khác và nó thuộc vào tầm quan trọng khác. Hệ khoảng cách Mahalanobis được định nghĩa như sau:

$$d(Q, I) = \sqrt{(H_Q - H_I)^T C^{-1} (H_Q - H_I)} \quad (2.45)$$

Trong đó C là ma trận tương quan của các véc tơ đặc trưng. Khoảng cách Mahalanobis có thể được đơn giản hoá nếu các kích thước đặc trưng độc lập.

Trong trường hợp đó chỉ có sự tương quan của mỗi thành phần đặc trưng, c_i được cần đến.

$$d(Q, I) = \sum_{i=1}^n (H_Q - H_I)^2 / c_i \quad (2.46)$$

2.4.2. Độ đo tương tự

2.4.2.1. Độ phân kỳ Kullback-Leibler

Độ phân kỳ Kullback-Leibler giữa hai ảnh I và J được định nghĩa là:

$$D(I, J) = \sum_i f_i(I) \log \frac{f_i(I)}{f_i(J)} \quad (2.47)$$

Độ phân kỳ Kullback – Leibler thường được sử dụng để đánh giá độ tương tự về kết cấu.

2.4.2.2. Độ phân kỳ Jeffrey

Độ phân kỳ Jeffrey được định nghĩa là:

$$D(I, J) = \sum_i f_i(I) \log \frac{f_i(I)}{f_i^{\wedge}} + f_i(J) \log \frac{f_i(J)}{f_i^{\wedge}} \quad (2.48)$$

trong đó $f_i^{\wedge} = [f_i(I) + f_i(J)]/2$.

Khác với độ phân kỳ Kullback – Leibler, độ phân kỳ Jeffrey có tính đối xứng.

2.4.2.3. Độ đo tương tự hình dạng toàn cục

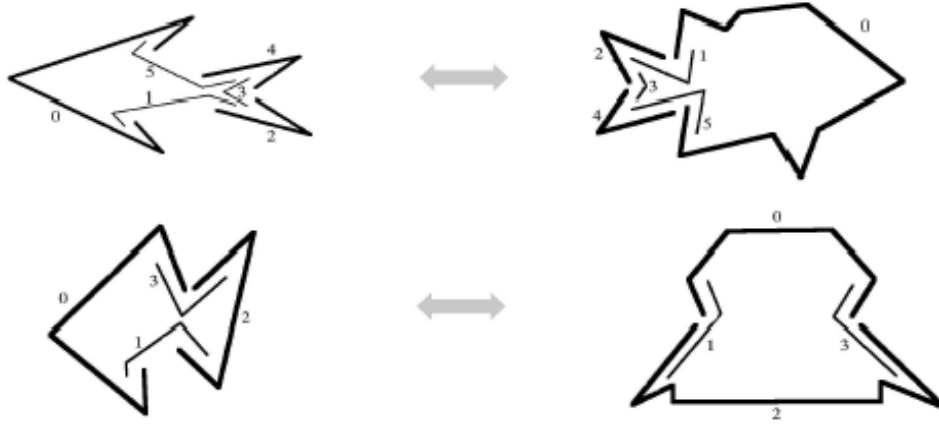
Độ đo tương tự hình dạng toàn cục là nền tảng của sự tương tự từng phần. Hiện có rất nhiều phương pháp để xác định độ tương tự giữa những đường cong đa giác. Mặc dù vậy, khi độ đo này được sử dụng để tìm cửa sổ tối ưu để điều khiển quá trình đơn giản hoá và cuối cùng cung cấp một giá trị bằng số biểu diễn sự tương tự từng phần của những hình dạng phức tạp bất kỳ thì nó phải hiệu quả với những biến dạng đường cong không đồng bộ. Phần này sử dụng phiên bản cải tiến của độ đo tương tự hình dạng cục bộ (VPS) được giới thiệu trong phần trước. Giống như tất cả các độ đo tương tự hình dạng toàn cục, VPS đòi hỏi toàn bộ đối tượng xuất hiện khi đưa vào.

Để tính toán VPS giữa hai đường cong, mối quan hệ hợp lý nhất của các cung lồi tối đa được thiết lập. Để đạt được điều này, đầu tiên người ta chia những đường cong đa giác thành những cạnh lồi lớn nhất. Chú ý rằng một cạnh lồi của đường bao đối tượng có thể là lồi hoặc lõm đối với bề mặt đối tượng. Vì sự so sánh đơn giản 1-1 của những cạnh lồi lớn của hai đường bao đa giác ít được sử dụng do thực tế các cạnh có thể chứa số lượng khác nhau các cung và thậm chí những hình dạng tương tự có thể có những đặc điểm nhỏ khác nhau, nên cho phép sự so sánh một - một, một - nhiều, nhiều - một với sự tương ứng của những cung lồi lớn. Ý chính ở đây là có ít nhất một trong số các đường bao cung lồi lớn nhất tương ứng với một phần các đường bao khác bao gồm những cung lồi lớn gần kề. Trong trường hợp này những phần tương ứng của những đường bao có thể được nhận ra với những phần đối tượng nhìn thấy. Hai ví dụ tương ứng được chỉ ra trong hình dưới.

Thuận lợi chính của VPS là dựa trên sự tương ứng của những phần nhìn thấy, chúng được định nghĩa như dãy các cặp bộ phận của những đường bao A và B

$$C = ((P_1A, P_1B), \dots, (P_nA, P_nB)) \quad (2.49)$$

Với $A = P_1A \cup \dots \cup P_nA$ và $B = P_1B \cup \dots \cup P_nB$



Hình 2.10 : Tính toán độ tương tự bởi độ đo tương tự hình dạng toàn cục.

Thứ tự của các phần là nhất định và ít nhất một phần trong số mỗi cặp là cung lồi. Sự tương tự hình dạng s được định nghĩa như là mức tối thiểu toàn cục của tổng các khoảng cách $d_{arcs}(P_iA, P_iB)$ trên toàn bộ những tương ứng có thể với độ đo tương tự đa giác được định nghĩa như bên dưới.

Cơ sở tương tự của những đa giác d_{arcs} được định nghĩa bằng việc sử dụng các hàm tag của chúng để biểu diễn. Hàm tag được gọi là hàm đảo là hàm gồm nhiều giá trị tạo nên đa giác trong khoảng $[0, 2\pi]$ bằng việc chỉ biểu diễn các hướng góc của những đoạn thẳng. Các độ dài cung được chuẩn hoá thành một trước khi ánh xạ trong không gian tag. Sự biểu diễn này đã được sử dụng trong đồ họa máy tính. Với kí hiệu hàm tag là Tg thì độ tương tự được định nghĩa như sau:

$$d_{arcs}(C, D) = \left(\int_0^1 |Tg_C(s) - Tg_D(s) + \theta(C, D)| ds \right) \cdot \max \left\{ \frac{l(C)}{l(D)}, \frac{l(D)}{l(C)} \right\} \quad (2.50)$$

Với $l(C)$ biểu thị độ dài của C và tích phân được tính trên độ dài cung s . Hằng số $\theta(C, D)$ được chọn để tối ưu hóa tích phân (nó tính cho các hướng khác nhau của những cung) và được tính như sau:

$$\theta(C, D) = \int_0^1 (Tg_C(s) - Tg_D(s))^2 ds \quad (2.51)$$

2.4.2.4. Sự tính toán hình dạng tương tự từng phần tối ưu

Cho một truy vấn đường gấp khúc Q và một đường gấp khúc gốc T, chúng ta đối diện với hai mục tiêu liên quan: (1) Khoanh vùng phần P của đường gấp khúc gốc T cái tương tự nhất với Q và (2) chuẩn hoá P thành truy vấn gấp khúc Q. Cả hai mục tiêu sẽ đạt được bởi quá trình đơn giản hoá của T trong ngữ cảnh của Q được mô tả trong phần này. Để đạt được những mục tiêu này chúng ta cần phải có độ đo hình dạng tương tự toàn cục chất lượng cao, cái được gọi là s và nó có thể được sử dụng để so sánh hai đường gấp khúc.

Một đường gấp khúc T có thể được xác định như là tập các đỉnh được sắp xếp $T = \{ t_1, \dots, t_n \}$. Mục đích là tìm và di chuyển một tập con S_Q^* của các đỉnh của T bởi vậy đường gấp khúc $T_Q^* = T - S_Q^*$ là đường gấp khúc con giống nhất giữa T và Q. vì vậy, S_Q^* được coi như là đối số của cực tiểu toàn cục.

$$S_Q^* = \operatorname{argmin}\{s(Q, T - S_Q) : S_Q \subseteq T\} \quad (2.52)$$

Và độ tương tự từng phần tối ưu giữa T và Q được định nghĩa như sau:

$$\operatorname{Ops}(Q, T) = \min\{s(Q, T - S_Q) : S_Q \subseteq T\} \quad (2.53)$$

Chiều dài của cả hai đường Q và $T - S_Q$ đều được chuẩn hoá thành một trước khi $s(Q, T - S_Q)$ được tính toán.

Quan sát thấy rằng có sự khác nhau cơ bản giữa phương pháp này với những phương pháp biến dạng của Basri[9]. Thứ nhất là phương pháp này chỉ cho phép đơn giản hoá của một hình dạng nhất định, nghĩa là không cho phép biến dạng tùy ý. Thứ hai là không đo giá trị biến dạng của hình dạng mà thay vào đó là hình dạng tương tự sau khi biến dạng. Một điểm thú vị của định nghĩa này là luôn luôn đạt được một mức tối thiểu toàn bộ của hình dạng tương tự, nhưng sự tính toán có thể dẫn đến sự bùng nổ phép toán tổ hợp.

Do vậy, một thuật toán tối ưu được giới thiệu để tính toán ops. Đầu tiên một chuỗi các cạnh đệ quy được tạo ra:

$$T = T^n, T^{n-1}, \dots, T^2$$

Trong đó T^{k-1} đạt được bằng cách di chuyển một đỉnh đơn từ T^k :

$$T_Q^{k-1} = \operatorname{argmin}\{s(Q, T^k - \{x\}) : x \in T^k\} \quad (2.54)$$

Sau đó mức cực tiểu toàn bộ của sự tương tự giữa Q và T_Q^k được tính toán

$$Ps(Q,T) = \min\{s(Q,T^k_Q): k=2\dots n\} \text{ và } T^*_Q = \operatorname{argmin}\{s(Q,T^k_Q): k=2\dots n\} \quad (2.55)$$

Chiều dài của cả hai đường Q và T^k_Q đều được chuẩn hoá thành 1 trước khi $s(Q,T^*_Q)$ được tính toán.

Một thuộc tính quan trọng của những phần tương tự ops và ps được xác định là thực tế chúng bất biến đối với sự khác nhau giữa Q và T, do chuẩn hoá độ dài các đường cong được so sánh bởi s. Quan sát thấy rằng nếu T và Q ở những tỷ lệ khác nhau tạo cho chúng có độ dài bằng nhau thì không giải quyết được vấn đề của những tỷ lệ khác nhau. Lý do là trên thực tế Q chỉ giống một phần đường cong T^*_Q của T. Do đó, việc tạo lên Q và T^*_Q có độ dài như nhau để giải quyết vấn đề những tỷ lệ khác nhau. Đây là tất cả những gì xảy ra trong suốt quá trình tính toán ops và ps, bởi vì khi độ đo tương tự toàn cục s được sử dụng để so sánh truy vấn Q với đường cong gốc T đã đơn giản hoá cả với những đường cong đã theo tỷ lệ cùng chiều dài. Việc tổng kết sự đơn giản hoá của T kết hợp với việc xác định tỷ lệ cùng độ dài cung cấp một giải pháp cho vấn đề tỷ lệ khác nhau.

Quá trình di chuyển đỉnh có độ phức tạp là $O(n^2)$, với n là số đỉnh của T. Điều này không làm mất đi tính phức tạp của độ đo tương tự hình dạng toàn cục s, cái được sử dụng trong mỗi bước. Vì s có thể được tính toán với độ phức tạp là $O(n \cdot \log(n))$, nên tổng độ phức tạp của việc đối sánh hình dạng từng phần là $O(n^3 \log(n))$.

2.5. Đánh giá hiệu năng của hệ thống tra cứu ảnh

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 mượn từ hệ thống tra cứu thông tin truyền thống. [7]

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 thực sự thích hợp với truy vấn.

$$precision = \frac{|Q(q) \cap R(q)|}{|Q(q)|} \quad (2.56)$$

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

$$recall = \frac{|Q(q)IR(q)|}{|R(q)|} \quad (2.57)$$

Thông thường phải có sự thoả hiệp giữa hai số đo này bởi vì nếu muốn tăng cường số đo này thì lại phải chịu giảm số đo kia và ngược lại. Trong các hệ thống tra cứu điển hình thì độ thu hồi có xu hướng tăng lên khi số lượng các kết quả thu được tăng lên trong khi đó thì độ chính xác dường như lại bị giảm đi.

Ngoài ra, 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 truy vấn mà thôi.

Gần đây MPEG7 có khuyến 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 $\max\{N(q_1), N(q_2), \dots, N(q_Q)\}$ 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 (ở đây $K = \min\{4N(q), 2M\}$) 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 như sau:

$$AVR(q) = \sum_{k=1}^{N(q)} \frac{rank(k)}{N(q)} \quad (2.58)$$

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

$$MRR(q) = AVR(q) - 0.5 - 0.5 * N(q) \quad (2.59)$$

$MRR(q)$ 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 như sau:

$$NMRR(q) = \frac{MRR(q)}{K + 0.5 - 0.5 * N(q)} \quad (2.60)$$

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

$$ANMRR = \frac{1}{Q} \sum_{q=1}^Q NMRR(q) \quad (2.61)$$

Chương 3 : TRA CỨU ẢNH Y TẾ

Sự phát triển của những công nghệ hiện đại cùng với sự mở rộng của multimedia làm lượng ảnh lưu trữ ngày càng được gia tăng một cách nhanh chóng trong đó ảnh y tế cũng tăng lên đáng kể, đòi hỏi phải có các công cụ hỗ trợ tìm kiếm ảnh hiệu quả và tiện lợi. Lấy thông tin từ dữ liệu ảnh có giúp ích đến rất nhiều các lĩnh vực, từ những kho lưu trữ ảnh chụp, kho lưu trữ ảnh tội phạm, cơ sở dữ liệu ảnh về địa lý và quan trọng hơn là trong lĩnh vực y học. Vấn đề tra cứu ảnh y tế đã trở thành cần thiết trong việc sử dụng hình ảnh y tế trong y học lâm sàng nghiên cứu bệnh tật và y học giáo dục.

3.1. Tìm hiểu đặc trưng của ảnh y tế

Nhà vật lý người Đức Wilhelm Conrad Rontgen được ghi nhận là người phát hiện ra tia X bởi ông là người có hệ thống nghiên cứu chúng mặc dù không phải là người đầu tiên quan sát thấy tác động của chúng. Ông cũng là người đặt ra cái tên X-Quang ngoài ra cũng có nhiều người gọi đây là “tia Rontgen”.

Các nhà nghiên cứu đầu tiên quan trọng trong X-quang là Ivan Pulyui, William Crookes, Johann Wilhelm Hittorf, Eugen Goldstein, Heinrich Hertz, Philipp Lenard, Hermann von Helmholtz, Nikola Tesla, Thomas Edison, Charles Glover Barkla, Max von Laue, và Wilhelm Conrad Rontgen .

Năm 1895. Rontgen đã phát hiện ra X- Quang và là người đi tiên phong trong hình ảnh y tế. Lần đầu tiên trên thế giới người ta có thể hình dung được phần bên trong của cơ thể người mà không cần phải trải qua phẫu thuật. Ngay lập tức khám phá này đã được công bố rộng rãi trên báo trí. Công bố của ông đã đưa ra hình ảnh X- Quang bàn tay bà Rontgen.

Bức xạ tia X trong đó bao gồm X-Quang là một hình thức của bức xạ điện từ. Bước sóng điện từ của X-Quang trong khoảng 0.01- 10 nanomet, tương ứng với tần số trong khoảng 3×10^{16} Hz- 3×10^{19} Hz và năng lượng trong phạm vi 120 eV- 120keV. Đây là những bước sóng ngắn hơn so với tia UV và dài hơn tia gamma. Bức xạ tia X được gọi là bức xạ Rontgen.



Hình : hình ảnh X-Quang bàn tay bà Rontgen

X-Quang được phân thành 3 loại:

X-Quang từ 0.12 – 12 keV (10-0.10 nm bước sóng) được phân là loại mềm

X-Quang từ 12 – 120 keV (0.10-0.01nm bước sóng) được phân là loại cứng

X-Quang cứng có thể xuyên qua các vật thể rắn và được sử dụng phổ biến nhất trong lĩnh vực xác định hình thể bên trong của các đối tượng và tinh thể.

Vì vậy trong các ứng dụng chẩn đoán y tế, năng lượng X-Quang mềm sẽ bị cơ thể hấp thụ do đó một tấm kim loại mỏng được làm bằng nhôm sẽ được đặt trên cửa sổ của ống X-ray (là một phần của quang phổ điện từ) để lọc ra những năng lượng thấp trong quang phổ chỉ để chụp ảnh và thu được 1 hình ảnh phóng xạ.

X-Quang rất hữu ích trong việc phát triển bệnh lý của hệ thống xương cũng như để phát hiện một số quy trình bệnh trong mô mềm. Đáng chú ý nhất là vùng ngực, nó có thể xác định được các bệnh về phổi như viêm phổi, phù phổi, ung thư phổi, X-Quang bụng còn phát hiện được một số bệnh về đường ruột như tắc ruột, cổ trướng và nhiều chẩn đoán khác. Tuy nhiên X-Quang không có tác dụng chẩn đoán nhiều trong các mô mềm như não hay cơ.

Trong những năm gần đây máy vi tính và chụp X-Quang kỹ thuật số đã thay thế ảnh phim trong các ứng dụng y tế và nha khoa mặc dù ảnh phim vẫn còn sử dụng rộng rãi trong các quá trình chụp X-Quang công nghiệp (ví dụ: Để kiểm tra đường nối hàn của kết cấu).

Vào cuối thế kỷ XIX công nghệ hình ảnh được phát triển đáng kể. Nhiều kỹ thuật hình ảnh khác nhau đã được phát triển và sử dụng dựa trên các nguyên tắc vật lý khác nhau. Trong lĩnh vực y học những kỹ thuật hình ảnh khác nhau gọi là phương thức:

+ Phương thức giải phẫu học: Cung cấp cái nhìn sâu sắc về hình thái giải phẫu bao gồm X-Quang, siêu âm, chụp cắt lớp (CT), hình ảnh cộng hưởng từ (MRI) đôi khi còn xuất hiện dưới các tên khác nhau như: Chụp động mạch cộng hưởng từ (MRA từ MRI) chụp cắt lớp vi tính động mạch (CTA từ CT).

+ Phương thức chức năng: Miêu tả quá trình trao đổi chất cơ bản của các mô hoặc bộ phận cơ thể bao gồm các phương thức y học hạt nhân.

Ngày nay hình ảnh y tế là một phần quan trọng và thường xuyên được sử dụng trong y học X-Quang được sử dụng để xác định một số bệnh lý và các cấu trúc trong cơ thể. Bệnh lý của bệnh nhân được quan sát trực tiếp chứ không phải suy ra từ các triệu chứng lâm sàng. Ngoài ra hình ảnh y tế còn được sử dụng khi lập kế hoạch thực hiện phẫu thuật.

3.2 Tác dụng của tra cứu ảnh y tế

Tra cứu ảnh dựa vào nội dung hình ảnh y tế để hỗ trợ lâm sàng, góp phần quản lý dữ liệu lâm sàng và các phương pháp tiếp cận trên nội dung hình ảnh được lưu trữ.

Trước khi có sự xuất hiện của việc tra cứu ảnh y tế thì hình ảnh y tế được chú thích bởi văn bản hoặc các thông tin nền như thời gian, địa điểm, hoặc nhiếp ảnh gia, cho phép hình ảnh y tế được truy cập bằng cách tìm kiếm dựa trên văn bản. Thông qua mô tả văn bản, hình ảnh y tế có thể được quản lý dựa trên việc phân loại ảnh, phương thức. Tuy nhiên với sự xuất hiện của các cơ sở dữ liệu hình ảnh lớn, phương pháp tìm kiếm truyền thống dựa vào văn bản còn nhiều mặt hạn chế bởi:

+ Tiêu đề mô tả phụ thuộc vào ý tưởng chủ quan của người dùng khi thu thập cơ sở dữ liệu.

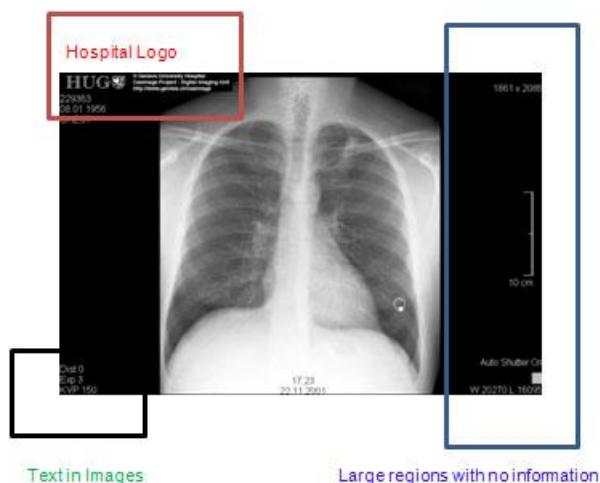
+ Một số tính năng hình ảnh, kết cấu và phân phối màu sắc khó thể mô tả một cách rõ ràng

+ Nhập chuỗi văn bản trong một cơ sở dữ liệu đòi hỏi nhiều thời gian và bất kỳ thay đổi trong quy tắc phân loại nào đều phải phân loại lại các hình ảnh

Để thực hiện chú thích bằng tay đòi hỏi nhiều thời gian. Trong khi số lượng của hình ảnh trong cơ sở dữ liệu tăng lên sẽ gây khó khăn trong việc tìm kiếm hình ảnh mong muốn và sẽ không khả thi để chú thích tất cả các thuộc tính của nội dung ảnh trên một khối lượng hình ảnh khổng lồ.

Thông thường một hình ảnh y tế thường có chứa một số đối tượng và thông tin. Để chú thích tất cả các kiến thức, hình dạng bất thường trong cơ thể, phạm vi bệnh lý khác nhau của bất kỳ hình ảnh gần như là không thể.

Ngoài ra tra cứu ảnh y tế cũng gặp một số khó khăn nhất định. Bởi hình ảnh y tế được thu từ nhiều thiết bị khác nhau. Mặt khác với đặc điểm của ảnh y tế có độ phân giải thấp và tiếng ồn mạnh là 2 đặc điểm phổ biến trong hầu hết các hình ảnh y tế.



Hình 2: Tiếng ồn xung quanh các đối tượng trong một hình ảnh y tế

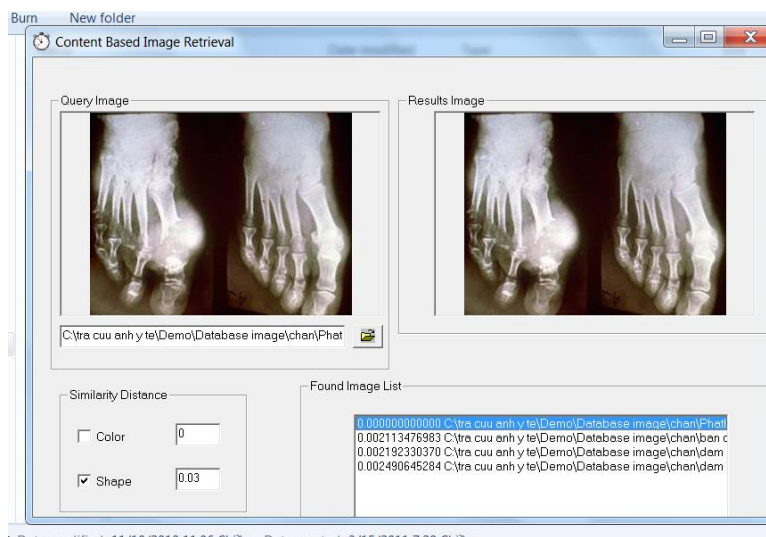
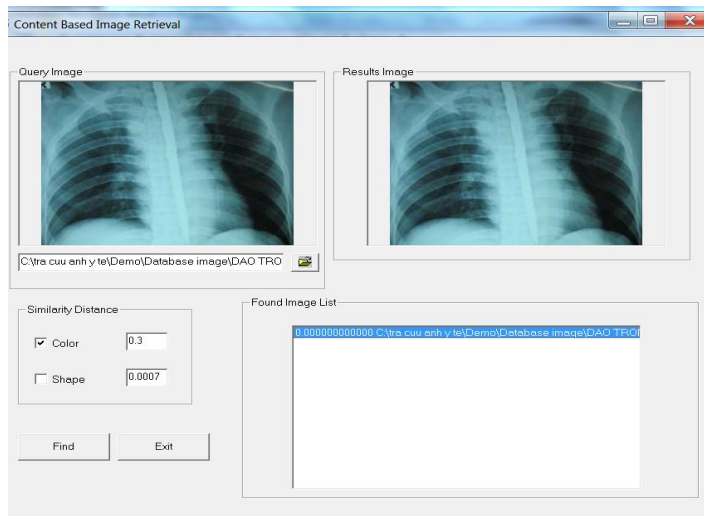
3.3. Chương trình mô phỏng

Khởi động chương trình, chọn file ảnh mẫu, ảnh mẫu sẽ được hiển thị ở khung "Query Image"

Chọn phương pháp so sánh: Bằng hình dạng hay bằng màu sắc, nhập số đo khoảng cách tối đa (giá trị ngưỡng) trong hộp Similarity Distance.

Nhấn nút Find, chương trình sẽ yêu cầu người dùng chọn thư mục chứa các file ảnh cần tìm. Chương trình sau đó sẽ liệt kê tất cả các ảnh có khoảng cách Euclide nhỏ hơn giá trị ngưỡng theo thứ tự tăng dần của khoảng cách Euclide, những ảnh được coi là giống hơn được xếp trước. Khi chọn một ảnh trong danh

sách ảnh tìm thấy (Found Image List), chương trình sẽ hiển thị ảnh đó trong khung Results Image.



Hình 3.4: Một số kết quả chạy thử chương trình

3.4. Khả năng mở rộng của chương trình

3.4.1. Những hạn chế của chương trình

Chương trình được giới thiệu mới chỉ đưa ra một thuật toán khá đơn giản để trích chọn đặc điểm (sinh véc tơ đặc trưng) nên chắc chắn sẽ không thể cho kết quả so sánh tốt.

Các ảnh ban đầu chưa được xử lý “sơ chế” như lọc nhiễu và yêu cầu ở một định dạng đơn giản nên phạm vi so sánh bị hạn chế.

Chương trình thiết lập kích thước cố định cho ảnh trước khi xử lý là 256×256 , kích thước các khối là 16×16 , số chiều của vector đặc trưng là 256 làm hạn chế tính mềm dẻo của chương trình.

3.4.2. Khả năng mở rộng

Kỹ thuật “tra cứu ảnh theo nội dung” có rất nhiều hướng nghiên cứu phát triển tạo ra những thuật toán rất hiệu quả làm cho máy tính “hiểu” được nội dung của ảnh. Chúng ta hoàn toàn có thể xây dựng một thuật toán tốt hơn để trích chọn được những đặc điểm đặc trưng khác như màu sắc, kết cấu, hình dạng của đối tượng ảnh để phát triển cho bài toán nhận dạng vật thể.

Thuật toán resample trong chương trình sử dụng phương pháp đơn giản nhất để thực hiện nội suy, nếu sử dụng các thuật toán cho độ chính xác cao hơn thì chắc chắn sẽ cải thiện đáng kể chất lượng chương trình.

Sử dụng các phương pháp làm giảm số chiều của véc tơ đặc trưng sẽ làm tăng đáng kể tốc độ tra cứu. Đối với các hệ thống máy tính có trang bị nhiều bộ vi xử lý hoặc có bộ xử lý lõi kép thì có thể xây dựng các thuật toán cho phép phân phối các tác vụ cho nhiều vi xử lý đồng thời cũng là một giải pháp tốt để tăng tốc độ tra cứu.

Sử dụng các phương pháp tính toán độ tương tự phù hợp hơn cho từng loại đặc điểm để có được những kết quả so sánh gần với trực giác hơn.

KẾT LUẬN

Đồ án đã trình bày một vài kỹ thuật nền tảng của các hệ thống tra cứu ảnh theo nội dung bao gồm mô tả nội dung trực quan, đánh giá độ tương tự và đánh giá hiệu năng hệ thống, trong đó nhấn mạnh vào các kỹ thuật mô tả các đặc điểm trực quan.

Các đặc điểm trực quan tổng quát được sử dụng nhiều nhất trong các hệ tra cứu ảnh theo nội dung là màu sắc, kết cấu, hình dạng.

Màu sắc thường được biểu diễn thông qua histogram màu, sơ đồ tương quan màu, véc tơ gắn kết màu và các moment màu trong một không gian màu nhất định.

Kết cấu có thể được biểu diễn thông qua đặc điểm Tamura, phân tích Wold, mô hình SAR, biến đổi Gabor và biến đổi dạng sóng.

Hình dạng có thể biểu diễn thông qua mô tả Fourier

Ngoài ra các đặc điểm trực quan của mỗi điểm ảnh lại có thể được sử dụng để phân tách mỗi ảnh thành các vùng đồng nhất hoặc các đối tượng ảnh. Các đặc điểm cục bộ của các vùng ảnh hoặc các đối tượng ảnh có thể dùng trong các hệ thống tra cứu ảnh theo vùng.

Có nhiều cách để đánh giá khoảng cách giữa các đặc điểm trực quan, một số cách được sử dụng phổ biến như khoảng cách Minkowski, khoảng cách toàn phương, khoảng cách Mahalanobis, độ phân kỳ Kullback-Leibler và độ phân kỳ Jeffrey. Đến thời điểm này thì phương pháp tính khoảng cách Minkowski và khoảng cách toàn phương được sử dụng rộng rãi nhất trong các hệ thống tra cứu ảnh.

Mặc dù kỹ thuật tra cứu ảnh theo nội dung đưa ra được một giải pháp thông minh và tự động để tìm kiếm ảnh một cách hiệu quả thì vấn đề chính của kỹ thuật này vẫn chỉ dựa trên những đặc điểm ở mức thấp. Nói chung mỗi đặc điểm mức thấp này chỉ có thể phản ánh được một khía cạnh nào đó của ảnh. Không có một đặc điểm nào có thể phản ánh được ngữ nghĩa của một bức ảnh, kể cả khi sử dụng kết hợp nhiều đặc điểm.

Ngoài ra sự đánh giá độ tương tự giữa các đặc điểm trực quan lại chưa liên quan đến đặc điểm sinh lý về thị giác của con người. Người sử dụng thường quan

tâm đến sự giống nhau về ngữ nghĩa nên kết quả truy vấn dựa trên các đặc điểm mức thấp thường không thoả mãn được yêu cầu và nói chung là khó đoán trước. Mặc dù phản hồi thích hợp là một cách để bù đắp vào những khoảng cách giữa tìm kiếm theo ngữ nghĩa và việc xử lý dữ liệu mức thấp nhưng vấn đề này vẫn tồn tại và cần phải có những kỹ thuật khác mới đáp ứng được.

Những vấn đề đã được giải quyết trong luận văn:

Giới thiệu chi tiết phương pháp tra cứu ảnh theo nội dung.

Sơ lược về cách đánh giá hiệu năng của một hệ thống tra cứu ảnh.

Xây dựng được chương trình thử nghiệm theo hai đặc điểm là: Hình dạng và màu sắc.

Những vấn đề còn tồn tại:

Do thời gian tìm hiểu về đề tài chưa được nhiều và những hạn chế về khả năng lập trình đồ hoạ nên một số mục tiêu đặt ra từ khi bắt tay nghiên cứu chưa thực hiện được trong chương trình chạy thử này, bao gồm:

Chưa xây dựng được chức năng tra cứu ảnh theo kết cấu (Texture).

Chưa có chức năng tra cứu ảnh bằng cách kết hợp nhiều đặc điểm.

Chưa cài đặt các kỹ thuật đánh chỉ số hiệu quả.

Chưa cài đặt các kỹ thuật tăng hiệu năng của hệ thống bằng cách giảm số chiều các véc tơ đặc trưng.

Trong thời gian tới, tôi hy vọng sẽ có thể giải quyết được những vấn đề còn tồn tại trên để có thể xây dựng được một chương trình thực sự hữu ích, đáp ứng được những yêu cầu của bài toán.

TÀI LIỆU THAM KHẢO

Tiếng Việt

1. Lương Mạnh Bá, Nguyễn Thanh Thủy (2003), *Nhập môn xử lý ảnh số*, Nhà xuất bản Khoa học Kỹ thuật.

2. Đỗ Năng Toàn, Phạm Việt Bình (2008). *Xử lý ảnh*. Nhà xuất bản Khoa học và kỹ thuật

Tiếng Anh

3. Cheng Chang, Liu Wenyin, Hongjiang Zhang (2001), *Image retrieval base on region shape similarity*, Microsoft Research China, www.eecs.berkeley.edu/~cchang/docs/Spie01.pdf.

4. Colin C. Venteres and Dr. Matthew Cooper, *A Review of Content-Based Image Retrieval Systems*, <http://www.jtap.ac.uk/reports/htm/jtap-054.html>

5. Dengsheng Zhang, Aylwin Wong, Maria Indrawan, Guojun Lu, *Content-based Image Retrieval Using Gabor Texture Features*, personal.gscit.monash.edu.au/~dengs/resource/papers/pcm00.pdf.

6. Eva M.van Rikxoort (2005), “*Content-based Image Retrieval Utilizing Color, Texture, and Shape*”, Master’s thesis in AI Radbound University Nijmegen The Netherlands.

7. Fuhui Long, Hongjiang Zhang, David Dagan Feng (2002), *Fundamentals of Content-based Image Retrieval*, in *Multimedia Information Retrieval and Management - Technological Fundamentals and Applications*.

8. Randy Crane (1997), *A simplified approach to Image Processing: classical and modern technique in C*, Prentice Hall, ISBN 0-13-226616-1.

9. J.Eakins, M.Graham (2004), “*Content-based Image Retrieval*”: A report to the JISC Technology Applications Programme, University of Northumbria at Newcastle.

10. Longin J. Latecki, R. Lamkaemper, D. Wolter (2005), “*Optimal Partial Shape Similarity*”, Dept. of Computer and Information Sciences Temple University.
11. Low (1991), *A introductory Computer Vision and Image Processin*, McGraw-hill, 244p. ISBN 0077074033
12. M. A. Stricker and M. Orengo (1995), “Similarity of color images. In *Proc. of the SPIE conference on the Storage and Retrieval for Image and Video Databases III*”, pages 381–392.
13. M. Partio, B. Cramariuc, M. Gabbouj, and A. Visa (2002), “*Rock Texture Retrieval using Gray Level Co-occurrence Matrix*”, NORSIG-2002, 5th Nordic Signal Processing Symposium, On Board Hurtigruten M/S Trollfjord, Norway.
14. Mike D Sutton (2005), *Using Device Independent Bitmaps (DIBs) in Visual Basic*, Part 1, EDAIS, <http://edais.mvps.org/>
15. Palaniraja Sivakumar (2004), “*Image Similarity Based on Color and Texture*”, Report for CIS751 MS Project.
16. Sebe N, Lew (2001), “*Texture Features for Content-based Retrieval*”: Principles of visual Information Retrieval.
17. Shengjiu Wang (2001), “*A Robust CBIR Approach Using Local Color Histogram*”, Technique Repost TR 01-13, Edmonton, Alberta, Canada.
18. Tanner Helland, *Graphics Programming in Visual Basic*, tannerhelland.tripod.com/VBGraphicsTutorial.htm
19. Vishal Chitkara (2001). “*Color-based image retrieval using compact binary signatures*”, Master’s thesis, University of Alberta.
20. Vittorio Castelli and Lawrence D. Bergman (2002) “*Image Database Search and Retrieval of Digital Imagery*”, John Wiley & Sons, Inc., New York.