

MỤC LỤC

Lời cảm ơn
Nhiệm vụ của đề tài
Giới thiệu cơ quan thực tập
Mục lục

NỘI DUNG BÁO CÁO

Chương 1: TỔNG QUAN VỀ XỬ LÝ ẢNH VÀ TRA CỨU ẢNH

1.1 Tổng quan về xử lý ảnh

1.1.1 Một số khái niệm

1.1.1.1 Pixel

1.1.1.2 Gray level

1.1.1.3 Định dạng ảnh

1.1.1.3.1 GIF

1.1.1.3.2 PNG

1.1.1.3.3 BMP

1.1.2 Biểu diễn ảnh

1.1.3 Tăng cường ảnh – khôi phục ảnh

1.1.4 Biến đổi ảnh

1.1.5 Phân tích ảnh

1.1.6 Nhận dạng ảnh

1.1.7 Nén ảnh

1.2 Tổng quan về tra cứu ảnh dựa trên nội dung

1.2.1 Những thành phần của một hệ thống tra cứu ảnh dựa trên nội dung

1.2.1.1 Công nghệ tự động trích chọn siêu dữ liệu

1.2.1.2 Giao diện để lấy yêu cầu truy vấn người sử dụng

1.2.1.3 Phương pháp so sánh độ tương tự giữa các ảnh

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

1.2.2 Những ứng dụng cơ bản của tra cứu ảnh

1.2.3 Những chức năng của hệ thống tra cứu ảnh dựa trên nội dung

1.2.4 Các phương pháp tra cứu ảnh dựa trên nội dung

1.2.4.1 Tra cứu ảnh dựa trên màu sắc

1.2.4.1 Tra cứu ảnh dựa trên kết cấu

1.2.4.1 Tra cứu ảnh dựa trên hình dạng

1.2.5 Những hệ thống tra cứu ảnh dựa trên nội dung

1.2.5.1 Hệ thống QBIC

- 1.2.5.2 *Hệ thống Photobook*
- 1.2.5.3 *Hệ thống Visual SEEK và WebSEEK*
- 1.2.5.4 *Hệ thống RetrievalWare*
- 1.2.5.5 *Hệ thống Imatch*
- 1.2.6 **Kết luận**

Chương 2: TÌM HIỂU CÁC PHƯƠNG PHÁP TRÍCH CHỌN ĐẶC TRƯNG HÌNH ẢNH

2.1 Màu sắc

2.1.1 *Không gian màu*

2.1.2 *Lược đồ màu*

2.2 Kết cấu

2.2.1 *Các đặc trưng Tamura*

2.2.1.1 *Thô*

2.2.1.2 *Độ tương phản*

2.2.1.3 *Hướng*

2.2.2 *Các đặc trưng Wold*

2.2.3 *Mô hình tự hồi quy đồng thời SAR*

2.2.4 *Các đặc trưng Gabor*

2.2.5 *Các đặc trưng biến đổi sóng*

2.3 Hình dạng

2.3.1 *Các bất biến mômen*

2.3.1 *Các góc quay*

2.3.1 *Các ký hiệu mô tả Fourier*

2.3.1 *Hình tròn, Độ lệch tâm và Hướng trục chính*

2.4 Thông tin không gian

2.5 Phân đoạn ảnh

Chương 3: CÁC ĐỘ ĐO TƯƠNG TỰ

3.1 Lược đồ giao

3.2 Khoảng cách Minkowski

3.2 Khoảng cách toàn phương

3.2 Khoảng cách EMD

Chương 4: CÀI ĐẶT THỬ NGHIỆM CHƯƠNG TRÌNH, KẾT LUẬN, TÀI LIỆU THAM KHẢO

4.1 Kết luận

4.2 Tài liệu tham khảo

Chương 1: TỔNG QUAN VỀ XỬ LÝ ẢNH VÀ TRA CỨU ẢNH

1.1 TỔNG QUAN VỀ XỬ LÝ ẢNH

Xử lý ảnh là một trong những mảng quan trọng nhất trong kỹ thuật thị giác máy tính, là tiền đề cho nhiều nghiên cứu thuộc lĩnh vực này. Hai nhiệm vụ cơ bản của quá trình xử lý ảnh là nâng cao chất lượng thông tin hình ảnh và xử lý số liệu cung cấp cho các quá trình khác trong đó có việc ứng dụng thị giác vào điều khiển.

Quá trình bắt đầu từ việc thu nhận ảnh nguồn (từ các thiết bị thu nhận ảnh dạng số hoặc tương tự) gửi đến máy tính. Dữ liệu ảnh được lưu trữ ở định dạng phù hợp với quá trình xử lý. Người lập trình sẽ tác động các thuật toán tương ứng lên dữ liệu ảnh nhằm thay đổi cấu trúc ảnh phù hợp với các ứng dụng khác nhau.

1.1.1 Một số khái niệm

1.1.1.1 Pixel (Picture Element): phân tử ảnh

Ảnh trong thực tế là một ảnh liên tục về không gian và về giá trị độ sáng. Để có thể xử lý ảnh bằng máy tính cần thiết phải tiến hành số hoá ảnh. Trong quá trình số hoá, người ta biến đổi tín hiệu liên tục sang tín hiệu rời rạc thông qua quá trình lấy mẫu (rời rạc hóa về không gian) và lượng hoá thành phân giá trị mà thể về nguyên tắc bằng mắt thường không phân biệt được hai điểm kề nhau. Trong quá trình này, người ta sử dụng khái niệm *Picture element* mà ta quen gọi hay viết là *Pixel* - phân tử ảnh. Ở đây cũng cần phân biệt khái niệm pixel hay đề cập đến trong các hệ thống đồ họa máy tính. Để tránh nhầm lẫn ta tạm gọi khái niệm pixel này là pixel thiết bị. Khái niệm pixel thiết bị có thể xem xét như sau: khi ta quan sát màn hình (trong chế độ đồ họa), màn hình không liên tục mà gồm nhiều điểm nhỏ, gọi là pixel. Mỗi pixel gồm một cặp tọa độ x, y và màu.

Cặp tọa độ x, y tạo nên *độ phân giải* (resolution). Như màn hình máy tính có nhiều loại với độ phân giải khác nhau: màn hình CGA có độ phân giải là 320 x 200; màn hình VGA là 640 x 350,...

Như vậy, một ảnh là một tập hợp các điểm ảnh. Khi được số hoá, nó thường được biểu diễn bởi bảng hai chiều $I(n,p)$: n dòng và p cột. Ta nói ảnh gồm n x p pixels. Người ta thường ký hiệu $I(x,y)$ để chỉ một pixel. Thường giá trị của n chọn

bằng p và bằng 256. Hình 1.2 cho ta thấy việc biểu diễn một ảnh với độ phân giải khác nhau. Một pixel có thể l- u trữ trên 1, 4, 8 hay 24 bit.

1.1.1.2 Gray level: Mức xám

Mức xám là kết quả sự mã hoá t- ong ứng một c- ờng độ sáng của mỗi điểm ảnh với một giá trị số - kết quả của quá trình l- ợng hoá. Cách mã hoá kinh điển th- ờng dùng 16, 32 hay 64 mức. Mã hoá 256 mức là phổ dụng nhất do lý do kỹ thuật. Vì $2^8 = 256$ (0, 1, ..., 255), nên với 256 mức, mỗi pixel sẽ đ- ợc mã hoá bởi 8 bit.

1.1.1.3 Định dạng ảnh

1.1.1.3.1 GIF (Graphics Interchange Format) là một định dạng tập tin hình ảnh bitmap cho các hình ảnh dùng ít hơn 256 màu sắc khác nhau và các hoạt hình dùng ít hơn 256 màu cho mỗi khung hình. GIF là định dạng nén dữ liệu đặc biệt hữu ích cho việc truyền hình ảnh qua đường truyền lưu lượng nhỏ. Định dạng này được CompuServe cho ra đời vào năm 1987 và nhanh chóng được dùng rộng rãi trên Word Wide Web cho đến nay.

Tập tin GIF dùng nén dữ liệu bảo toàn trong đó kích thước tập tin có thể được giảm mà không làm giảm chất lượng hình ảnh, cho những hình ảnh có ít hơn 256 màu. Số lượng tối đa 256 màu làm cho định dạng này không phù hợp cho các hình chụp (thường có nhiều màu sắc), tuy nhiên các kiểu nén dữ liệu bảo toàn cho hình chụp nhiều màu cũng có kích thước quá lớn đối với truyền dữ liệu trên mạng hiện nay. Định dạng JPEG là nén dữ liệu thất thoát có thể được dùng cho các ảnh chụp, nhưng lại làm giảm chất lượng cho các bức vẽ ít màu, tạo nên những chỗ nhòe thay cho các đường sắc nét, đồng thời độ nén cũng thấp cho các hình vẽ ít màu. Như vậy, GIF thường được dùng cho sơ đồ, hình vẽ nút bấm và các hình ít màu, còn JPEG được dùng cho ảnh chụp.

Định dạng GIF đã được đăng ký sở hữu trí tuệ bởi Unisys, và những ai muốn viết chương trình để tạo ra hoặc hiển thị tập tin GIF phải trả tiền bản quyền. Tiêu chuẩn định dạng PNG đã ra đời để thay thế GIF, giảm các hạn chế luật pháp và hạn chế công nghệ. Nay giấy phép sở hữu trí tuệ của Unisys đã hết hạn, nhưng PNG vẫn được ưa chuộng do có nhiều tính năng kỹ thuật vượt trội, và đã trở thành định dạng phổ biến thứ 3 trên mạng.

1.1.1.3.2 PNG (Portable Network Graphics) là một dạng hình ảnh sử dụng phương pháp nén dữ liệu mới - không làm mất đi dữ liệu gốc. PNG được tạo ra nhằm cải thiện và thay thế định dạng ảnh GIF với một định dạng hình ảnh không đòi hỏi phải có giấy phép sáng chế khi sử dụng. PNG được hỗ trợ bởi thư viện tham chiếu libpng, một thư viện nền tảng độc lập bao gồm các hàm của C để quản lý các hình ảnh PNG.

Những tập tin PNG thường có phần mở rộng là PNG and png và đã được gán kiểu chuẩn MIME là image/png (được công nhận vào ngày 14 tháng 10 năm 1996).

Phần đầu của tập tin

Một tập tin PNG bao gồm 8-byte kí hiệu (89 50 4E 47 0D 0A 1A 0A) được viết trong hệ thống có cơ số 16, chứa các chữ "PNG" và 2 dấu xuống dòng, ở giữa là sắp xếp theo số lượng của các *thành phần*, mỗi thành phần đều chứa thông tin về hình ảnh. Cấu trúc dựa trên các thành phần được thiết kế cho phép định dạng PNG có thể tương thích với các phiên bản cũ khi sử dụng.

Các "thành phần" trong tập tin

PNG là cấu trúc như một chuỗi các thành phần, mỗi thành phần chứa kích thước, kiểu, dữ liệu, và mã sửa lỗi CRC ngay trong nó.

Chuỗi được gán tên bằng 4 chữ cái phân biệt chữ hoa chữ thường. Sự phân biệt này giúp bộ giải mã phát hiện bản chất của chuỗi khi nó không nhận dạng được.

Với chữ cái đầu, viết hoa thể hiện chuỗi này là thiết yếu, nếu không thì ít cần thiết hơn ancillary. Chuỗi thiết yếu chứa thông tin cần thiết để đọc được tệp và nếu bộ giải mã không nhận dạng được chuỗi thiết yếu, việc đọc tệp phải được hủy.

Thành phần cơ bản

Một bộ giải mã (decoder) phải có thể thông dịch để đọc và hiển thị một tệp PNG.

- IHDR phải là thành phần đầu tiên, nó chứa đựng header
- PLTE chứa đựng bảng màu (danh sách các màu)
- IDAT chứa đựng ảnh. Ảnh này có thể được chia nhỏ chứa trong nhiều phần IDAT. Điều này làm tăng kích cỡ của tệp lên một ít nhưng nó làm cho việc phát sinh ảnh PNG mượt hơn (streaming manner).

- IEND đánh dấu điểm kết thúc của ảnh.

Ảnh động

PNG không hỗ trợ ảnh động. Nhưng một định dạng khác phức tạp hơn dựa trên ý tưởng và các chunk của PNG là MNG được thiết kế cho ảnh động, tuy nhiên định dạng này không cho phép 'tương thích lùi' tức là hiển thị một ảnh trong trường hợp hệ thống không hỗ trợ được hình động. Một định dạng khác là APNG cũng dựa trên PNG hỗ trợ ảnh động và tương thích lùi, nhưng đơn giản hơn MNG. Tuy nhiên, đến thời điểm 2005 những định dạng này vẫn chưa được hỗ trợ rộng rãi.

1.1.1.3.3 BMP Trong đồ họa máy vi tính, BMP, còn được biết đến với tên tiếng Anh khác là *Windows bitmap*, là một định dạng tập tin hình ảnh khá phổ biến. Các tập tin đồ họa lưu dưới dạng BMP thường có đuôi là .BMP hoặc .DIB (*Device Independent Bitmap*).

Các thuộc tính tiêu biểu của một tập tin ảnh BMP (cũng như file ảnh nói chung) là

- số bit trên mỗi điểm ảnh (*bit per pixel*), thường được ký hiệu bởi **n**. Một ảnh BMP n-bit có 2^n màu. Giá trị n càng lớn thì ảnh càng có nhiều màu, và càng rõ nét hơn. Giá trị tiêu biểu của n là 1 (ảnh đen trắng), 4 (ảnh 16 màu), 8 (ảnh 256 màu), 16 (ảnh 65536 màu) và 24 (ảnh 16 triệu màu). Ảnh BMP 24-bit có chất lượng hình ảnh trung thực nhất.
- chiều cao của ảnh (*height*), cho bởi điểm ảnh (*pixel*).
- chiều rộng của ảnh (*width*), cho bởi điểm ảnh.

Cấu trúc tập tin ảnh BMP bao gồm 4 phần

- Bitmap Header (14 bytes): giúp nhận dạng tập tin bitmap.
- Bitmap Information (40 bytes): lưu một số thông tin chi tiết giúp hiển thị ảnh.
- Color Palette (4*x bytes), x là số màu của ảnh: định nghĩa các màu sẽ được sử dụng trong ảnh.
- Bitmap Data: lưu dữ liệu ảnh.

Đặc điểm nổi bật nhất của định dạng BMP là tập tin hình ảnh thường không được nén bằng bất kỳ thuật toán nào. Khi lưu ảnh, các điểm ảnh được ghi trực tiếp vào tập tin - một điểm ảnh sẽ được mô tả bởi một hay nhiều byte tùy thuộc vào giá

trị n của ảnh. Do đó, một hình ảnh lưu dưới dạng BMP thường có kích cỡ rất lớn, gấp nhiều lần so với các ảnh được nén (chẳng hạn GIF, JPEG hay PNG).

Định dạng BMP được hỗ trợ bởi hầu hết các phần mềm đồ họa chạy trên Windows, và cả một số ứng dụng chạy trên MS-DOS. Ngay từ Windows 3.1, Microsoft đã cho ra đời phần mềm PaintBrush, một phần mềm hỗ trợ vẽ hình ảnh đơn giản và lưu hình ảnh được vẽ dưới dạng BMP 16 hay 256 màu. Tuy nhiên, do kích thước tập tin ảnh BMP quá lớn, định dạng BMP không phù hợp để trao đổi hình ảnh qua mạng Internet (do hạn chế về tốc độ truyền dữ liệu). Do đó, các trang web thường sử dụng ảnh dạng GIF, JPEG hay PNG. Các định dạng này hỗ trợ các thuật toán nén hình ảnh, vì vậy có thể giảm bớt kích cỡ của ảnh.

1.1.2 Biểu diễn ảnh

Trong biểu diễn ảnh, người ta thường dùng các phần tử đặc trưng của ảnh là pixel. Nhìn chung có thể xem một hàm hai biến chứa các thông tin về biểu diễn của một ảnh. Các mô hình biểu diễn ảnh cho ta một mô tả logic hay định lượng các tính chất của hàm này. Trong biểu diễn ảnh cần chú ý đến tính trung thực của ảnh hoặc các tiêu chuẩn “thông minh” để đo chất lượng ảnh hoặc tính hiệu quả của các kỹ thuật xử lý.

Việc xử lý ảnh số yêu cầu ảnh phải được mẫu hoá và lượng tử hoá. Thí dụ một ảnh ma trận 512 dòng gồm khoảng 512×512 pixel. Việc lượng tử hoá ảnh là chuyển đổi tín hiệu tương tự sang tín hiệu số (Analog Digital Convert) của một ảnh đã lấy mẫu sang một số hữu hạn mức xám. Vấn đề này sẽ trình bày chi tiết trong chương 2.

Một số mô hình thường được dùng trong biểu diễn ảnh: Mô hình toán, mô hình thống kê. Trong mô hình toán, ảnh hai chiều được biểu diễn nhờ các hàm hai biến thực giao gọi là các *hàm cơ sở*. Các biến đổi này sẽ trình bày kỹ trong chương 3. Với mô hình thống kê, một ảnh được coi như một phần tử của một tập hợp đặc trưng bởi các đại lượng như: kỳ vọng toán học, hiệp biến, phương sai, moment.

1.1.3 Tăng cường ảnh - khôi phục ảnh

Tăng cường ảnh là bước quan trọng, tạo tiền đề cho xử lý ảnh. Nó gồm một loạt các kỹ thuật như: lọc độ tương phản, khử nhiễu, nổi màu, v...v.

Khôi phục ảnh là nhằm loại bỏ các suy giảm (degradation) trong ảnh. Với một hệ thống tuyến tính, ảnh của một đối tượng có thể biểu diễn bởi:

$$g(x,y) = \int_{-\infty-\infty}^{+\infty+\infty} h(x,y;\alpha,\beta) f(\alpha,\beta) d\alpha d(\beta + \eta(x,y))$$

Trong đó:

- $\eta(x,y)$ là hàm biểu diễn nhiễu cộng.
- $f(\alpha,\beta)$ là hàm biểu diễn đối tượng.
- $g(x,y)$ là ảnh thu nhận.
- $h(x,y; \alpha,\beta)$ là hàm tán xạ điểm (Point Spread Function - PSF).

Một vấn đề khôi phục ảnh tiêu biểu là tìm một xấp xỉ của $f(\alpha,\beta)$ khi PSF của nó có thể đo lường hay quan sát được, ảnh mờ và các tính chất đặc xuất của quá trình nhiễu.

1.1.4 Biến đổi ảnh

Thuật ngữ biến đổi ảnh (Image Transform) thường dùng để nói tới một lớp các ma trận đơn vị và các kỹ thuật dùng để biến đổi ảnh. Cũng như các tín hiệu một chiều được biểu diễn bởi một chuỗi các hàm cơ sở, ảnh cũng có thể được biểu diễn bởi một chuỗi rời rạc các ma trận cơ sở gọi là *ảnh cơ sở*.

Phương trình ảnh cơ sở có dạng:

$A^*_{k,l} = a_k a_l^{*T}$, với a_k là cột thứ k của ma trận A . A là ma trận đơn vị. Có nghĩa là $A A^{*T} = I$. Các $A^*_{k,l}$ định nghĩa ở trên với $k,l = 0,1, \dots, N-1$ là ảnh cơ sở. Có nhiều loại biến đổi được dùng như :

- Biến đổi Fourier, Sin, Cosin, Hadamard,...
- Tích Kronecker (*)
- Biến đổi KL (Karhunen Loeve): biến đổi này có nguồn gốc từ khai triển của các quá trình ngẫu nhiên gọi là phương pháp trích chọn các thành phần chính.

Do phải xử lý nhiều thông tin, các phép toán nhân và cộng trong khai triển là khá lớn. Do vậy, các biến đổi trên nhằm làm giảm thứ nguyên của ảnh để việc xử lý ảnh được hiệu quả hơn

1.1.5 Phân tích ảnh

Phân tích ảnh liên quan đến việc xác định các độ đo định lượng của một ảnh để đưa ra một mô tả đầy đủ về ảnh. Các kỹ thuật được sử dụng ở đây nhằm mục đích xác định biên của ảnh. Có nhiều kỹ thuật khác nhau như lọc vi phân hay dò theo quy hoạch động.

Ngời ta cũng dùng các kỹ thuật để phân vùng ảnh. Từ ảnh thu được, người ta tiến hành kỹ thuật tách (split) hay hợp (fusion) dựa theo các tiêu chuẩn đánh giá như: màu sắc, cường độ, v...v. Các phương pháp được biết đến như Quad-Tree, mảnh hoá biên, nhị phân hoá đường biên. Cuối cùng, phải kể đến các kỹ thuật phân lớp dựa theo cấu trúc.

1.1.6 Nhận dạng ảnh

Nhận dạng ảnh là quá trình liên quan đến các mô tả đối tượng mà người ta muốn đặc tả nó. Quá trình nhận dạng thường đi sau quá trình trích chọn các đặc tính chủ yếu của đối tượng. Có hai kiểu mô tả đối tượng:

- Mô tả tham số (nhận dạng theo tham số).
- Mô tả theo cấu trúc (nhận dạng theo cấu trúc).

Trên thực tế, người ta đã áp dụng kỹ thuật nhận dạng khá thành công với nhiều đối tượng khác nhau như: nhận dạng ảnh vân tay, nhận dạng chữ (chữ cái, chữ số, chữ có dấu).

Nhận dạng chữ in hoặc đánh máy phục vụ cho việc tự động hoá quá trình đọc tài liệu, tăng nhanh tốc độ và chất lượng thu nhận thông tin từ máy tính.

Nhận dạng chữ viết tay (với mức độ ràng buộc khác nhau về cách viết, kiểu chữ, v...,v) phục vụ cho nhiều lĩnh vực.

Ngoài 2 kỹ thuật nhận dạng trên, hiện nay một kỹ thuật nhận dạng mới dựa vào kỹ thuật mạng nơ ron đang được áp dụng và cho kết quả khả quan.

1.1.7 Nén ảnh

Dữ liệu ảnh cũng như các dữ liệu khác cần phải lưu trữ hay truyền đi trên mạng. Như đã nói ở trên, lượng thông tin để biểu diễn cho một ảnh là rất lớn. Trong phần 1.1 chúng ta đã thấy một ảnh đen trắng cỡ 512 x 512 với 256 mức xám chiếm 256K bytes. Do đó làm giảm lượng thông tin hay nén dữ liệu là một nhu cầu cần thiết. Nhiều phương pháp nén dữ liệu đã được nghiên cứu và áp dụng cho loại dữ liệu đặc biệt này.

1.2 TỔNG QUAN VỀ TRA CỨU ẢNH DỰA TRÊN NỘI DUNG

Tra cứu ảnh là quá trình tìm kiếm trong một cơ sở dữ liệu ảnh những ảnh thoả mãn một yêu cầu nào đó. Tra cứu ảnh được sử dụng trong nhiều lĩnh vực khác nhau: y tế, khoa học hình sự, bảo tồn, ngân hàng... Vấn đề tra cứu ảnh cũng nhận được sự quan tâm của nhiều nhà nghiên cứu.

Tra cứu ảnh theo nội dung chính thức xuất hiện từ năm 1992, đánh dấu bằng Hội thảo về các hệ thống quản lý thông tin trực quan của Quỹ Khoa học Quốc gia của Hoa Kỳ. Một số hệ tra cứu ảnh theo nội dung tiêu biểu: QBIC, VIR Image Engine, VisualSEEK, NeTra, MARS, Viper

Tra cứu ảnh theo nội dung dựa vào các đặc điểm *nội dung trực quan* của chính bức ảnh để tra cứu: màu sắc, kết cấu, hình dạng và bố cục không gian. Đây là các đặc điểm mức thấp, chưa phản ánh được ngữ nghĩa của ảnh.

1.2.1 Những thành phần của một hệ thống tra cứu ảnh dựa trên nội dung.

1.2.1.1 Công nghệ tự động trích chọn siêu dữ liệu.

Mỗi đặc điểm nguyên thủy của ảnh có định dạng đặc trưng của nó như biểu đồ màu được sử dụng rộng rãi để biểu thị đặc điểm màu sắc, đặc điểm hình dạng có thể biểu thị bằng một tập các đoạn biên liền nhau. Với siêu dữ liệu thích hợp, hệ thống có thể tìm kiếm ảnh dựa trên màu sắc.

1.2.1.2. Giao diện để lấy yêu cầu truy vấn người sử dụng.

Trong bất kỳ một hệ thống tìm kiếm nào thì quá trình tìm kiếm đều bắt đầu từ một yêu cầu tìm kiếm. Vì vậy, nó là vấn đề cốt yếu để lấy yêu cầu truy vấn của người sử dụng một cách chính xác và dễ dàng. Tìm kiếm dựa trên text đã được sử dụng rộng rãi trong các hệ thống tìm kiếm.

Ví dụ: Tìm một quyển sách mà mình mong muốn với từ khóa nào đó trong thư viện. Với hệ thống tìm kiếm ảnh dựa trên nội dung thì quá trình tìm kiếm được thực hiện thông qua một hình ảnh mẫu được cung cấp bởi người sử dụng gọi là truy vấn bởi mẫu. Mặc dù vậy, người sử dụng không thể luôn luôn đưa ra một ảnh mẫu cho hệ thống tìm kiếm. Hệ thống tìm kiếm ảnh dựa trên màu sắc đưa ra một giao diện để chỉ định hoặc chọn một số đặc điểm cơ bản cho việc cung cấp ảnh mẫu như

sử dụng hệ thống QBIC của IBM người sử dụng có thể chỉ định truy vấn đặc điểm màu sắc bằng cách chọn ra số lượng thành phần RED, BLUE, GREEN liên quan hoặc là có thể lựa chọn màu sắc ảnh mong muốn từ bảng màu.

1.2.1.3 Phương pháp so sánh độ tương tự giữa các ảnh.

Hệ thống tìm kiếm ảnh dựa trên màu sắc yêu cầu các phương pháp dựa trên những đặc điểm nguyên thủy để so sánh độ tương tự giữa ảnh mẫu và tất cả những hình ảnh trong tập ảnh. Mặc dù vậy, sự tương tự hoặc sự khác nhau giữa các ảnh không chỉ xác định theo một cách duy nhất. Số lượng của ảnh tương tự sẽ thay đổi khi yêu cầu truy vấn thay đổi. Chẳng hạn, trong trường hợp hai bức tranh, một là “biển xanh với mặt trời mọc” và trường hợp khác là “núi xanh với mặt trời mọc”. Khi “mặt trời” được xem xét thì độ tương tự giữa hai ảnh này là cao nhưng nếu đối tượng quan tâm là “biển xanh” thì độ tương tự giữa hai ảnh này là thấp. Như vậy rất khó khăn để tìm ra phương pháp đo độ tương tự giữa hai hình ảnh một cách chính xác đối với tất cả các kiểu yêu cầu của truy vấn. Hay nói cách khác, mỗi một phương pháp tìm kiếm sẽ có giới hạn của chính nó. Ví dụ, rất khó cho công nghệ tìm kiếm dựa trên màu sắc để tìm ra điểm khác nhau giữa một ảnh là bầu trời màu xanh với một ảnh là mặt biển xanh.

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

Đối với một tập dữ liệu ảnh lớn thì không gian lưu trữ cho siêu dữ liệu là rất cần thiết. Một hệ thống tìm kiếm ảnh dựa trên nội dung phải có những công nghệ hiệu quả để quản lý siêu dữ liệu, đồng thời phải có chuẩn để mô tả nó.

Khi một truy vấn được xử lý trên một cơ sở dữ liệu lớn, việc so sánh độ tương tự giữa ảnh truy vấn và tất cả các hình ảnh từng cặp là không thể thực hiện được bởi người dùng chỉ cần những ảnh có độ tương tự cao so với ảnh mẫu. Những chỉ số cấu trúc có thể giúp tránh được việc tìm kiếm tuần tự và cải thiện tìm kiếm một cách hiệu quả nên được sử dụng trong hệ thống tìm kiếm ảnh dựa trên màu sắc. Hơn nữa, với những cơ sở dữ liệu ảnh thường xuyên thay đổi thì chỉ số cấu trúc động là rất cần thiết. Khi nội dung của ảnh được thể hiện bởi các vector ít chiều và khoảng cách giữa các ảnh được định nghĩa (như khoảng không gian được tính toán

bằng khoảng cách Euclidean) cây R và các thành phần của nó có thể được sử dụng để đánh chỉ số cho ảnh.

Khi khoảng cách không được định nghĩa như không gian vector hoặc khi không gian vector là nhiều chiều hoặc khi mà những gì chúng ta có chỉ là một hàm khoảng cách tức là khoảng cách metric thì những phương pháp để đánh chỉ số ảnh dựa trên hàm khoảng cách trong không gian metric là thích hợp.

1.2.2 Những ứng dụng cơ bản của tra cứu ảnh.

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

- Ngăn chặn tội phạm.
- Quân sự.
- Quản lý tài sản trí tuệ.
- Thiết kế kiến trúc máy móc.
- Thiết kế thời trang và nội thất.
- Báo chí quảng cáo.
- Chuẩn đoán y học.
- Hệ thống thông tin địa lý.
- Di sản văn hóa.
- Giáo dục và đào tạo.
- Giải trí.
- Tìm kiếm trang web.

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

Một hệ thống tra cứu ảnh dựa trên nội dung tiêu biểu không chỉ liên quan tới các nguồn thông tin trong những dạng khác nhau (ví dụ như văn bản, ảnh, video) mà còn liên quan đến nhu cầu của người sử dụng. Về cơ bản nó phân tích cả nội dung của nguồn thông tin cũng như truy vấn của người sử dụng và sau đó đối sánh chúng để tìm ra những tiêu chí có liên quan này. Những chức năng chính của một hệ thống tra cứu ảnh bao gồm:

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

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

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

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

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

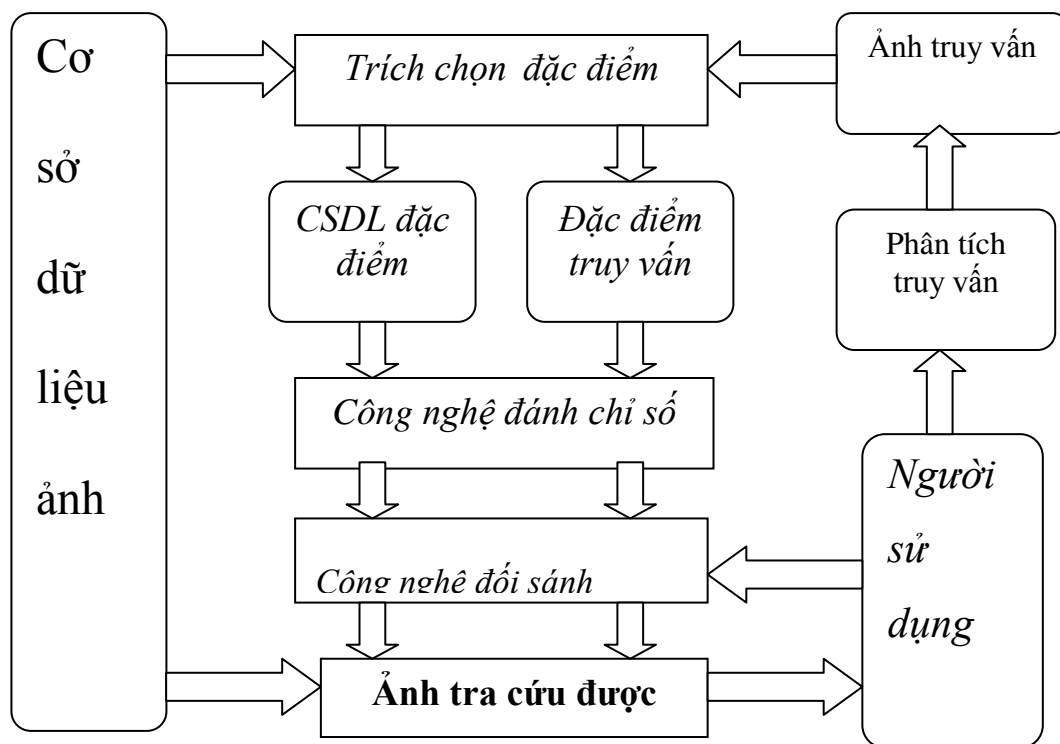
Yêu cầu của người sử dụng: Có rất nhiều cách có thể đưa truy vấn trực quan. Một phương pháp truy vấn tốt là phương pháp tự nhiên với người sử dụng tức là cung cấp đầy đủ thông tin từ người sử dụng để trích chọn những kết quả có ý nghĩa. Những phương pháp dưới đây thường được sử dụng trong kỹ thuật tra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

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

Truy vấn bởi đặc điểm (QBF- Query By Feature): Trong hệ thống kiểu này người dùng chỉ định câu hỏi bởi những đặc điểm chỉ định rõ ràng đó là những đặc điểm được quan tâm trong tìm kiếm. Ví dụ người dùng có thể truy vấn cơ sở dữ liệu ảnh bởi việc đưa ra một câu lệnh “Đưa ra tất cả những ảnh có góc bên trên trái chứa 25% điểm màu vàng”. Truy vấn này được người dùng chỉ định bởi việc sử dụng công cụ giao diện đồ họa đặc biệt. Những người sử dụng chuyên nghiệp thì có thể tìm kiếm kiểu truy vấn tự nhiên này nhưng những người không chuyên thì rất khó. QBIC là một ví dụ về hệ thống tra cứu ảnh dựa trên nội dung mà người sử dụng truy vấn kiểu này.

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

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



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

1.2.4 Các phương pháp tra cứu ảnh dựa trên nội dung.

1.2.4.1. Tra cứu ảnh dựa trên màu sắc.

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.

1.2.4.2. Tra cứu ảnh dựa trên kết cấu.

Khả năng tra cứu ảnh dựa trên kết cấu tương tự dường như không hiệu quả nhưng khả năng đối sánh dựa trên đặc điểm này thường có lợi cho việc phân biệt các vùng ảnh với màu tương tự (ví dụ như bầu trời và biển hoặc lá cây và cỏ). Một loạt các kỹ thuật đã được sử dụng cho việc đo kết cấu tương tự. Công nghệ tốt nhất được thiết lập dựa trên mô hình thống kê có thể tính toán được khoảng cách của kết cấu như mức độ tương phản, độ thô, phương hướng và tính cân đối hoặc chu kỳ, phương hướng và tính ngẫu nhiên. Các phương pháp phân tích kết cấu cho tra cứu thường sử dụng những bộ lọc Gabor. Các truy vấn kết cấu có thể được trình bày tương tự như truy vấn màu sắc bằng việc lựa chọn những mẫu kết cấu như mong muốn từ bảng màu hoặc bằng việc cung cấp ảnh truy vấn mẫu. Hệ thống sau đó sẽ tra cứu những ảnh với giá trị độ đo kết cấu giống nhau nhất với truy vấn.

1.2.4.3. Tra cứu ảnh dựa trên hình dạng.

Khả năng tra cứu bởi hình dạng có lẽ là nhu cầu hiển nhiên nhất ở mức độ nguyên thủy. Không như kết cấu, hình dạng là một khái niệm hoàn toàn rõ ràng, và bằng chứng là những vật thể tự nhiên đầu tiên được nhận thấy bởi hình dạng của chúng. Số lượng đặc điểm cơ bản của hình dạng đối tượng được tính toán cho mỗi đối tượng xác định trong mỗi ảnh. Hai kiểu chính của đặc điểm hình dạng thường được sử dụng là đặc điểm tổng thể (như tỷ lệ bên ngoài), và những đặc điểm cục bộ (như tập các đoạn biên liên tiếp). Các phương pháp khác đề cập tới sự đối sánh hình dạng bao gồm sự biến dạng co giãn của các khuôn dạng....

Những truy vấn đối với hệ thống tra cứu hình dạng thường được biểu diễn bằng cách xác định một hình ảnh mẫu để thực hiện như là hình thức truy vấn hoặc như là một bản phác thảo được vẽ ra bởi người sử dụng.

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

1.2.5.1. Hệ thống QBIC (Query By Image Content).

Hệ thống QBIC cho phép người sử dụng tra cứu ảnh dựa vào màu sắc, hình dạng và kết cấu. QBIC cung cấp một số phương pháp: Simple, Multi-feature, và Multi-pass.

Phương pháp truy vấn Simple chỉ sử dụng một đặc điểm cụ thể để tra cứu ảnh.

Truy vấn Multi-feature bao gồm nhiều hơn một đặc điểm và mọi đặc điểm đều có trọng số như nhau trong suốt quá trình tìm kiếm.

Truy vấn Multi-pass sử dụng đầu ra của các truy vấn trước làm cơ sở cho bước tiếp theo. Người sử dụng có thể vẽ ra và chỉ định màu, kết cấu mẫu của hình ảnh yêu cầu.

Trong hệ thống QBIC màu tương tự được tính toán bằng độ đo bình phương sử dụng biểu đồ màu k phần tử và màu trung bình được sử dụng như là bộ lọc để cải tiến hiệu quả của truy vấn.

1.2.5.2. Hệ thống Photobook.

Hệ thống này được phát triển ở viện kỹ thuật Massachusetts. Nó cho phép người sử dụng tra cứu ảnh dựa trên màu sắc, hình dạng và kết cấu. Hệ thống này cung cấp một tập các thuật toán đối sánh gồm: Euclidean, mahalanobis, histogram, vector space angle, Fourier peak, và wavelet tree distance... Hệ thống như là một công cụ bán tự động và có thể sinh ra một mẫu truy vấn dựa vào những ảnh mẫu được cung cấp bởi người sử dụng. Điều này cho phép người sử dụng trực tiếp đưa những yêu cầu truy vấn của họ với những lĩnh vực khác nhau, và mỗi lĩnh vực họ có thể thu được những mẫu truy vấn tối ưu.

1.2.5.3. Hệ thống VisualSEEK và WebSEEK.

Cả hai hệ thống này đều được phát triển tại Trường Đại học Colombia. VisualSEEK là hệ thống cơ sở dữ liệu ảnh; Nó cho phép người sử dụng tra cứu ảnh dựa trên màu sắc, không gian miền và đặc điểm kết cấu. Thêm vào đó VisualSEEK còn cho phép người sử dụng tạo truy vấn bằng việc chỉ định vùng màu và những

không gian vị trí của chúng. WebSEEK là một catalog ảnh và là công cụ tìm kiếm trên website.

1.2.5.4. Hệ thống RetrievalWare.

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

1.2.5.5. Hệ thống Imatch .

Hệ thống này cho phép người sử dụng tra cứu ảnh bởi nội dung màu, hình dạng, và kết cấu. Nó cung cấp một số phương pháp để tra cứu ảnh tương tự: Màu tương tự, màu và hình dạng, màu và hình dạng mờ, và phân bố màu. Màu tương tự để tra cứu những ảnh tương tự với ảnh mẫu dựa trên sự phân bố màu toàn cục. Màu và hình dạng thực hiện tra cứu bởi việc kết hợp cả hình dạng, kết cấu và màu. Màu và hình dạng mờ thực hiện thêm những bước xác định đối tượng trong ảnh mẫu. Phân bố màu cho phép người sử dụng xác định tỷ lệ phần trăm của một màu trong hình ảnh mong muốn. Imatch cũng cung cấp những đặc điểm khác nội dung để xác định ảnh: ảnh nhị phân, lưu trữ trong những định dạng khác và những ảnh có tên tương tự.

1.2.6. Kết luận

Trong chương này đã cung cấp cái nhìn tổng quan về một hệ thống tra cứu ảnh dựa trên nội dung. Những thành phần, những đặc điểm cũng như những ứng dụng cơ bản của một hệ thống tra cứu ảnh đã được xem xét. Thêm vào đó các chức năng chính của một hệ thống tra cứu ảnh cũng đã được đề cập. Và cuối cùng là một số hệ thống tra cứu ảnh dựa trên nội dung điển hình trên thế giới tiếp cận theo những hướng khác nhau cũng đã được xem xét.

Chương 2 : TÌM HIỂU CÁC PHƯƠNG PHÁP TRÍCH CHỌN ĐẶC TRƯNG HÌNH ẢNH

Trích chọn đặc trưng là cơ sở của tra cứu ảnh dựa vào nội dung. Theo nghĩa rộng, các đặc trưng có thể bao gồm cả các đặc trưng dựa vào văn bản và các đặc trưng trực quan như màu, kết cấu, hình dạng. Trong phạm vi đặc trưng trực quan, các đặc trưng có thể được phân loại tiếp thành các đặc trưng chung và các đặc trưng lĩnh vực cụ thể. Các đặc trưng trực quan chung gồm màu, kết cấu, và hình dạng trong khi các đặc trưng lĩnh vực cụ thể là phụ thuộc ứng dụng. Các đặc trưng lĩnh vực cụ thể bao gồm nhiều tri thức lĩnh vực.

Nhìn chung, không tồn tại một biểu diễn đơn tốt nhất cho một đặc trưng đã cho. Với mọi đặc trưng được cho tồn tại nhiều biểu diễn mô tả đặc trưng từ các cảnh hướng khác nhau.

2.1 Màu sắc

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

Một ảnh màu tiêu biểu được thu từ một camera số, hoặc được tải xuống từ Internet thường có ba kênh màu (các ảnh xám chỉ có một kênh, các ảnh đa phổ có thể có nhiều hơn ba kênh). Tuy nhiên, các giá trị của dữ liệu ba chiều (3 kênh màu) từ ảnh màu không cho chúng ta một mô tả chính xác của màu trong ảnh, nhưng cho vị trí của các điểm ảnh này trong không gian màu. Các điểm ảnh có các giá trị (1,1,1) sẽ xuất hiện khác nhau về màu trong các không gian màu khác nhau.

2.1.1 Không gian màu

Không gian màu là sự biểu diễn tập hợp các màu, một số không gian màu được sử dụng rộng rãi trong đồ họa máy tính. Những mô hình không gian màu có thể được phân biệt thành hướng phần cứng và hướng người sử dụng. Mô hình không gian màu hướng phần cứng bao gồm: RGB, IYQ, CMY dựa trên học thuyết 3 màu. Mô hình không gian màu hướng người sử dụng gồm: HLS, HCV, HSV, ... dựa trên 3 tỷ lệ phần trăm của màu đó là: sắc màu, độ bão hoà, cường độ sáng.

Nếu chúng ta coi thông tin màu của một ảnh như tín hiệu một, hai, hoặc ba chiều, phân tích tín hiệu bằng việc sử dụng ước lượng mật độ xác suất là cách đơn giản nhất để mô tả thông tin màu của ảnh. Lược đồ là công cụ đơn giản nhất.

2.1.2 Lược đồ màu

Lược đồ màu là một tập hợp các mức, mỗi mức biểu thị xác suất của những điểm ảnh trong ảnh. Một biểu đồ màu trong ảnh được định nghĩa là một vectơ:

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

Trong đó :

- i là một màu trong biểu đồ màu và tương ứng với một hình lập phương nhỏ của không gian màu RGB.
- $H[i]$ là số lượng điểm ảnh có màu i trong ảnh.
- N là số mức trong biểu đồ màu.

Trong biểu đồ màu, giá trị của mỗi mức sẽ là tổng số điểm ảnh có cùng màu tương ứng. Để so sánh những ảnh có kích thước khác nhau, biểu đồ màu cần phải chuẩn hoá và được định nghĩa như sau:

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

Trong đó : $H' = \frac{H}{P}$ với P tổng số điểm ảnh trong ảnh.

Lược đồ màu là dễ dàng để tính toán và hiệu quả trong mô tả cả phân bố màu toàn cục và cục bộ trong ảnh. Hơn nữa, nó mạnh với quay và dịch chuyển về trục quan sát và thay đổi chậm với tỷ lệ và góc quan sát.

Rõ ràng, một lược đồ màu chứa nhiều mức hơn, nó có khả năng phân biệt tốt hơn. Tuy nhiên, một lược đồ với một số lượng lớn các mức sẽ không chỉ tăng chi phí tính toán, mà cũng sẽ không thích hợp để xây dựng các cơ chế đánh chỉ số hiệu quả cho các cơ sở dữ liệu ảnh. Chính vì thế cần phải có sự cân nhắc trong việc xác định bao nhiêu mức nên được sử dụng trong biểu đồ màu.

2.2 Kết cấu

Kết cấu là một đặc tính quan trọng khác của ảnh. Các biểu diễn kết cấu đa dạng đã được nghiên cứu trong nhận dạng mẫu và thị giác máy tính. Về cơ bản, các phương pháp biểu diễn kết cấu có thể được phân ra thành hai loại: cấu trúc và thống

kê. Các phương pháp cấu trúc, gồm toán tử hình thái và đồ thị kê, mô tả kết cấu bởi nhận dạng cấu trúc gốc và các luật sắp đặt của chúng. Chúng có khuynh hướng hiệu quả nhất khi được áp dụng với các kết cấu đều. Các phương pháp thống kê, gồm các kỹ thuật phổ năng lượng Fourier, các ma trận đồng khả năng, phân tích thành phần chính bất biến - trượt, đặc trưng Tamura, phân rã Wold, trường ngẫu nhiên Markov mô hình fractal, và lọc đa phân giải như biến đổi Gabor và sóng, mô tả kết cấu bằng phân bố thống kê của cường độ ảnh. Một số biểu diễn kết cấu được sử dụng thường xuyên và đã được chứng minh là hiệu quả trong tra cứu ảnh dựa vào nội dung.

2.2.1 Các đặc trưng Tamura

Các đặc trưng Tamura [39], bao gồm *thô*, *tương phản*, *hướng*, *giống nhất*, *tính chất đều*, và *nhám*, được thiết kế phù hợp với các nghiên cứu tâm lý về nhận thức của người đối với kết cấu. Ba thành phần đầu tiên của các đặc trưng Tamura được sử dụng trong một số hệ thống tra cứu ảnh nổi tiếng ban đầu, như QBIC và Photobook. Các tính toán của ba đặc trưng này được cho ở dưới.

2.2.1.1 Thô

Thô là một độ đo tính chất hệt của kết cấu. Để tính toán thô, các trung bình động $A_k(x, y)$ được tính đầu tiên sử dụng cỡ $2^k \times 2^k$ ($k = 0, 1, \dots, 5$) các cửa sổ tại mỗi điểm ảnh (x, y) , tức là,

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

Ở đây $g(i, j)$ là cường độ điểm ảnh tại (i, j) .

Sau đó, các sự khác nhau giữa các cặp trung bình động không chồng theo hướng ngang và đứng cho mỗi điểm ảnh được tính toán, tức là,

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

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

Sau đó, giá trị của k cực đại hoá E theo một trong hai hướng được sử dụng để đặt cỡ tốt nhất cho mỗi điểm ảnh, tức là,

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

Sau đó thô được tính bằng trung bình S_{best} trên toàn bộ ảnh, tức là,

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

Thay vì lấy trung bình của S_{best} , một phiên bản được cải tiến của đặc trưng thô có thể thu được bởi sử dụng một lược đồ để mô tả phân bố của S_{best} . So với sử dụng một giá trị đơn để biểu diễn thô, sử dụng biểu diễn thô dựa vào lược đồ có thể tăng đáng kể hiệu năng tra cứu. Sự điều chỉnh này làm cho đặc trưng có khả năng xử lý với một ảnh hoặc vùng có đa đặc tính kết cấu, và do đó là hữu ích hơn đối với các ứng dụng tra cứu ảnh.

2.2.1.2 Độ tương phản

Công thức cho tương phản là như sau:

$$F_{con} = \frac{\sigma}{\alpha_4^{1/4}}$$

Ở đây $\alpha_4 = \mu_4 / \sigma^4$, μ_4 là mômen thứ tư về trung bình, và phương sai σ^2 .

Công thức này có thể được sử dụng cho cả toàn bộ ảnh và một vùng của ảnh.

2.2.1.3 Hướng

Để tính hướng, ảnh được chập với hai dãy 3×3 , tức là, $\begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}$ và

$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$ là một véc tơ gradient tại mỗi điểm ảnh được tính.

Độ lớn và góc của véc tơ này được định nghĩa bằng:

$$|\Delta G| = \sqrt{|\Delta_H|^2 + |\Delta_V|^2}$$

$$\theta = \tan^{-1}(\Delta_V / \Delta_H) + \pi / 2$$

Ở đây Δ_H và Δ_V là các khác biệt ngang và dọc của chập.

Sau đó, bằng lượng hoá θ và đếm số các điểm ảnh với độ lớn tương ứng $|\Delta G|$ lớn hơn một ngưỡng, một lược đồ của θ , biểu thị bằng H_D , có thể được xây dựng.

Lược đồ này sẽ cho biết các đỉnh bền vững cho các ảnh hưởng cao và sẽ là tương đối phẳng với các ảnh không có hướng bền vững. Sau đó toàn bộ lược đồ được tóm lược để thu toàn bộ độ đo hướng dựa trên tính nhọn của các đỉnh:

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

Trong tổng các phạm vi p này trên n_p đỉnh; và mỗi đỉnh p , w_p là tập các bin được phân bố trên nó; trong khi ϕ_p là bin nhận giá trị đỉnh.

2.2.2 Các đặc trưng Wold

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

Đối với một trường ngẫu nhiên đều $\{y(m,n), m,n \in Z^2\}$, phân rã Wold cho phép trường được phân rã thành ba thành phần trực giao lẫn nhau:

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

Ở đây $u(m,n)$ là thành phần vô định; và $d(m,n)$ là thành phần tiền định và thành phần tạm thời $e(m,n)$ nó có thể được phân rã tiếp thành thành phần điều hoà $h(m,n)$ và thành phần tạm thời $e(m,n)$. Trong miền tần số, một biểu diễn tương tự tồn tại:

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

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

Trong miền không gian, ba thành phần trực giao có thể thu được bởi ước lượng khả năng nhất (MLE), nó gồm sự điều chỉnh một quá trình AR bậc cao, cực tiểu hoá hàm giá, và giải một tập các phương trình tuyến tính. Trong miền tần số, các thành phần Wold có thể thu được bởi ngưỡng toàn cục của các độ lớn phổ

Fourier của ảnh. Phương pháp sử dụng trích rút đỉnh điều hoà và mô hình tự hồi quy đồng thời đa phân giải (MRSAR) thiếu một phân rã thực sự của ảnh được giới thiệu. Phương pháp này được thiết kế để dung sai sự đa dạng của các sự không đồng nhất trong các mẫu kết cấu tự nhiên.

2.2.3 Mô hình tự hồi quy đồng thời SAR

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

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

Ở đây μ là giá trị xiên được xác định bởi trung bình của toàn bộ ảnh; D là tập lân cận của (x, y) ; $\theta(x', y')$ là tập các trọng số được kết hợp với mỗi điểm ảnh lân cận; $\varepsilon(x, y)$ là một biến ngẫu nhiên độc lập Gaussian với trung bình không và phương sai σ^2 . Các tham số θ và σ được sử dụng để đo kết cấu. Thí dụ, một giá trị σ cao hơn hàm ý tính chất hạt tốt hơn hoặc thô kém hơn; các giá trị $\theta(x, y + 1)$ và $\theta(x, y - 1)$ cao hơn chỉ ra rằng kết cấu được hướng thẳng đứng. Kỹ thuật sai số bình phương tối thiểu (LSE) hoặc phương pháp MLE thường được sử dụng để đánh giá các tham số của mô hình SAR.

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

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

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

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

Ở đây N_i là lân cận tròn thứ i của (x, y) ; $w_i(x', y')$ là một tập các trọng số được tính trước chỉ ra đóng góp của điểm ảnh (x', y') trong vòng tròn thứ i .

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

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

2.2.4 Các đặc trưng lọc Gabor

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

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

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

Ở đây, σ_x và σ_y là các độ lệch chuẩn của phân bố Gaussian theo hướng x và y .

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

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

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

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

Ở đây $a > 1, \theta = n\pi / K, n = 0, 1, \dots, K-1$, và $m = 0, 1, \dots, S-1$. K và S là số các hướng và các tỷ lệ. Nhân tố tỷ lệ a^{-m} là để đảm bảo rằng năng lượng là độc lập của m .

Một ảnh $I(x, y)$ đã cho, biến đổi Gabor của nó được định nghĩa bằng:

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

Ở đây * chỉ ra số liên hợp phức. Sau đó trung bình μ_{mn} và độ lệch chuẩn σ_{mn} của độ lớn $W_{mn}(x, y)$, tức là, $f = [\mu_{00}, \sigma_{00}, \dots, \mu_{mn}, \sigma_{mn}, \Lambda, \mu_{S-1, K-1}, \sigma_{S-1, K-1}]$ có thể được sử dụng để biểu diễn đặc trưng kết cấu của một vùng kết cấu thuần nhất.

2.2.5 Các đặc trưng biến đổi sóng

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

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

Ở đây, m và n là các tham số giãn và dịch chuyển. Một tín hiệu $f(x)$ có thể được biểu diễn bằng:

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

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

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

2.3 Hình dạng

Các đặc trưng hình của các đối tượng hoặc các vùng đã được sử dụng trong nhiều hệ thống tra cứu ảnh dựa vào nội dung. So với các đặc trưng màu và kết cấu, các đặc trưng hình thường được mô tả sau khi các ảnh được phân đoạn thành các vùng hoặc các đối tượng. Do phân đoạn ảnh mạnh và chính xác là khó đạt được, sử dụng các đặc trưng hình cho tra cứu ảnh bị giới hạn đối với các ứng dụng chuyên biệt, ở đó các đối tượng hoặc các vùng đã sẵn có. Các phương pháp state-of-art cho mô tả hình có thể được phân thành hoặc là các phương pháp dựa vào đường bao hoặc các phương pháp dựa vào vùng. Một biểu diễn đặc trưng hình tốt cho một đối tượng phải bất biến với dịch chuyển, quay và tỷ lệ.

2.3.1 Các bất biến mômen

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

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

Ở đây (x_c, y_c) là tâm của đối tượng. Mômen trung tâm này có thể được chuẩn hoá để bất biến tỷ lệ :

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

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

$$\phi_1 = \mu_{2,0} + \mu_{0,2}$$

$$\phi_2 = (\mu_{2,0} - \mu_{0,2})^2 + 4\mu_{1,1}^2$$

$$\phi_3 = (\mu_{3,0} - 3\mu_{1,2})^2 + (\mu_{0,3} - 3\mu_{2,1})^2$$

$$\phi_4 = (\mu_{3,0} + \mu_{1,2})^2 + (\mu_{0,3} + \mu_{2,1})^2$$

$$\phi_5 = (\mu_{3,0} - 3\mu_{1,2})(\mu_{3,0} + \mu_{1,2}) \left[(\mu_{3,0} + \mu_{1,2})^2 - 3(\mu_{0,3} + \mu_{2,1})^2 \right] -$$

$$+ (\mu_{0,3} - 3\mu_{2,1})(\mu_{0,3} + \mu_{2,1}) \left[(\mu_{0,3} + \mu_{2,1})^2 - 3(\mu_{3,0} + \mu_{1,2})^2 \right]$$

$$\phi_6 = (\mu_{2,0} - \mu_{0,2}) \left[(\mu_{3,0} + \mu_{1,2})^2 - (\mu_{0,3} + \mu_{2,1})^2 \right] + 4\mu_{1,1}(\mu_{3,0} + \mu_{1,2})(\mu_{0,3} + \mu_{2,1})$$

$$\phi_7 = (3\mu_{2,1} - \mu_{0,3})(\mu_{3,0} + \mu_{1,2}) \left[(\mu_{3,0} + \mu_{1,2})^2 - 3(\mu_{0,3} + \mu_{2,1})^2 \right] -$$

2.3.2 Các góc quay

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

$$\theta(s) = \tan^{-1} \left(\frac{y'_s}{x'_s} \right)$$

$$y'_s = \frac{dy_s}{ds}$$

$$x'_s = \frac{dx_s}{ds}$$

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

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

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

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

2.3.3 Các ký hiệu mô tả Fourier

Các ký hiệu mô tả Fourier mô tả hình của một đối tượng với biến đổi Fourier của đường biên của nó. Xét đường biên đóng của một đối tượng hai chiều bằng một dãy đóng các điểm biên liên tiếp (x_s, y_s) , ở đây $0 \leq s \leq N-1$ và N là tổng số các điểm ảnh trên đường biên. Sau đó ba loại biểu diễn đường biên đóng, tức là, độ cong, khoảng cách trọng tâm, và hàm tọa độ phức hợp, có thể được định nghĩa.

Độ cong $K(s)$ tại một điểm s dọc theo đường biên đóng được định nghĩa như tỷ lệ thay đổi theo hướng \tan của đường biên đóng, tức là,

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

Ở đây $\theta(s)$ là hàm quay của chu tuyến.

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

$$R(s) = \sqrt{(x_s - x_c)^2 + (y_s - y_c)^2} \text{ Tọa độ phức hợp thu được bởi biểu diễn đơn giản các tọa độ của các điểm biên như các số phức hợp:}$$

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

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

Các ký hiệu mô tả Fourier của đường cong là:

$$f_K = [|F_1|, |F_2|, \dots, |F_{M/2}|]$$

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

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

Ở đây F_i biểu thị thành phần thứ i của các hệ số biến đổi Fourier. Ở đây chỉ các trục tần số dương được xem xét bởi vì đường cong và các hàm khoảng cách trọng tâm là thực và, do đó, các biến đổi Fourier của nó biểu lộ tính đối xứng, tức là, $|F_{-i}| = |F_i|$.

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

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

Ở đây F_1 là thành phần tần số khác không đầu tiên được sử dụng để chuẩn hoá các hệ số biến đổi. Ở đây cả hai thành phần tần số dương và âm được xem xét. Hệ số DC là phụ thuộc vào vị trí của hình, và do đó, bị loại bỏ.

Để đảm bảo các đặc trưng hình kết quả của tất cả các đối tượng trong một cơ sở dữ liệu có cùng độ dài, đường biên $((x_s, y_s), 0 \leq s \leq N-1)$ của mỗi đối tượng được lấy mẫu lại với M mẫu trước khi thực hiện biến đổi Fourier. Thí dụ, M có thể đặt tới $2^m = 64$ sao cho biến đổi có thể được thực hiện hiệu quả sử dụng biến đổi Fourier nhanh.

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

Hình tròn được tính toán bằng:

$$\alpha = \frac{4\pi S}{P^2}$$

Ở đây S là cỡ và P là chu vi của một đối tượng. Giá trị này có phạm vi giữa 0 và 1.

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

2.4 Thông tin không gian

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

Thu thông tin không gian của các đối tượng trong một ảnh là một quá trình quan trọng đối với các hệ thống GIS. Quá trình này bao gồm việc biểu diễn vị trí không gian tuyệt đối và cũng bao gồm vị trí không gian tương đối của các đối tượng. Các thao tác như giao và chồng được sử dụng. Bộ cục màu kết hợp thông tin không gian với thông tin màu xuất hiện trong ảnh và tạo ra một đặc trưng rất quan trọng trong quá trình tra cứu.

Tuy nhiên, tìm kiếm các ảnh dựa trên các quan hệ không gian của các vùng còn lại một vấn đề nghiên cứu khó trong tra cứu ảnh dựa vào nội dung, do phân đoạn tin cậy của các đối tượng hoặc các vùng thường là không khả thi ngoại trừ các ứng dụng rất giới hạn.

2.5 Phân đoạn

Phân đoạn là quá trình phân ảnh ra thành các vùng mà về lý tưởng nó sẽ tương ứng với các đối tượng xuất hiện trong ảnh. Đây là bước rất quan trọng đối với tra cứu ảnh. Cả đặc trưng hình và đặc trưng bố cục phụ thuộc vào phân đoạn tốt.

Trong phân đoạn các yêu cầu chính xác phân đoạn là rất khác nhau cho các đặc trưng hình và các đặc trưng bố cục. Với các đặc trưng hình, phân đoạn chính xác là mong muốn cao trong khi các đặc trưng bố cục, một phân đoạn thô có thể là đủ.

Chương 3: CÁC ĐỘ ĐO TƯƠNG TỰ

Đầu tiên các đặc trưng của các ảnh trong cơ sở dữ liệu được trích rút và được mô tả bởi các véc tơ đặc trưng nhiều chiều. Các véc tơ đặc trưng của các ảnh trong cơ sở dữ liệu tạo thành một cơ sở dữ liệu đặc trưng. Sau đó để tra cứu các ảnh, người sử dụng cung cấp cho hệ thống tra cứu ảnh mẫu hoặc hình phác thảo, các đặc trưng của ảnh mẫu hoặc hình phác thảo (gọi là ảnh truy vấn) được trích rút. Các kết quả tìm kiếm thu được bởi độ đo sự tương tự giữa các đặc trưng của ảnh cơ sở dữ liệu và ảnh truy vấn.

Độ đo tương tự lý tưởng sẽ có một số hoặc tất cả các đặc tính cơ sở sau:

Tương tự nhận thức :Khoảng cách đặc trưng giữa hai ảnh chỉ lớn nếu hai ảnh là không tương tự, và nhỏ nếu các ảnh là tương tự. Các ảnh thường hay được mô tả trong không gian đặc trưng và độ tương tự giữa các ảnh thường được đo bởi một độ đo khoảng cách trong không gian đặc trưng. Đem vào bản miêu tả các đặc tính của không gian này với nhận thức của con người và các đặc tính cơ sở của các véc tơ đặc trưng biểu diễn các ảnh là rất quan trọng trong cải tiến đặc tính tương tự nhận thức của độ đo tương tự được đề xuất.

Hiệu quả: Độ đo cần được tính toán nhanh để có phản hồi nhanh trong pha tìm kiếm. Các ứng dụng CBIR tiêu biểu đòi hỏi một phản hồi rất nhanh, không lâu hơn vài giây. Trong chu kỳ thời gian ngắn đó, máy tìm kiếm thường phải tính toán hàng ngàn khoảng cách phụ thuộc vào cỡ của cơ sở dữ liệu ảnh. Do đó độ phức tạp của độ đo khoảng cách là quan trọng.

Khả năng: Hiệu năng của hệ thống không được giảm quá nhiều cho các cơ sở dữ liệu lớn do một hệ thống có thể tìm kiếm trong các cơ sở dữ liệu chứa hàng triệu ảnh. Một sự thực hiện của hệ thống CBIR tính toán tất cả các khoảng cách giữa ảnh truy vấn và các ảnh trong cơ sở dữ liệu. Sau đó các khoảng cách này được lưu trữ để tìm ra các ảnh tương tự nhất đối với ảnh truy vấn. Do đó sự phức tạp của máy tìm kiếm phải tương ứng với cỡ của cơ sở dữ liệu ảnh (hoặc nếu chúng ta coi N là số các ảnh). Các kỹ thuật đánh chỉ số nhiều chiều (như được đề cập trong phần 1.2) có thể được sử dụng để giảm sự phức tạp xuống. Tuy nhiên, đã có báo cáo rằng hiệu

năng của các kỹ thuật đánh chỉ số hiện tại bị giảm về tuyến tính khi số chiều cần được đánh chỉ số là lớn hơn 20. Vậy chúng ta phải xem xét nhân tố này khi giải quyết với các cơ sở dữ liệu ảnh rất lớn.

Khoảng cách: Vấn đề khoảng cách tương tự có nên là độ đo hay không vẫn chưa được quyết định do thị giác người là rất phức tạp và các cơ chế của hệ thống trực quan người chưa được hiểu đầy đủ. Chúng ta muốn khoảng cách tương tự là độ đo do chúng ta xem các đặc tính dưới đây như các yêu cầu rất tự nhiên.

Sự bất biến của bản thân sự tương tự: Khoảng cách giữa một ảnh với bản thân nó phải bằng với một hằng số độc lập với ảnh.

Tối thiểu: Một ảnh phải tương tự với chính nó hơn là với các ảnh khác.

Tính đối xứng: Nếu ảnh tương tự với ảnh thì ảnh phải tương tự với ảnh A .

Tính bắc cầu: Cũng là vô lý nếu ảnh rất tương tự với ảnh, và lại rất tương tự với ,nhưng lại rất khác với A .

Tuy nhiên, tính chất bắc cầu này có thể không giữ cho một chuỗi các ảnh. Ngay khi nếu ảnh tương tự với ảnh với. Điều này không có nghĩa rằng ảnh tương tự với ảnh. Thí dụ, trong một cảnh video mỗi frame tương tự với các frame lân cận của nó nhưng frame đầu tiên và frame cuối cùng của cảnh có thể rất khác nhau.

Tính chất mạnh: Hệ thống phải mạnh đối với các thay đổi trong các điều kiện ảnh của các ảnh cơ sở dữ liệu. Thí dụ nếu các ảnh trong cơ sở dữ liệu thu được dưới ánh sáng đèn điện, hệ thống tra cứu phải có thể tìm thấy các đối tượng này ngay cả khi đối tượng truy vấn thu được dưới ánh nắng ban ngày.

Nhiều độ đo tương tự đã được đề xuất, nhưng không có độ đo nào có tất cả các đặc tính nói trên. Dưới đây là một số độ đo tương tự được sử dụng phổ biến nhất.

3.1 Lược đồ giao

Đây là một trong những độ đo khoảng cách đầu tiên trong tra cứu ảnh dựa vào màu. Khoảng cách được xác định dựa trên cơ phần chung của hai lược đồ màu. Giả sử hai lược đồ màu được ký hiệu là h_1 và h_2 , khoảng cách giữa chúng có thể được định nghĩa bằng:

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

Độ đo khoảng cách này là nhanh. Tuy nhiên, nó không là độ đo và thông tin màu không được sử dụng khi thu khoảng cách. Điều này có thể dẫn đến các kết quả không mong muốn.

3.2 Khoảng cách Minkowski

Khoảng cách L_1 , khoảng cách dạng Minkowski L_p : khoảng cách dạng Minkowski L_p giữa hai lược đồ được định nghĩa như sau:

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

3.3 Khoảng cách dạng toàn phương : khoảng cách giữa hai lược đồ màu N chiều h_1 và h_2 được định nghĩa như

$$dist_{QF} = \sqrt{A \|h_1 - h_2\|^2}$$

Ở đây $A = [a_{ij}]$ là một ma trận và các trọng số a_{ij} biểu thị sự tương tự giữa các bin i và j . Thông thường a_{ij} được cho bằng

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

Ở đây d_{ij} là khoảng cách giữa màu i và màu j (thông thường d_{ij} là khoảng cách O Co Lít giữa hai màu trong một số không gian màu đồng nhất như La^*b^* hoặc Lu^*v^*) và $d_{\max} = \max_{ij} (d_{ij})$. k là một hằng số điều khiển trọng số giữa các màu lân cận.

Sự lựa chọn thông thường khác cho a_{ij}

$$a_{ij} = \exp \left(-k \left(d_{ij} / d_{\max} \right)^2 \right)$$

3.4 Khoảng cách EMD : là dựa trên giá trị tối thiểu để biến đổi một phân bố thành một phân bố khác. Nếu giá trị dịch chuyển một đơn vị đặc trưng trong không gian đặc trưng là khoảng cách nền, thì khoảng cách giữa hai phân bố được cho bởi tổng các giá trị tối thiểu để di chuyển tất cả các đặc trưng riêng lẻ. EMD có thể được định nghĩa như giải pháp của vấn đề vận tải có thể được giải quyết bằng tối ưu tuyến tính:

$$dist_{EMD} = \frac{\sum_{ij} g_{ij} d_{ij}}{\sum_{ij} g_{ij}}$$

Ở đây d_{ij} biểu thị sự không tương tự giữa các bin i và j , và $g_{ij} \geq 0$ là luồng tối ưu giữa hai phân bố sao cho tổng giá trị

$$dist_{EMD} = \sum_{ij} g_{ij} d_{ij}$$

là cực tiểu, tùy vào các ràng buộc sau:

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

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

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

Với tất cả i và j . Mẫu số trong phương trình là một hệ số chuẩn hoá cho phép đối sánh các phần của các phân bố với tổng khối lượng khác nhau. Nếu khoảng cách nền là một độ đo và hai phân bố có cùng số lượng tổng khối lượng, EMD xác định một độ đo. Như một ưu điểm chính của EMD mỗi ảnh có thể được biểu diễn bởi các bin khác nhau mà thích nghi với phân bố cụ thể của chúng. Khi các lược đồ lè được sử dụng. Các giá trị không tương tự thu được cho các chiều riêng lẻ phải được kết hợp thành một giá trị không tương tự liên kết.

Các độ đo khoảng cách khác cũng được quan tâm là:

- *Khoảng cách Kolmogorov-Smirnov* được đề xuất trong . Nó được định nghĩa như sự khác nhau cực đại giữa các phân bố tích lũy

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

Ở đây h^c là lược đồ tích lũy của lược đồ h

- *Thống kê kiểu Cramer/Von Mises* dựa trên các phân bố tích lũy được định nghĩa

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

- Thống kê χ^2 được cho bởi

$$dist_{\chi} = \sum_i \frac{\left(h_{1i} - \hat{h}_i \right)^2}{\hat{h}_i}$$

Ở đây $\hat{h}_i = \frac{h_{1i} + h_{2i}}{2}$ biểu thị ước lượng chung.

- *Kullback-Leibler divergence* được định nghĩa bởi

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

- *Jeffrey-divergence* được định nghĩa bởi

$$dist_{JD} = \sum \left(h_{1i} \log \frac{h_{1i}}{\hat{h}_i} + h_{2i} \log \frac{h_{2i}}{\hat{h}_i} \right)$$

- *Weighted-Mean-Variance* được đề xuất trong . Khoảng cách này được định nghĩa bởi

$$dist_{WMV} = \frac{\mu_1 - \mu_2}{\sigma(\mu)} + \frac{\sigma_1 - \sigma_2}{\sigma(\sigma)}$$

Ở đây μ_1, μ_2 là các tham số thực nghiệm và σ_1, σ_2 là các độ lệch chuẩn của hai lược đồ h_1, h_2 . $\sigma(\cdot)$ biểu thị sự ước lượng của độ lệch chuẩn của thực thể tương ứng.

- *Khoảng cách Bhattacharyya* được định nghĩa

$$d_B^2(N(\mu_1, \Sigma_1), N(\mu_2, \Sigma_2)) = \frac{1}{8} (\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2) + \frac{1}{2} \ln \frac{\det \Sigma}{\sqrt{\det \Sigma_1 \det \Sigma_2}}$$

Ở đây $\Sigma = 0.5 \times (\Sigma_1 + \Sigma_2)$

- *Khoảng cách Mahalanobis* được cho bởi

$$d_B^2(N(\mu_1, \Sigma), N(\mu_2, \Sigma)) = (\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2)$$

Với các mô tả chi tiết hơn, chúng ta tham khảo các bài báo được trích dẫn. Trong cung cấp một sự so sánh toàn diện trên nhiều độ đo khoảng cách khác nhau.

Chương 4 :CÀI ĐẶT THỬ NGHIỆM CHƯƠNG TRÌNH,KẾT LUẬN ,TÀI LIỆU THAM KHẢO

4.1 KẾT LUẬN

Với sự đi lên ngày càng mạnh của công nghệ thông tin nói chung và ứng dụng xử lý ảnh trong thực tiễn nói riêng đã cho thấy rằng đây là một đề tài không hề cũ và là một phần hết sức quan trọng trong các ứng dụng thực tế đặt ra. Qua quá trình thực hiện đề tài em đã học tập được rất nhiều kiến thức bổ ích, nâng cao khả năng tri thức của mỗi người. Việc tìm hiểu một số phương pháp trích chọn đặc trưng hình ảnh theo nội dung và các độ đo tương tự đã góp phần hệ thống hoá kiến thức đồng thời giúp em tiến gần hơn đến các ứng dụng thực tiễn mà cuộc sống đang dần đòi hỏi.

Mặc dù đã cố gắng hết sức để hoàn thành đề tài một cách tốt nhất; song xử lý ảnh là đề tài lần đầu tiên chúng em làm nên đã gặp rất nhiều khó khăn về tài liệu cũng như kinh nghiệm trong lĩnh vực này. Với hiểu biết còn hạn hẹp , em rất mong sự giúp đỡ ,đóng góp ý kiến của các thầy cô giáo và các bạn để đề tài ngày càng hoàn thiện hơn.

Em xin chân thành cảm ơn PGS.TS Ngô Quốc Tạo đã tận tình hướng dẫn em hoàn thành đề tài và qua đây em cảm ơn tất cả thầy cô giáo và bạn bè đã tạo điều kiện, giúp đỡ em trong thời gian qua.

4.2 TÀI LIỆU THAM KHẢO

1. A Robust CBIR Approach Using Local Color Histograms by Shengjiu Wang
2. Feature Extraction from Images
3. Comparing Images Using the Hausdor Distance by Daniel P. Huttenlocher, Gregory A. Klanderman and William J. Rucklidge

MỤC LỤC

Chương 1: TỔNG QUAN VỀ XỬ LÝ ẢNH VÀ TRA CỨU ẢNH	1
1.1 TỔNG QUAN VỀ XỬ LÝ ẢNH.....	1
1.1.1 Một số khái niệm	1
1.1.1.1 Pixel (Picture Element):	1
1.1.1.2 Gray level:	2
1.1.1.3 Định dạng ảnh	2
1.1.2 Biểu diễn ảnh.....	5
1.1.3 Tăng c-ờng ảnh - khôi phục ảnh	5
1.1.4 Biến đổi ảnh.....	6
1.1.5 Phân tích ảnh	7
1.1.6 Nhận dạng ảnh.....	7
1.1.7 Nén ảnh	7
1.2 TỔNG QUAN VỀ TRA CỨU ẢNH DỰA TRÊN NỘI DUNG	8
1.2.1 Những thành phần của một hệ thống tra cứu ảnh dựa trên nội dung.	8
1.2.1.1 Công nghệ tự động trích chọn siêu dữ liệu.	8
1.2.1.2. Giao diện để lấy yêu cầu truy vấn người sử dụng.....	8
1.2.1.3 Phương pháp so sánh độ tương tự giữa các ảnh.....	9
1.2.1.4 Công nghệ tạo chỉ số và lưu trữ dữ liệu hiệu quả.	9
1.2.2 Những ứng dụng cơ bản của tra cứu ảnh.	10
1.2.3 Các chức năng của hệ thống tra cứu ảnh dựa trên nội dung.....	10
1.2.4 Các phương pháp tra cứu ảnh dựa trên nội dung.....	13
1.2.4.1. Tra cứu ảnh dựa trên màu sắc.	13
1.2.4.2. Tra cứu ảnh dựa trên kết cấu.	14
1.2.4.3. Tra cứu ảnh dựa trên hình dạng.	14
1.2.5. Những hệ thống tra cứu ảnh dựa trên nội dung.	15
1.2.5.1. Hệ thống QBIC (Query By Image Content).	15
1.2.5.2. Hệ thống Photobook.....	15
1.2.5.3. Hệ thống VisualSEEK và WebSEEK.	15
1.2.5.4. Hệ thống RetrievalWare.....	16
1.2.5.5. Hệ thống Imatch	16
1.2.6. Kết luận	16
Chương 2 : TÌM HIỂU CÁC PHƯƠNG PHÁP TRÍCH CHỌN ĐẶC TRƯNG HÌNH ẢNH	17
2.1 Màu sắc	17
2.1.1 Khung gian màu	17

2.1.2 Lược đồ màu	18
2.2 Kết cấu.....	18
2.2.1 Các đặc trưng Tamura.....	19
2.2.1.1 Thô	19
2.2.1.2 Độ tương phản.....	20
2.2.1.3 Hướng.....	20
2.2.2 Các đặc trưng Wold.....	21
2.2.3 Mô hình tự hồi qui đồng thời SAR	22
2.2.4 Các đặc trưng lọc Gabor	23
2.2.5 Các đặc trưng biến đổi sóng	24
2.3 Hình dạng	25
2.3.1 Các bất biến mômen.....	25
2.3.2 Các góc quay.....	26
2.3.3 Các ký hiệu mô tả Fourier.....	27
2.3.4 Hình tròn, độ lệch tâm, và hướng trục chính	28
2.4 Thông tin không gian	29
2.5 Phân đoạn	29
Chương 3: CÁC ĐỘ ĐO TƯƠNG TỰ	30
3.1 Lược đồ giao	31
3.2 Khoảng cách Minkowski.....	32
3.3. Khoảng cách dạng toàn phương.....	32
3.4 Khoảng cách EMD :	32
Chương 4 :CÀI ĐẶT THỬ NGHIỆM CHƯƠNG TRÌNH,KẾT LUẬN ,TÀI LIỆU THAM KHẢO	35
4.1 KẾT LUẬN.....	35
4.2 TÀI LIỆU THAM KHẢO.....	35