MỤC LỤC

PHẦN MỞ ĐẦU 3

Chương 1: KHÁT QUÁT VỀ TƯƠNG QUAN VÀ ĐỘ ĐỌ TƯƠNG QUAN GIỮA HAI ẢNH 5

1.1. Tính "ghép đúng" và tính "tương quan" 5

1.1.1. Khái niệm về độ tương quan giữa hai ảnh 5

1.1.2. Độ đo tương quan 6

1.2. Xác định độ đo nội dung ảnh 10

1.2.1. Độ đo thuộc tính màu sắc 11

1.2.1.1. Histogram 11

1.2.1.2. Moment màu 14

1.2.1.3. Vectơ gắn kết màu 15

1.2.1.4. Tương quan màu 15

1.2.2. Độ đo thuộc tính hình dạng 16

1.2.2.1. Cơ sở vùng 16

1.2.2.2. Cơ sở biên 20

1.2.3. Độ đo thuộc tính cấu trúc bề mặt 22

1.2.3.1. Các phương pháp không gian 22

1.2.3.2. Phương pháp tần số 24

1.2.3.2. Phương pháp moment 25

Chương 2: MỘT SỐ KỸ THUẬT ĐÁNH GIÁ ĐỘ TƯƠNG QUAN 27

2.1. Mô hình không gian vector VSM 27

2.1.1. Phép so sánh histogram 27

2.1.1.1. So sánh ngang các bin histogram (bin-by-bin) 27

2.1.1.2. So sánh chéo các bin histogram (cross-bin) 30

2.1.1.3. Phép so sánh qua giá trị điểm ảnh 32

2.2. Mô hình Vector 34

2.2.1. SVM 34

2.2.2. SVM trong kỹ thuật tra cứu ảnh 36
2.3. Mô hình k-phần tử kề cận (k-NN).....39
2.3.1. Thuật toán k-NN 39
2.3.2. k-NN trong so khớp điểm ảnh 42
Chương 3: CHƯƠNG TRÌNH THỬ NGHIỆM.....44
3.1. Bài toán 44
3.2. Xây dựng chương trình 44
3.2.1. Lựa chọn môi trường 44
3.2.2. Phân tích về lôgô 44
3.2.3. Đánh giá độ tương quan về lôgô.....45
3.2.3.1. Trích chọn các đặc trưng cho lôgô 45
3.2.3.2. So sánh độ đo tương quan giữa các cặp lôgô 45
3.2.4. Một số kết quả 46
PHẦN KẾT LUẬN 51
TÀI LIỆU THAM KHẢO.....52

PHẦN MỞ ĐẦU

Trong những năm gần đây, lĩnh vực xử lý ảnh số ngày càng được nhiều người quan tâm, sự phát triển nhanh chóng của các thiết bị đồ hoạ cũng như dung lượng của các thiết bị lưu trữ ngày càng tăng nhanh là những nhân tố tích cực thúc đẩy nghiên cứu các ứng dụng thực tế từ công nghệ xử lý ảnh.

Nhận dạng ảnh chính là việc dùng chương trình máy tính để phân tích nội dung của ảnh. Có rất nhiều các hướng nghiên cứu xác định nội dung của ảnh cũng như phân định các đối tượng thuộc bức ảnh. Các kết quả đạt được ở mỗi phương pháp đều có thể được đưa vào ứng dụng trong thực tiễn để giải quyết một vấn đề cụ thể nào đó.

Với việc không còn bị hạn chế về độ lớn của bộ nhớ dùng để lưu trữ, vấn đề đặt ra là khai thác kho dữ liệu ảnh như thế nào cho hiệu quả. Bằng trực quan con người có thể dễ dàng hiểu được nội dung một bức ảnh, nhưng để máy tính cũng hiểu được nội dung đó thì quả thực là một vấn đề khó. Ngay cả khi chúng ta hạn chế vấn đề ở chỗ chỉ đem so sánh nội dung các bức ảnh với nhau theo góc độ đánh giá tính tương quan thì cũng đã đặt ra nhiều vấn đề lớn cho việc nghiên cứu: so sánh dựa trên các đặc trưng nào, giống nhau hay khác nhau ở mức độ bao nhiêu? Để tìm hiểu sâu hơn vấn đề này, thực tế đặt ra cho chúng ta bài toán là hãy xác định xem liệu một đối tượng (được mô tả dưới dạng ảnh – có thể là trích một phần từ một bức ảnh lớn hơn) có xuất hiện ở một ảnh nào trong tập ảnh (cơ sở dữ liệu đầu vào) cho trước hay không? Trong các nghiên cứu khoa học hiện nay, vấn đề này được xếp vào nhóm các kỹ thuật *tra cứu ảnh theo nội dung*. Các kỹ thuật này cho phép trích chọn đặc điểm dựa vào nội dung trực quan bao gồm *màu sắc, kết cấu, hình dạng*, *bố cục không gian*... của ảnh, từ đó làm cơ sở cho việc tra cứu, sắp xếp, tổ chức cơ sở dữ liệu ảnh.

Nghiên cứu - tìm hiểu - đánh giá các phương pháp đã có để đi tìm lời giải cho bài toán trên đây chính là nội dung của đề tài "Tìm hiểu bài toán đánh giá sự tương quan giữa hai ảnh". Để đánh giá được độ tương quan của ảnh như đã phân tích, trước hết là phải xác định được *các độ đo nội dung* cần thiết để so sánh, sau đó phải xây dựng được *hàm đánh giá*. Đó chính là hai mục tiêu cơ bản được nghiên cứu trong đề tài này. Trên cơ sở các nghiên cứu đó, đề tài sẽ thử nghiệm một phương pháp cụ thể để xây dựng một chương trình phần mềm cho phép tra cứu một mẫu

lôgô thương mại xem nó đã có hay chưa có trong kho cơ sở dữ liệu về ảnh lôgô thương mại đã lưu trữ (đã đăng ký) bằng cách liệt kê ra 20 mẫu lôgô có nội dung ảnh gần giống nhất với mẫu lôgô đưa vào, qua đó cho phép người dùng quan sát và quyết định có cho đăng ký (lưu trữ) mẫu lôgô đó hay không. Vì vậy, nội dung chính của đồ án được trình bày bao gồm Phần mở đầu, Phần kết luận và ba chương nội dung, cụ thể:

Chương 1: KHÁT QUÁT VỀ TƯƠNG QUAN VÀ ĐỘ ĐỌ TƯƠNG QUAN GIỮA HAI ẢNH

Nội dung chương này đi vào phân tích các chi tiết cấu thành nội dung của bức ảnh theo khía cạnh nhận thức của thị giác con người. Đồ án giới thiệu các phương pháp cũng như các vector đặc trưng dùng để mô tả nội dung của một bức ảnh. Đây chính là cơ sở để chúng ta thực hiện các phép tính toán so sánh các bức ảnh với nhau ở chương 2.

Chương 2: MỘT SỐ KỸ THUẬT ĐÁNH GIÁ ĐỘ TƯƠNG QUAN

Đưa ra các kỹ thuật đánh giá độ tương quan giữa các bức ảnh dựa trên độ đo nội dung ảnh (vector đặc trưng). Tính đến nay, đã có rất nhiều các kỹ thuật được giới thiệu. Để hệ thống hoá và phân loại, các kỹ thuật này sẽ được trình bày theo tiêu chí phân loại các mô hình độ tương quan.

Chương 3: CHƯƠNG TRÌNH THỬ NGHIỆM

Giới thiệu chương trình phần mềm tự xây dựng nhằm mô phỏng cho các lý thuyết đã đề cập trong đồ án.

Chương 1:

KHÁT QUÁT VỀ TƯƠNG QUAN VÀ ĐỘ ĐỌ TƯƠNG QUAN GIỮA HAI ẢNH

1.1. Tính "ghép đúng" và tính "tương quan"

Việc so sánh hai bức ảnh với nhau không dừng lại ở chỗ chỉ so sánh các điểm ảnh với nhau, nó cần phải được xem xét dựa trên sự nhìn nhận trực quan của con người, tức là quan tâm đến nội dung trực quan của ảnh.

Quá trình đánh giá độ tương quan giữa các ảnh có thể phân chia thành hai bước chính. Trước tiên là xác định định lượng một độ đo về nội dung của bức ảnh dựa trên tập các đặc trưng được hình thành từ việc mã hoá các thuộc tính của ảnh. Tập các đặc trưng đó hình thành nên vector đặc trưng hay độ đo nội dung của ảnh trên không gian các đặc trưng. Bước thứ hai sau đó là đánh giá so sánh độ tương quan dựa trên các định lượng về số đo ở bước trước. Nếu cùng một không gian đặc trưng mà cho một độ đo khác biệt lớn thì có thể kết luận chúng không tương quan nhau hay chúng không được xếp trong cùng một lớp. Thông thường, khi "so sánh" hai bức ảnh thì hoặc là ta quan tâm đến việc đánh giá "độ giống nhau" hay "độ tương quan" giữa chúng là bao nhiêu hoặc cũng có thể quan tâm đến khía cạnh hai bức ảnh đó khác biệt nhau đến mức độ nào.

Ở đây ta thấy có một khái niệm mang ý nghĩa hẹp hơn, sâu hơn khái niệm "tương quan", đôi khi mập mờ, đó là "khớp đúng" hay "ghép đúng"(matching). Tính tương phản của hai khái niệm này có thể phân biệt được qua các ứng dụng của chúng. Các kỹ thuật ghép đúng được phát triển chủ yếu cho các ứng dụng về nhận dạng đối tượng biến đổi hình dạng, trong khi các kỹ thuật về độ tương quan thì lại thấy chủ yếu trong các ứng dụng sử dụng cơ sở dữ liệu về nội dung trực quan của ảnh, đặc biệt là các ứng dụng về tra cứu ảnh.

1.1.1. Khái niệm về độ tương quan giữa hai ảnh

Gọi $\{F(x, y); x, y = 1,2,..., N\}$ là mảng hai chiều mô tả các điểm ảnh của một bức ảnh. Giá trị hàm $F(x, y)$ là giá trị màu của điểm ảnh có vị trí xác định tại toạ độ (x, y). Giả sử ảnh xây dựng trên hệ màu RGB thì hàm $F(x, y)$ có thể viết:

$F(x,y)=\{F_R(x,y),F_G(x,y),F_B(x,y)\}$.

Với ảnh đen trắng thì giá trị hàm $F(x,y)$ mô tả giá trị xám của điểm ảnh tại toạ độ (x,y). Gọi f là ánh xạ từ không gian ảnh vào không gian các đặc trưng N chiều của ảnh, $X=\{x_1,x_2,...,x_n\}$ tức là:

$$f:F\to X$$

trong đó n là số các đặc trưng trích chọn trong ảnh. Sự khác biệt nhau giữa hai ảnh F_1 và F_2 có thể hiểu như là khoảng khác biệt D đo được giữa các vector đặc trưng tương ứng X_1 và X_2. Khi đó việc mô tả quá trình tra cứu ảnh được thể hiện như sau:

Cho biết một ảnh cần tra cứu P, hãy tra cứu ra một ảnh M trong cơ sở dữ liệu ảnh S thoả mãn:

$D(f(P),f(M))\le D(f(P),f(F))$ với mọi F thuộc S và F khác M.

Hiệu quả của hệ tra cứu ảnh phụ thuộc vào sự thể hiện các vector đặc trưng cụ thể và sự lựa chọn mô hình đánh giá độ tương quan.

1.1.2. Độ đo tương quan

Trong phần này trình bày một số cách nhìn nhận của con người về tính tương quan qua một số phân tích của các nhà tâm lý học, chúng ta cũng sẽ đề cập đến các ưu -khuyết điểm của một số cách tiếp cận nghiên cứu khác nhau. Các lý thuyết này được gom lại theo một khung nhìn thống nhất.

Các tiên đề về đo

Một số đo về độ tương quan đưa ra là để giải thích về tính tương quan được coi là khoảng cách trong không gian các đặc trưng đó, và nó được coi là không gian độ đo.

Nhận thức tính tương quan thông qua khoảng cách d, còn đo tính tương quan thông qua độ tương quan δ. Nếu A và B là đại diện cho các tác nhân a và b và $d(A,B)$ là khoảng cách mang tính tri giác giữa hai tác nhân đó thì đánh giá khoảng cách (đo) sẽ là:

$$\delta(A,B)=g[d(A,B)]$$

trong đó g là hàm không giảm đơn điệu thích hợp với tham số của nó. Lưu ý rằng chỉ có giá trị của δ là có thể sử dụng được cho thực nghiệm.

Các tác nhân được thể hiện ở đây chính là các điểm trong không gian độ đo, và $d(A,B)$ chính là hàm khoảng cách trong không gian này. Mô hình này thừa nhận rằng khoảng cách tri giác d thoả mãn tiên đề độ đo. Các ràng buộc mang tính kinh nghiệm được thay đổi thực tế theo một số nhà nghiên cứu.

Tiên đề thứ nhất đối với hàm khoảng cách đó là:

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

đối với mọi tác nhân (tính thống nhất của tự tương quan). Giả thiết này có thể được dùng đề kiểm tra thử việc đánh giá độ đo tương quan, nó ngụ ý $\delta(A,A) = \delta(B,B)$. Tiên đề nói nên tính thống nhất của tự tương quan.

Tiên đề thứ hai về mô hình khoảng cách là tính tối thiểu:

$$d(A,B) \geq d(A,A)$$

Lần nữa, giả thiết này đưa ra cách để kiểm tra thực nghiệm. Do tính quan hệ đơn điệu giữa d và δ, nó chỉ ra $\delta(A,B) \geq \delta(A,A)$. Tversky thì lại chỉ ra rằng giả giả thiết này vi phạm một số thực nghiệm về nhận dạng.

Tiên đề thứ ba nói rằng khoảng cách giữa các tác nhân có tính đối xứng:

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

Cũng như tiên đề trước, nó cũng mang hàm nghĩa kiểm tra thực nghiệm. Cũng ngụ ý rằng $\delta(A,B) = \delta(A,B)$. Một số nghiên cứu đã kiểm chứng giả thiết này bằng các thực nghiệm tương quan trực tiếp và quan sát tính không đối xứng của các ma trận hỗn độn. Hiện tượng này thường được quy kết cho tính khác biệt nổi trội hoặc mạnh mẽ của các tác nhân. Nói chung thì độ tương quan của tác nhân ít trội với tác nhân trội nhiều (nguyên bản hơn) lớn hơn so với độ tương quan của tác nhân trội nhiều so với tác nhân trội ít.

Tiên đề cuối cùng là bất đẳng thức tam giác:

$$d(A,B) + d(B,S_C) \geq d(A,S_C)$$

Về mặt trực quan rõ là tiên đề này là yếu nhất. Hàm tương quan giữa d và δ không đảm bảo được rằng đồng ý hay bác bỏ bất đẳng thức tam giác ứng với d sẽ chuyển sang tương quan cho ứng với δ.

Thứ tự trong quan hệ khoảng cách là bất biến với tất cả các biến đổi trong kiểu $\delta(A,B) = g[d(A,B)]$ nếu g là tăng đơn điệu. Một hệ quả của nó là bất đẳng thức tam giác không thể được thử mà chỉ dựa trên thứ tự đo. Tuy nhiên ít ra trong một số loại tác nhân, bất đẳng thức tam giác là không ổn.

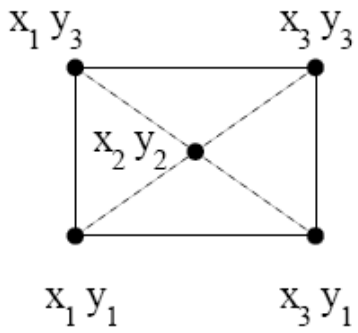
Tversky và Krantz chứng minh rằng nếu các tiên đề về khoảng cách đã được kiểm chứng và các khoảng cách là dương dọc theo các đường thẳng trong không gian đặc trưng, chẳng hạn d là khoảng cách Minkowski, nó có dạng:

$$d_p(A,B) = \left[\sum_i (A_i - B_i)^p\right]^{\frac{1}{p}}$$

trong đó $A = \{A_1,..., A_N\}, B = \{B_1,..., B_N\}$ và $p > 0$ là một hằng số đặc trưng cho hàm khoảng cách.

Tập các lý thuyết về độ tương quan

Trong một bài báo công bố năm 1977, Amos Tvesky giới thiệu mô hình độ tương phản đặc trưng rất nổi tiếng. Thay vào việc xem xét các tác nhân như là các điểm trong không gian độ đo, Tvesky đã đặc trưng hoá các tác nhân đó như là các tập đặc trưng nhị phân. Nói cách khác một tác nhân là một tập các đặc trưng của bản thân tác nhân đó. Nghĩa là, một tập đặc trưng là một tập hợp các dự đoán logic được xem là đúng đối với tác nhân đó. Gọi a,b là 2 tác nhân. A,B là 2 bộ đặc trưng tương ứng của chúng và $s(a,b)$ là mức độ giống nhau giữa a và b. Lý thuyết của Tvesky dựa trên các giả thuyết sau:



Hình 1.1. Bất đẳng thức tam giác

Hình 1.1. cho thấy đường dẫn từ góc $x_1y_1 \to x_3y_1 \to x_3y_3$ dài hơn từ $x_1y_1 \to x_2y_2 \to x_3y_3$ khi x_2y_2 nằm trong hình chữ nhật.

Tính ghép đúng:

$s(a,b) = F(A \cap B, A - B, B - A)$

Tính đơn điệu:

$s(a,b) > s(a,c)$ bất cứ khi nào mà $A \cap B \subseteq A \cap B, A - B \subseteq A - C, B - A \subseteq C - A$

Một hàm mà thoả tính ghép đúng và tính đơn điệu thì được gọi là hàm ghép đúng. Biểu thức $F(X,Y,Z)$ được xác định bất cứ khi nào có A,B mà $X = A \cap B, Y = A - B, Z = B - A$. Xác định $V \approx \mathrm{W}$ nếu tồn tại X,Y,Z mà ít nhất một trong các biểu thức sau đúng:

$F(V,Y,Z) = F(\mathrm{W},\mathrm{Y},\mathrm{Z})$

$F(X,V,Z) = F(Z,\mathrm{W},\mathrm{Z})$

$F(X,Y,V) = F(X,Y,\mathrm{W})$

Các cặp tác nhân (a,b) và (c,d) được gọi là đồng nhất trong một (hai, ba) thành phần khi một (tương ứng, hai, ba) những biểu thức sau đây đúng:

$(A \cap B) \approx (C \cap D)$

$(A - B) \approx (C - D)$

$(B - A) \approx (D - C)$

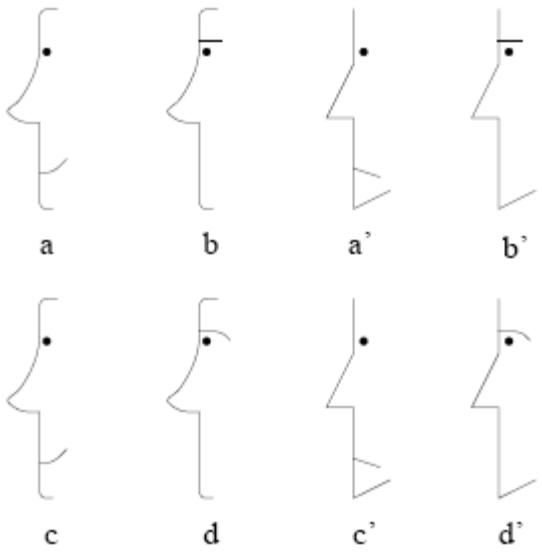
Dựa trên các định nghĩa này, Tvesky thừa nhận một thuộc tính thứ 3 của độ đo tương quan, tính độc lập.

Tính độc lập:

Giả sử cặp (a,b) và (c,d) tựa cho các cặp (a',b') và (c',d') đồng nhất trên 2 thành phần trong khi các cặp (a,b) và (a',b') tựa cho (c,d) và (c',d') đồng nhất trên thành phần thứ 3 còn lại, khi đó:

$s(a,b) \geq s(a',b') \Leftrightarrow s(c,d) \geq s(c',d')$

Một ví dụ về tính độc lập được cho ở 0. Trong trường hợp này thuộc tính độc lập cho thấy là nếu (a,b) là "gần đúng" hơn (c,d) thì (a',b') là "gần đúng" hơn (c',d'). Giả thuyết này (với một số các dữ kiện đã biết trước về sự lựa chọn các đặc trưng) có thể kiểm tra qua thực tế.



Hình 1.2. Ví dụ về tính độc lập

Hình 1.2. cho thấy, nếu a và b được coi là tương quan hơn so với a' và b' thì c và d sẽ tương quan hơn so với c' và d'.

1.2. Xác định độ đo nội dung ảnh

Để có thể đánh giá độ tương quan giữa các ảnh, trước hết phải xác định độ đo nội dung ảnh. Các đặc trưng được trích chọn để xác định nên nội dung cần phải lựa chọn giống như trong giai đoạn nhận thức chú tâm của con người. Cảm nhận của giác quan là rộng lớn, các thuộc tính có thể chia thành các mức thấp và cao. Các thuộc tính mức thấp bao gồm *màu sắc, cấu trúc, hình dạng, bố cục không gian*... còn ở mức cao - mức ngữ nghĩa đó là *khái niệm, từ khoá*. Việc chỉ sử dụng các đặc trưng mức thấp có thể không đưa ra được kết quả như ý muốn, do đó có thể đưa thêm ngữ nghĩa ở mức cao vào để nâng cao kết quả trong việc xác định nội dung ảnh. Ngữ nghĩa có thể hoặc là các ghi chú thủ công, hoặc được cấu trúc tự động từ các đặc trưng ở mức thấp. Trong phạm vi nghiên cứu này chúng ta chỉ tập trung vào việc xác định các độ đo dựa trên các đặc trưng ở dạng mức thấp của thuộc tính ảnh.

1.2.1. Độ đo thuộc tính màu sắc

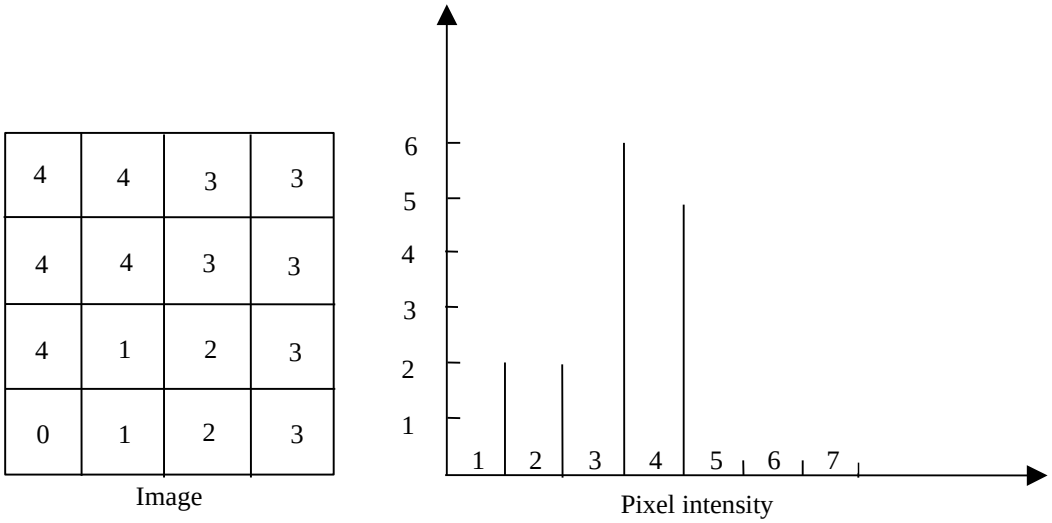
Màu sắc là một trong những thuộc tính trực quan quan trọng nhất trên một bức ảnh. Các hệ tra cứu ảnh nổi tiếng như QBIC, Netra, VisualSEEK cho thấy việc sử dụng thuộc tính màu sắc là tốt nhất. Nghiên cứu thuộc các ngành tâm lý học và nghệ thuật đều đã chứng tỏ rằng sự hiện diện và phân bố màu sắc đem lại cảm giác và mang chuyển ý nghĩa đến cho người quan sát một cách hiệu quả nhất.

Phương pháp phân tích màu phổ biến nhất là dùng Histogram. Histogram màu phản ánh sự phân bố mang tính thống kê, hay sự ghép nối xác suất các cường độ của 3 kênh màu. Histogram màu được tính toán qua việc thống kê rời rạc màu sắc trong phạm vi ảnh và đếm số lượng của các điểm ảnh theo mỗi màu. Nhân tố về màu sắc nhìn chung mô tả điểm ảnh trong không gian màu 3 chiều. Trước khi xây dựng nên histogram, không gian màu thường được định hướng chuyển đổi thành một số không gian màu đồng nhất chẳng hạn như hệ HSV (hue, saturation,value). Hue mô tả bước sóng thực của màu sắc nhận thức qua tri giác, saturation (độ bão hoà) chỉ ra lượng ánh sáng trắng trong sắc màu và value (giá trị) là độ sáng biểu thị cường độ của sắc màu.

1.2.1.1. Histogram

Histogram là thống kê về điểm ảnh, nó mô tả đặc trưng về mật độ của ảnh. Histogram cung cấp thông tin về độ tương phản và cường độ phát tán nói chung trên một bức ảnh. Nó đơn giản là một đồ thị thanh về mật độ các điểm ảnh. Cường độ sáng điểm ảnh được vẽ dọc theo trục x và số lượng các điểm ảnh cho mỗi giá trị cường độ sáng thể hiện trên trục y. 0 cho thấy một ví dụ về histogram của một bức ảnh đơn giản.

Ảnh sẫm màu có histogram phân tán lệch về phía trái, ảnh sáng hơn sẽ cho ta histogram phân tán lệch về phía phải. Một ảnh lý tưởng, sự phân bố điểm ảnh trên histogram là đồng đều. Một số thao tác trên Histogram vẫn thường dùng trong xử lý ảnh có thể được áp dụng cho việc xác định độ đo cho ảnh như sau.



Hình 1.3. Ảnh và histogram của ảnh

Histogram cân bằng

Histogram cân bằng hoá hay phép cân bằng histogram là một trong những xử lý quan trọng trong một phần mềm về bất cứ công việc xử lý ảnh nào. Để cải thiện độ tương phản tức là nhằm đạt được một histogram đồng đều cho ảnh chính là mục đích của việc cân bằng histogram. Kỹ thuật này có thể sử dụng trong toàn bộ bức ảnh hoặc có thể chỉ trên phạm vi một phần nhỏ của bức ảnh.

Cân bằng histogram sẽ không có nghĩa là "san bằng" ảnh. Nó chỉ có nghĩa là phân bố lại mật độ ảnh. Nếu như histogram của bất cứ một ảnh nào có nhiều các đỉnh và các vũng, nó sẽ vẫn giữ nguyên các đỉnh và vũng sau khi cân bằng, nhưng các đỉnh và vũng lúc này đã được dịch chuyển. Vì vậy, dùng khái niệm "dàn trải" sẽ là tốt hơn dùng khái niệm "san bằng" để mô tả cho cân bằng histogram.

Do cân bằng histogram là việc xử lý điểm, quá trình xử lý không đưa thêm giá trị cường độ mới vào trong ảnh. Các giá trị đã có sẽ được ghép với các giá trị mới nhưng số lượng thực sự về các giá trị cường độ trong bức ảnh kết quả sẽ vẫn bằng hoặc ít hơn so với số lượng về giá trị cường độ trong bức ảnh ban đầu.

Các thao tác chính khi cân bằng histogram:

+ Tính toán histogram

+ Tính toán chuẩn hoá tổng của các histogram

+ Biến đổi ảnh vào thành ảnh kết quả

Bước đầu tiên là đếm mỗi giá trị khác nhau của các điểm ảnh trong ảnh. Ta có thể bắt đầu với một mảng các số 0. Với điểm ảnh 8-bit, kích cỡ của một mảng là 256 phần tử (0-255). Phân tích ảnh và tăng mỗi phần tử mảng tương ứng với mỗi điểm ảnh khi xét đến.

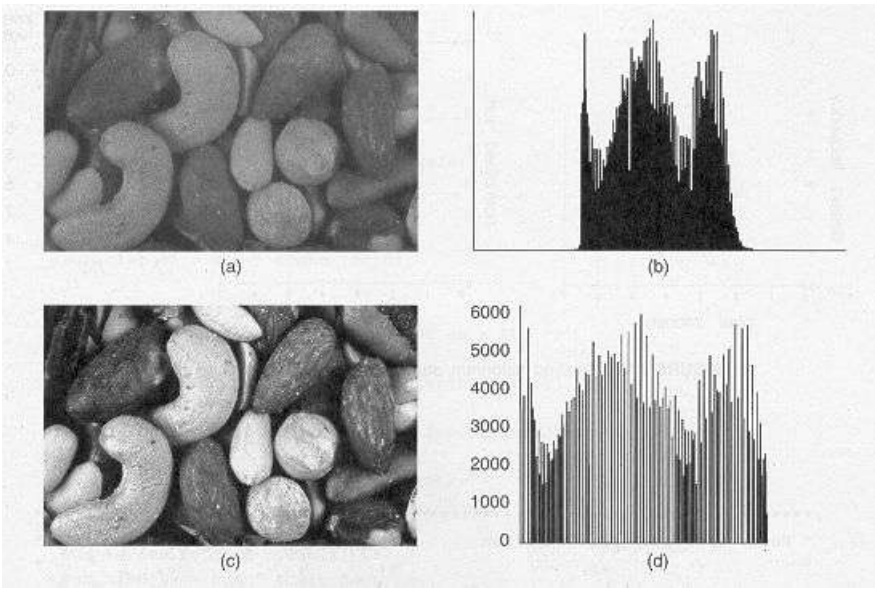
Bước thứ 2 đòi hỏi một mảng khác để lưu tổng của tất cả các giá trị histogram. Trong mảng này phần tử 1 sẽ lưu tổng histogram của các phần tử 1 và 0. Phần tử 255 sẽ lưu tổng histogram của các phần tử 255, 254, …, 1, 0. Mảng này sau đó sẽ được chuẩn hoá bằng cách nhân mỗi phần tử với giá trị tính:

= giá trị điểm ảnh lớn nhất/tổng số các điểm ảnh. (Chẳng hạn ảnh 8-bit cỡ 512x512 sẽ có giá trị đó là 255/262144).

Kết quả của bước 2 cho ta một ảnh LUT mà ta có thể sử dụng để chuyển dịch vào ảnh ban đầu.

0 cho thấy bước 2 và 3 của tiến trình xử lý và ảnh kết quả. Từ tổng chuẩn hoá trong 0(a) ta có thể xác định các giá trị ánh xạ bằng cách làm tròn tới giá trị nguyên gần nhất. 0 sẽ ghép với 0; 1 sẽ ghép với 1; 2 sẽ ghép với 2; 3 sẽ ghép với 5…

Cân bằng histogram cho phép các chi tiết rõ ràng hơn trong vùng tối. Trong một số trường hợp người ta thực hiện cân bằng histogram trên tất cả các bức ảnh trước khi tiến hành các thao tác xử lý ảnh khác. Thực tế đó không nên thực hiện việc làm này khi mà chất lượng của các bức ảnh đã cho là tốt. Sử dụng đúng lúc, cân bằng histogram có thể là một công cụ rất mạnh.



Hình 1.4. (a) Ảnh gốc; (b) Histogram gốc; (c) Ảnh cân bằng ; (d) Histogram cân bằng

Histogram đặc tả

Cân bằng hoá histogram xấp xỉ tới một histogram đồng nhất. Khi một histogram đồng nhất không mang lại kết quả như ý muốn, hoặc có lúc ta muốn làm sáng hoặc tối một ảnh hoặc ta cần một ảnh có độ tương phản tốt hơn. Thay đổi này có thể thực hiện thông qua histogram đặc tả.

Để có được Histogram đặc tả, người ta dùng các thông số đầu vào là histogram và ảnh gốc và nó được thực hiện theo 2 bước đơn giản như sau: Đầu tiên bức ảnh ban đầu được tính histogram cân bằng. Sau đó tính histogram nghịch đảo của histogram cân bằng.

Khi tính nghịch đảo của histogram cân bằng đòi hỏi phải sinh được ảnh LUT sau đó tính biến đổi nghịch đảo cho LUT. Biến đổi nghịch đảo được tính bằng cách phân tích các đầu ra của LUT. Đầu ra gần nhất đối với đầu vào cụ thể sẽ là giá trị nghịch đảo.

1.2.1.2. Moment màu

Môment màu được sử dụng rất thành công trong rất nhiều hệ tra cứu ảnh (chẳng hạn như hệ QBIC, đặc biệt khi mà ảnh bao gồm chỉ các đối tượng). Các thành phần của môment màu bao gồm:

+ thứ nhất : trung vị (mean)

+ thứ hai : biến trạng (variance)

+ thứ ba : độ lệch (skewness) của môment màu

Các thành phần này tỏ ra rất có hiệu quả trong việc thể hiện sự phân bố màu sắc của ảnh. Về mặt toán học, 3 thành phần môment đầu tiên được định nghĩa như sau:

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

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

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

trong đó f_{ij} là giá trị của thành phần màu thứ *i* của điểm ảnh *j*, và *N* là số các điểm ảnh của ảnh. Thông thường mômen màu thể hiện tốt hơn nếu nó được xác định bởi cả không gian L*u*v* và L*a*b*. Ngược lại không gian HSV chỉ đơn độc sử dụng mômen thành phần thứ ba có thể làm tăng hiệu năng tra cứu so với việc sử dụng chỉ mômen thành phần thứ nhất và thứ hai. Tuy nhiên, mômen thành phần thứ ba này đôi khi tạo ra các đặc trưng quá nhạy cảm đối với sự thay đổi cảnh quan và vì vậy làm giảm hiệu năng.

Do chỉ có 9 con số (3 mômen cho mỗi trong số 3 thành phần màu) được dùng để thể hiện màu cho ảnh nên các thành phần mômen thể hiện khá cô đọng so với các đặc trưng khác của màu. Do tính cô đọng như vậy, nó có thể hạ thấp hơn mức năng lượng nhận biết. Thông thường các mômen màu được dùng trong quá trình tiền nghiên cứu trước khi sử dụng các đặc tính phức tạp khác về màu sắc cho quá trình tra cứu.

1.2.1.3. Vectơ gắn kết màu

Đây là cách để tích hợp thêm các thông tin về tương quan không gian vào trong histogram màu (CCV – color coherence vectors). Mỗi bin màu trong histogram được phân đoạn thành 2, một phần là gắn kết (các điểm ảnh mang tính mạch lạc) – nó thuộc về một vùng màu rộng lớn đồng nhất, phần kia là không gắn kết (thiếu mạch lạc). Gọi α_i, β_i tương ứng là số các điểm ảnh gắn kết và không gắn kết trong bin màu *i* thì vector gắn kết màu CCV của ảnh sẽ là $\langle(\alpha_1,\beta_1),(\alpha_2,\beta_2),...,(\alpha_N,\beta_N)\rangle$ và nên nhớ rằng $\langle(\alpha_1+\beta_1),(\alpha_2+\beta_2),...,(\alpha_N+\beta_N)\rangle$ chính là histogram màu của ảnh.

Do tích hợp được thêm các thông tin về tương quan không gian nên vector gắn kết màu hiệu quả hơn so với histogram màu trong việc tra cứu ảnh, đặc biệt là đối với những ảnh có màu sắc hoặc cấu trúc chủ đạo (tính đồng nhất cao).

1.2.1.4. Tương quan màu

Tương quan màu được dùng để mô tả đặc trưng không chỉ về sự phân bố màu của các điểm ảnh mà nó còn mô tả tương quan không gian của từng cặp điểm ảnh. Thành phần thứ nhất và thứ hai trong histogram ba chiều mô tả màu sắc của mọi cặp điểm ảnh, thành phần thứ ba mô tả khoảng cách tương quan giữa chúng. Một vector tương quan màu là một bảng chỉ số về các cặp màu, trong đó số γ của cặp *(i,j)* là xác suất tìm một điểm màu *j* với khoảng cách *k* tới điểm màu *i* trong ảnh. Với

ảnh I, gọi $I_{c(i)}$ là tập các điểm ảnh của I có màu $c(i)$ thì tương quan màu được định nghĩa như sau:

$$\gamma_{i,j}^{(k)} = \Pr_{P1 \in I_{c(i)}, P2 \in P2 \in I} P2 \in I_{c(j)} \| P1 - P2 | = k$$

trong đó $i, j \in 1,2,..., N$, $k \in 1,2,..., d$ và $|P1-P2|$ là khoảng cách giữa điểm $P1$ và $P2$. Nếu tính cho kết hợp tất cả các cặp màu thì vector tương quan màu sẽ rất lớn ($O(N^2d)$). Trong thực tế người ta sử dụng phiên bản tương quan gọi là vector màu tương quan tự động, nó chỉ xét đến tương quan không gian của những điểm nhất định và do đó giảm được độ lớn xuống còn ($O(Nd)$).

So với histogram màu và vector gắn kết màu CCV thì vector tương quan màu cho kết quả tốt hơn trong tra cứu nhưng tính toán thì lại phức tạp hơn.

1.2.2. Độ đo thuộc tính hình dạng

Hình dạng của đối tượng là một ảnh nhị phân thể hiện hình dáng của đối tượng. Nhận thức thị giác và hiểu biết của con người về đối tượng tập trung nhiều vào thuộc tính hình dạng, các đặc trưng về hình dạng đóng một vai trò quan trọng trong việc nghiên cứu quá trình phân tích nội dung ảnh.

Trước kia người ta coi hình dạng bao gồm bộ của mảng hai chiều, mang đầy đủ thông tin vùng miền, sau này thể hiện hình dạng theo hình dáng đường nét ngoài (đường biên). Xác định các độ đo dựa trên hình dáng đường biên ngoài này thường cho các vector ngắn gọn hơn và do đó các thuật toán tìm kiếm sau đó cũng đơn giản hơn, tuy nhiên nếu xét về khía cạnh nội dung cũng như hiệu quả cho các phép tra cứu thì chúng không tốt bằng. Theo góc độ phân loại trên, việc xác định độ đo nội dung thuộc tính hình dạng sẽ tiếp cận theo hai hướng tương ứng là cơ sở vùng và cơ sở biên.

1.2.2.1. Cơ sở vùng

Đo các đặc tính hình học cơ bản

Việc mô tả các tính chất hình học của vùng có thể đưa ra được các độ đo tính chất các điểm ảnh thuộc phạm vi trong vùng xét đến. Các đặc tính đó là:

Vùng:

Có thể đo được bằng cách đếm số điểm trong vùng.

Đường viền chữ nhật:

Là hình chữ nhật nhỏ nhất gần đúng với đối tượng.

Tỉ lệ hình dạng:

Là sự thay đổ về tỉ lệ của đối tượng, do nó được tính là tỉ lệ giữa chiều rộng trên chiều dài của hình chữ nhật.

Độ tròn

Độ tròn còn gọi là dáng tròn, được định nghĩa:

$$Roundness = \frac{1}{Formfactor} = \frac{P^2}{4\pi A}$$

trong đó P là chu vi của đường biên và A là diện tích của vùng.

Độ cô đọng:

Tương quan độ tròn định nghĩa ở trên. Nó được định nghĩa như là tỉ lệ của chu vi hình tròn trên chu vi thực của đối tượng:

$$comp = \frac{P_{circle}}{P} = \frac{2\sqrt{A\pi}}{P}$$

Độ dãn dài:

Được định nghĩa như là tỉ lệ giữa bình phương chu vi với diện tích.

Độ lồi:

Một khoảng lồi là phần lồi tối thiểu bao bọc được đối tượng. Nó có thể được hình dung như là một miếng ruybăng co dãn được bao bọc vòng quanh đối tượng. Độ lồi có thể vì vậy được định nghĩa là tỉ lệ của đường kính của vỏ lồi trên chu vi thực của đối tượng:

$$conv = \frac{P_{convexhull}}{P_{contour}}$$

Tỉ số trục chính:

Các trục chính được định nghĩa là duy nhất là các đoạn thẳng giao vuông góc với nhau tại trung tâm của đối tượng. Độ dài của các trục chính là bằng eigenvalues $\lambda_{1,2}$ của ma trận đồng khả năng C.

Độ biến dạng tròn:

Mô tả hình dạng tròn đến mức độ nào. Tỉ lệ cân xứng hay sai số đối với hình tròn hay độ tròn được định nghĩa là:

$$c\,\mathrm{var} = \frac{1}{N\mu_r^2}\sum_i \left(\|\bar{p}_i - \mu\| - \mu_r\right)^2$$

trong đó $\mu_r = 1/N = \sum_i \|\bar{p}_i - \mu\|$ là bán kính trung bình, $p_i = (x_i, y_i)$ là điểm biên thứ i , $\bar{\mu}$ là tâm của vùng và N là số các điểm trên đường biên.

Độ elip:

Là mở rộng của độ biến dạng tròn, nó cho thấy độ kéo dãn dài của hình dạng, tức là hình elip điền khít có ma trận đồng khả năng *C* và đo sai số khớp *evar*:

$$e\,\mathrm{var} = \frac{1}{N\mu_{rc}}\sum_i \left(\sqrt{(\bar{p}_i - \mu)\mathbf{C}^{-1}(\bar{p}_i - \mu)} - \mu_{rc}\right)^2$$

trong đó $\mu_{rc} = 1/N\sum_i \sqrt{(\bar{p}_i - \mu)\mathbf{C}^{-1}(\bar{p}_i - \mu)}$.

Các thuộc tính hình học được sử dụng rộng rãi trong tra cứu ảnh. Các bộ mô tả đơn giản, chẳng hạn như là đo diện tích và độ lệch tâm, với trọng số hàm khoảng cách Euclidean được dùng trong QBIC. Các bộ mô tả hình dạng đơn giản rất mạnh trong việc xử lý nhiễu và thường cũng rất mạnh đối với việc thể hiện hướng, góc quay và tỉ lệ. Hơn nữa các thuộc tính hình dạng thường rất dễ tính toán và kết quả cho ta là các vector ngắn gọn. Tuy nhiên các bộ mô tả này thường không ổn định, do nhận biết về sự thay đổi không đáng kể trong hình dạng có thể dẫn đến kết quả là những thay đổi lớn về tín hiệu trong một số bộ mô tả.

Đo các mômen bất biến

Cho hàm 2D liên tục *f(x,y)*, mômen của (*p*+*q*) được định nghĩa:

$$m_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^q f(x,y)\,dx\,dy$$

Môments m_{pq} là duy nhất được định nghĩa theo hàm hình dạng *f(x,y)*, và môments m_{pq} là đủ để tái cấu trúc lại hàm miền ban đầu *f(x,y)*. Nói cách khác, mô tả hình dạng trên cơ sở môment chính là việc duy trì bảo tồn thông tin. Môment trung tâm được định nghĩa:

$$M_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-x_c)^p (y-y_c)^q f(x,y)dxdy$$

trong đó $x_c=M_{10}(R)/M_{00}(R)$ và $y_c=M_{01}(R)/M_{00}(R)$ xác định trung tâm của vùng lớn (trọng tâm) và *R* là vùng quan tâm đến.

Nếu f(x,y) là ảnh số , thì M_{pq} sẽ là:

$$M_{pq} = \sum_x \sum_y (x-x_c)^p (y-y_c)^q f(x,y)$$

Đó chính là yếu tố quan trọng để bộ mô tả hình dạng có thể là bất biến với tỉ lệ, chuyển vị và góc quay. Tuy nhiên một chuẩn hoá định nghĩa về môment là cần thiết:

$$\eta_{pq} = \frac{M_{pq}}{M_{00}^{\gamma}}, \quad where\ \gamma = (p+q)/2+1 \quad for\ p+q = 2,3,...$$

Một bộ môment 7 bất biến có thể được dẫn xuất ra từ các môment chuẩn hoá thứ tự thứ 2 và thứ 3 như sau:

$$\phi_1 = \eta_{20} + \eta_{02}$$

$$\phi_2 = (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})\left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right]$$

$$\phi_6 = (\eta_{20} + \eta_{02})\left[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$$

$$\phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})\left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right] + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right]$$

Các mômen này là bất biến với sự thay đổi về chuyển dịch, góc quay và tỉ lệ. Ưu điểm chính khác nữa là ảnh không cần phải được phân mảnh thứ tự nhằm mô tả hình dạng. Các mômen bất biến có thể có được bằng cách tích hợp trực tiếp từ giá trị cường độ thực sự của ảnh (f(x,y)). Do ưu thế đó của các mômen bất biến, chúng đã được dùng trong các hệ CBIR chẳng hạn như QBIC.

1.2.2.2. Cơ sở biên

Đường biên được hiểu là đường viền của đối tượng. Khi đi từ vùng miền của đối tượng này sang vùng miền của đối tượng khác sẽ gặp biên giữa hai đối tượng. Nhận diện hai đối tượng dựa trên sự khác biệt về độ sáng của màu sắc giữa hai đối tượng, tại ranh giới giữa hai đối tượng luôn có sự biến đổi đột ngột về độ sáng. Do đó, hầu hết các thuật toán dò biên sẽ đều phải dựa trên đặc điểm qui định đường biên này.

Đường biên thể hiện hình dạng đối tượng. Thông thường có 2 bước để thể hiện đường biên. Đầu tiên, một hàm 1D được xây dựng để tham số hóa đường viền của hình dạng 2D. Sau đó cấu trúc hàm 1D đó được dùng để trích ra vector các đặc trưng mô tả hình dạng của đối tượng.

Mã vòng (chain code)

Mã vòng được dùng để thể hiện đường biên bằng một thứ tự kết nối chiều dài và hướng của các đoạn thẳng. Thường thì nó được dựa trên 4 hoặc 8 liên kết của các đoạn.

Tạo ra các mã vòng dùng tất cả các cặp điểm ảnh sẽ dẫn đến hai bất lợi. Một là mã vòng đạt được sẽ dài, và thứ hai là sự nhiễu loạn trên đường biên có thể dẫn đến thay đổi mã. Tuy nhiên có một cách để tránh những vấn đề này là tái tạo lại đường biên qua việc lựa chọn không gian lưới lớn hơn.

Mã vòng của biên phụ thuộc vào điểm bắt đầu. Mã có thể chuẩn hoá dễ dàng bằng cách sử dụng một số hàm dưới đây. Mã vòng được coi như là chuỗi vòng các con số và điểm bắt đầu được xác định lại sao cho chuỗi thu được có biên độ nguyên tối thiểu. Tuy nhiên cách chuẩn hoá chỉ chính xác khi và chỉ khi đường biên là bất biến với sự thay đổi của góc quay và tỉ lệ.

Hàm mô tả Fourier (FD)

Đường biên của đối tượng có thể được biểu diễn như là thứ tự của các trục $u(k)=[x(k), y(k)]$, với $k = 0, 1, 2, \dots, K-1$. Hơn thế nữa mỗi cặp trục có thể được coi như là một bộ số phức để cho:

$$u(k) = x(k) + jy(k)$$

Hàm biến đổi rời rạc Fourier (DFT) của u(k) và nghịch đảo của nó được viết như sau:

$$F(n) = \sum_{k=0}^{K-1} u(k) \exp\left[\frac{-j2\pi nk}{K}\right] = M(n)e^{j\theta(k)}, \quad 0 \le n \le K-1$$

$$u(k) = \frac{1}{K} \sum_{n=0}^{K-1} F(n) \exp\left[\frac{j2\pi nk}{K}\right], \quad 0 \le n \le K-1$$

trong đó K là số các mẫu đường biên và M(n) là độ lớn của các ký hiệu mô tả Fourier.

Các hệ số phức F(n) được gọi là các hàm mô tả Fourier của đường biên. Thay tất cả F(n) và chỉ dùng hệ số đầu tiên M, dẫn đến kết quả sau:

$$\hat{u}(k) = \sum_{n=0}^{M-1} F(n) \exp\left[\frac{j2\pi nk}{K}\right], \quad 0 \le k \le K-1$$

Mặc dù chỉ có M là được dùng để tính được mỗi phần tử của *û(k), k* vẫn trong phạm vi từ 0 tới *K*-1. Điều đó có nghĩa là cùng với một số các điểm tồn tại trong biên xấp xỉ, nhưng có ít hơn các điểm cần cho việc tái cấu trúc cho mỗi điểm. Bởi thực tế các thành phần tần số cao tính toán cho các chi tiết nhỏ mịn và các thành phần tần số thấp xác định hình dáng tổng thể, M càng nhỏ thì các chi tiết càng khuyết hơn trên đường biên.

Ưu điểm chính của FD là ở chỗ nó dễ dàng ứng dụng, khá tốt đối với nhiễu và bất biến đối với các biến đổi hình học. Theo công thức trên, phương pháp mô tả Fourier không có hiệu quả trong tra cứu độ tương quan. Lý do có thể là do cảm nhận về tần số qua thị giác của con người là không được rõ ràng. Một nhược điểm khác với FD được thấy ở chỗ: các hàm cơ bản là các hàm sin tổng quát, nó có thể đưa ra sai số về vị trí ở các hệ số cụ thể nào đó, vấn đề này xảy ra khi thực hiện tra cứu đối với các ảnh có vùng che khuất.

Xấp xỉ đa giác

Như đã đề cập, thị giác của con người chia đối tượng bằng cách phát hiện các đường biên nơi có biến đổi cao về cường độ sáng. Việc xấp xỉ đường biên bởi các đường thẳng liên kết các điểm cong cao (HCP) đảm bảo đầy đủ các thông tin cần

thiết cho việc nhận dạng thành công hình dạng. Vì vậy việc xấp xỉ đa giác của đường biên tại các điểm cong cao thu được các thông tin hình dạng một cách hiệu quả trong một số lĩnh vực, nó có thể được ứng dụng trong kỹ thuật nhận dạng hình dạng dựa trên biến đổi Wavelet cực đại module.

1.2.3. Độ đo thuộc tính cấu trúc bề mặt

Mặc dù không có định nghĩa chính thức về cấu trúc bề mặt, nhưng có một số trực giác về đặc điểm của cấu trúc bề mặt, có thể tạm hiểu khái niệm này như là các biến đổi vùng của nền ảnh về cường độ mang tính lặp đi lặp lại và nhìn nhận chung đó là cấu trúc bề mặt. Cấu trúc bề mặt là thuộc tính vùng, định nghĩa nó phải bao hàm giá trị độ xám trong không gian kề cận. Kích cỡ của xung quanh phục thuộc vào kiểu của cấu trúc hoặc kích cỡ cơ sở xác định nên cấu trúc. Cấu trúc bề mặt bao gồm cả sự phân bố không gian của mức xám và do đó histogram 2D hoặc ma trận đồng khả năng đều có thể là các công cụ tốt để phân tích cấu trúc bề mặt. Có một số đặc tính, chẳng hạn như độ thô, độ tương phản, độ định hướng… đóng vai trò quan trọng trong việc mô tả cấu trúc. Độ đo độ thô (kích cỡ trung bình của vùng có cùng cường độ), độ đo độ tương phản (phụ thuộc vào sự biến thiên về histogram mức xám) và độ định hướng cho ta hướng chính của cấu trúc bức ảnh. Phân tích cấu trúc bề mặt là rất quan trọng bởi vì cấu trúc bề mặt là rất hữu ích trong các ứng dụng như kiểm duyệt tự động, xử lý ảnh trong y học, phán đoán từ xa, tự động dò tìm, đánh giá độ tương quan. Trong các nghiên cứu từ trước đến nay, người ta đưa ra một số đặc tính cho khái niệm cấu trúc trên cở sở phân chia thành các nhóm như: đặc tính về không gian, đặc tính về tần số, đặc tính trên cơ sở môment...

1.2.3.1. Các phương pháp không gian

Ma trận đồng khả năng

Ban đầu, ma trận đồng khả năng mức xám (GLCM) được Haralick giới thiệu cho phép ước lượng các thuộc tính ảnh liên quan đến các thống kê mức thứ cấp, nó tính đến việc sắp xếp không gian theo các mức xám cơ bản. Mỗi đầu vào (i,j) trong GLCM tương ứng với số các sự kiện của cặp mức xám mức i và j chính là khoảng cách d trong ảnh gốc. Các thống kê về xác suất cùng xảy ra được dùng để đặc trưng hoá các thuộc tính của vùng cấu trúc.

Hàm tương quan tự động

Một đặc tính quan trọng của cấu trúc bề mặt là tính lặp đi lặp lại tự nhiên của các phần tử cấu trúc. Hàm tương quan tự động của ảnh có thể được dùng để đánh giá chỉ số độ hạt và được coi là độ mịn và độ thô của bề mặt. Nếu như bề mặt là thô

thì hàm tương quan tự động sẽ giảm chậm theo khoảng cách; ngược lại nó sẽ giảm rất nhanh. Công thức về hàm tương quan tự động của ảnh I(x,y) được định nghĩa như sau:

$$\rho(x,y) = \frac{\sum_{u=0}^{M}\sum_{v=0}^{N} I(u,v)\, I(u+x, v+y)}{\sum_{u=0}^{M}\sum_{v=0}^{N} I^2(u,v)}$$

trong đó x,y là vị trí khác nhau trên các hướng u,v và M,N là kích thước của ảnh.

Phân mảnh

Kích cỡ phân mảnh (fractal) có thể được đo bởi độ nhám bề mặt. Trước tiên chúng ta định nghĩa một thuyết phân số nhằm giới thiệu một số khái niệm cơ bản. Tính tự tương quan theo tỷ lệ trong hình học được coi là một khái niệm chính. Một phân mảnh nguyên tố được xác định như sau: Nếu A được bao bọc trong không gian Euclidean n chiều, A được gọi là tự tương quan khi A là liên kết của N các mảnh khác biệt của chính nó, mỗi trong chúng được nén xuống với tỉ lệ r. Kích cỡ phân mảnh D liên quan đến N và tỉ lệ r:

$$D = \frac{\log N}{\log(1/r)}$$

Có một số phương pháp để ước lượng kích cỡ phân mảnh D. Hai phương pháp đưa ra ở đây mô tả như sau: Giả sử rằng ta đang tính kích cỡ phân mảnh của một ảnh A. Gọi P(m,L) là xác suất mà m điểm trong hình chữ nhật dài L trọng tâm tại một điểm bất kỳ trên bề mặt A. Gọi M là tổng số điểm ảnh của ảnh. Khi phủ ảnh bởi các hình vuông kích cỡ dài L thì (M/m)P(m,L) là số các hộp có m điểm bên trong. Số các hình hộp cần để phủ một ảnh là:

$$E[N(L)] = M\sum_{m=1}^{N}(1/m)P(m,L$$

Giá trị của N(L) là cân xứng với L^{-D} và do đó nó có thể được dùng để tính toán kích cỡ phân mảnh D. Tuy nhiên, kích cỡ phân mảnh tự nó không đủ để sao chụp tất cả các thuộc tính cấu trúc bề mặt. Bởi vậy người ta còn đưa ra một độ đo khác gọi là lacunarity để phân biệt giữa tính mịn và thô của cấu trúc có cùng kích cỡ phân mảnh.

1.2.3.2. Phương pháp tần số

Phổ năng lượng

Giải pháp cho phương pháp cơ sở tần số là phân chia ảnh thành tập các khối không chồng đè (*nxn* khối) sau đó tính toán phổ năng lượng cho từng khối này. Độ lớn cực đại của phổ có thể dùng để làm tham số cho mô hình các thuộc tính của cấu trúc. Mỗi mẫu hình có chu kỳ nhất định trong vùng không gian ban đầu được thể hiện bởi một đỉnh (peak) trong phổ năng lượng. Với các ảnh mà các mẫu không theo chu kỳ hoặc ngẫu nhiên thì việc xác định được đỉnh của phổ sẽ không đơn giản.

Phân tích hàm bước sóng Gabor có thể đồng thời xác định tiềm năng của cả phạm vi không gian và tần số. Việc giải mã bước sóng Gabor có thể đồng thời xác định tiềm năng của vùng không gian và vùng tần số. Cách xác định này cho thấy có thể tối ưu trong nhận thức về tính tối thiểu của liên kết hai chiều không chắc chắn trong không gian và tần số. Hàm Gabor được dùng là phần cơ bản trong chuẩn MPEG-7, nó sử dụng "Bộ mô tả duyệt qua dấu trúc" và "Bộ mô tả cấu trúc thuần nhất".

Như tâm lý học lôgic cho thấy, hệ trực quan của con người phân tích các ảnh cấu trúc theo kiểu giải mã ảnh thành các ảnh lọc, mỗi trong chúng có sự thay đổi về cường độ sáng khi qua các vùng tần số hẹp có độ định hướng thấp. Tuy nhiên phương pháp lọc đa kênh là xu hướng của trực giác bởi vì nó cho phép chúng ta khám phá tính định hướng và kích cỡ trội khác nhau. Bộ lọc Gabor đã được dùng trong một số ứng dụng phân tích ảnh như phân chia cấu trúc, dò tìm khuyết tật, nhận dạng khuôn mặt, giám sát máy móc và tra cứu ảnh.

Nghiên cứu thêm về hàm Gabor ta thấy, hàm Gabor là một hàm Gausian điều chỉnh số mũ phức tạp. Nói chung, một hàm Gabor g(x,y) dạng 2D và biến đổi Fourier G(u,v) của nó có thể được viết như sau:

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

$$G(u,v)=\exp\left\{-\frac{1}{2}\left[\frac{(u-W)^2}{\sigma_u^2}+\frac{v^2}{\sigma_v^2}\right]\right\}$$

trong đó W đại diện cho tần số của hàm Gabor. Hằng số không gian σ_x và σ_y xác định hình bao Gausian dọc theo trục *x* và *y*. Có thể xác định $\sigma_u = 1/(2\pi\sigma_x)$ và $\sigma_v = 1/(2\pi\sigma_y)$. Một lớp các hàm tự tương quan liên quan đến bước sóng Gabor đã được dùng trong việc tra cứu ảnh sẽ. Biểu thức tính g(x,y) trên gọi là bước sóng mẹ, kho bộ lọc tự tương quan có thể có được bằng cách giãn nở xấp xỉ và phép quay cho g(x,y) qua hàm sinh:

$$g_{mn}(x,y) = a^{-m} G(x', y'), \quad a > 1, \; m, n = \text{integer}$$

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

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

trong đó $\theta = n\pi/K$ và *K* là số các hướng. Thừa số a^{-m} cho thấy năng lượng không phụ thuộc vào *m*.

1.2.3.2. Phương pháp moment

Xác định độ đo qua phương pháp môment

Trong phương pháp cơ sở môment, các đặc trưng có được trực tiếp từ hàm mức xám *f(x,y)* qua bước tính các môment ảnh trong phạm vi các vùng cục bộ. Các thành phần môment thứ *(p+q)* của hàm 2 biến *f(x,y)* đối với thành phần gốc được xác định theo biểu thức (*). Gọi (*i*,*j*) là toạ độ điểm ảnh đã tính môment. Một cửa sổ độ rộng *W*, các chiều được chuẩn hoá trong phạm vi [-1,1] và toạ độ chuẩn hóa (x_m, y_n) được cho bởi:

$$x_m = \frac{m-i}{W/2}, \quad y_n = \frac{n-j}{W/2}.$$

Các môment trong phạm vi cửa sổ trung tâm tại điểm (*i*,*j*) được tính toán bởi xấp xỉ tổng rời rạc sử dụng toạ độ chuẩn (x_m, y_n):

$$x_m = \sum_{-W/2}^{W/2} \sum_{-W/2}^{W/2} f(m,n) x_m^p y_n^q$$

Do tính toán rời rạc của các tập môment đối với điểm ảnh đã cho trên cửa sổ hình chữ nhật xác định, phép tính đó tương ứng với toán tử lân cận và nó có có thể xem giống như việc nhân chập với một mặt lạ. Dưới đây là các mặt lạ tương ứng với các môment với kích cỡ cửa sổ là 3:

$$m_{00}=\begin{bmatrix}1&1&1\\1&1&1\\1&1&1\end{bmatrix}\quad m_{10}=\begin{bmatrix}-1&0&1\\-1&0&1\\-1&0&1\end{bmatrix}\quad m_{01}=\begin{bmatrix}-1&-1&-1\\0&0&0\\1&1&1\end{bmatrix}\quad m_{11}=\begin{bmatrix}1&0&-1\\0&0&0\\-1&0&1\end{bmatrix}$$

$$m_{12}=\begin{bmatrix}-1&0&1\\0&0&0\\1&0&-1\end{bmatrix}\quad m_{02}=\begin{bmatrix}1&1&1\\0&0&0\\1&1&1\end{bmatrix}\quad m_{20}=\begin{bmatrix}1&0&1\\1&0&1\\1&0&1\end{bmatrix}\quad m_{21}=\begin{bmatrix}-1&0&-1\\0&0&0\\1&0&1\end{bmatrix}$$

Trước khi tính toán các mômen, tỉ lệ cần phải được lựa chọn bằng cách chọn kích cỡ của cửa sổ. Nếu chọn kích cỡ của sổ càng lớn thì các đặc trưng được trích chọn càng tổng thể hơn. Ảnh với dấu hiệu cấu trúc rộng hơn sẽ đòi hỏi kích cỡ cửa sổ lớn hơn trong khi các cấu trúc mịn hơn sẽ có được từ các cửa sổ nhỏ hơn.

Tập các giá trị cho mỗi mômen trên ảnh đưa vào có thể được coi như một ảnh đặc trưng mới. Tuy nhiên chỉ các mômen không thôi thì không đủ để tạo ra được các đặc trưng cho một ảnh nhất định. Một số nghiên cứu đã đề xuất sử dụng một bộ chuyển đổi không tuyến tính để ghép các mômen với các đặc trưng cấu trúc. Chẳng hạn hàm lượng giác *tan* có thể được dùng cho việc chuyển đổi không tuyến tính, nó chuyển các ảnh mômen M_k với thành phần trung vị thành các ảnh đặc trưng cấu tương ứng. Phép chuyển đổi có thể viết như sau:

$$F_k(i,j)=\frac{1}{N}\sum_{(a,b)\in W_{i,j}}\left|\tanh\left(\sigma\left(M_k(a,b)-\overline{M}\right)\right)\right|$$

trong đó N là số lượng điểm ảnh trong cửa sổ $W_{i,j}$, (i,j) là trung tâm của cửa sổ và σ điều khiển hình dạng của hàm lôgíc.

Mômen Zernike

Sử dụng hàm cơ bản Zernike để thay thế hàm cơ bản không trực giao chúng ta sẽ có được cách xác định các mômen Zernike trực giao thành phần n và sự lặp lại l:

$$A_{nl}=\frac{n+1}{n}\iint_{D^2}f(x,y)V_{nl}*(x,y)dxdy$$

trong đó $V_{nl}(x,y)$ là hàm cơ bản Zernike của thành phần thứ n và sự lặp lại l:

$$V_{nl}(x,y)=V_{nl}(\rho\cos\theta,\rho\sin\theta)=R_{nl}(\rho)e^{il\theta}$$

trong đó (ρ,θ) là toạ độ cực của (x,y), n = 0, 1, 2, …, ∞, và l lấy các giá trị dương và âm tuỳ theo trạng thái.

Chương 2:

MỘT SỐ KỸ THUẬT ĐÁNH GIÁ ĐỘ TƯƠNG QUAN

Các mô hình độ tương quan về ảnh đều có thể nghiên cứu dựa trên bài toán tra cứu ảnh trong một tập ảnh cho trước. Bài toán yêu cầu đưa ra kết quả là bảng xếp hạng về độ đo tính tương quan của tập ảnh so với ảnh tra cứu. Với bài toán này, chúng ta sẽ tiến hành nghiên cứu một số kỹ thuật đánh giá độ tương quan dựa trên những mô hình tiêu biểu dưới đây.

2.1. Mô hình không gian vector VSM

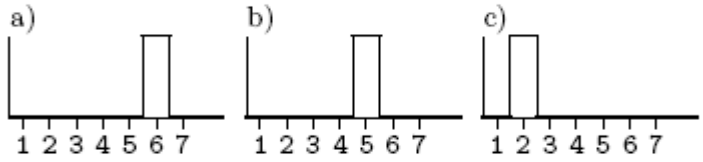
Trong mô hình này, nội dung mỗi ảnh được trích chọn thành vector đặc trưng, cơ sở dữ liệu ảnh là cơ sở dữ liệu lưu trữ các vector đặc trưng. Phần lớn các đặc trưng của ảnh đều có thể được xác định qua histogram (*color histogram, texture histogram, invariant feature histogram, local feature histogram, gabor feature histogram*). Trong mô hình không gian vector, các histogram được chuẩn hoá tức là các bin có giá trị tới hạn là 1. Từ trước đến nay, trong các bài toán tính toán độ tương quan giữa 2 ảnh dựa trên histogram, việc sử dụng chuẩn l_1 có dạng $\sum_{j=0}^{N-1}|h_1(j)-h_2(j)|$ trong đó phép tổng được tính trên tất cả các bin, luôn cho kết quả tốt. Vì dạng chuẩn l_1 được định nghĩa là không gian vector cho nên chúng được đặt cái tên là mô hình không gian vector (VSM).

2.1.1. Phép so sánh histogram

2.1.1.1. So sánh ngang các bin histogram (bin-by-bin)

Việc so sánh histogram (vector đặc trưng ảnh) theo kiểu bin-by-bin thường được tính toán nhanh chóng bởi phép so sánh chỉ tính đến độ lớn của bin mà không tính đến vị trí của bin.

Hình 2.1. cho thấy nếu so sánh kiểu bin-by-bin thì phép so sánh histogram của hình a) và b) so với của a) và c) là tương quan nhau, song về trực quan rõ ràng là độ tương quan của hình a) và b) lớn hơn so với a) và c) – (xét trên góc độ đây là histgram màu của ảnh xám chẳng hạn).



Hình 2.1. Độ tương quan giữa a) và b) phải lớn hơn giữa b) và c)

Độ tương quan qua histogram đồng nhất (histogram intersection)

Như ta đã biết, dựa vào không gian màu của ảnh, histogram của ảnh được xác định bằng cách tách biệt các màu sắc của ảnh theo các thành phần màu sắc (3 thành phần màu sắc trên hệ 3 chiều) và đếm số lượng có được từ ảnh của các màu này. Vì vậy, màu của bức ảnh được ánh xạ tới không gian màu gồm bộ n màu. Histogram màu của một ảnh I là một véc tơ n chiều $H_i(j)$, trong đó mỗi phần tử đại diện cho tần suất xuất hiện của màu j trong ảnh I.

Phép chỉ số hoá màu sắc cho phép nhận dạng ảnh hoặc các thành phần ảnh dựa trên sự phân bố histogram màu của các điểm ảnh, Swain và Ballard giới thiệu phương pháp đối sánh histogram và gọi là histogram intersection. Cho cặp histogram $H(I)$ và $H(I')$ của các ảnh I và I' tương ứng, mỗi trong chúng gồm n bin. Histogram intersection định nghĩa như sau:

$$H(I) \cap H(I') = \sum_{j=1}^{n} \min(H_j(I), H_j(I'))$$

Với hai ảnh, giá trị của histogram intersection càng lớn thì độ tương quan của hai ảnh càng cao.

Một số nghiên cứu thêm cho thấy rằng, hiệu quả của phương pháp histogram intersection chịu ảnh hưởng của các yếu tố như: hệ màu lựa chọn là gì và số lượng bin màu chọn là bao nhiêu.

Với lựa chọn hệ RGB, cách xây dựng công thức đánh giá độ tương quan của một cặp ảnh như sau:

Gọi I_R, I_G, I_B là các histogram màu chuẩn hoá của một ảnh trong cơ sở dữ liệu và tương ứng Q_R, Q_G, Q_B là các histogram màu chuẩn hoá của ảnh dùng để tra cứu. Độ tương quan giữa 2 ảnh trên là $SHI_C(I,Q)$ được tính theo công thức sau:

$$SHI_C(I,Q) = \frac{\sum_r \min(I_R(r), Q_R(r)) + \sum_g \min(I_G(g), Q_G(g)) + \sum_b \min(I_B(b), Q_B(b))}{\min(|I|, |Q|) * 3}$$

Lưu ý rằng giá trị đo độ tương quan trên nằm trong khoảng [0,1]. Nếu các histogram I và Q là như nhau thì rõ ràng giá trị trên là 1. Rõ ràng rằng, nếu một trong 2 ảnh trên nằm trong ảnh còn lại thì giá trị độ tương quan cũng sẽ là 1.

Độ tương quan qua khoảng cách Minkowski

Nếu mỗi chiều của vector đặc trưng của ảnh phụ thuộc vào chiều khác và cùng đóng vai trò quan trọng như nhau thì khoảng cách Minkowski sẽ xấp xỉ cho tính toán khoảng cách giữa hai ảnh. Khoảng cách này được định nghĩa là:

$$D(I,J) = (\sum_i |f_i(I) - f_i(J)|^P)^{1/P}$$

khi $p = 1,2$ và ∞ thì $D(I,J)$ là các khoảng cách L_1, L_2 (cũng còn gọi là khoảng cách Euclidean) và L_∞ tương ứng.

Độ tương quan qua khoảng cách Euclidean

Khoảng cách Euclidean được dùng như là độ đo tính toán khoảng cách giữa các vector đặc trưng, nó chính là dạng đặc biệt của khoảng cách Minkowski (*p=2*) thường được sử dụng phổ biến để tính toán độ tương quan giữa hai ảnh.

Mô tả phương pháp này, chúng ta lấy lại ví dụ tra cứu ảnh ở phần trước.

Gọi I_R, I_G, I_B là các histogram màu chuẩn hoá của một ảnh trong cơ sở dữ liệu và tương ứng Q_R, Q_G, Q_B là các histogram màu chuẩn hoá của ảnh dùng để tra cứu. Độ tương quan giữa 2 ảnh I và Q theo phương pháp Euclidean là $SED_C(I,Q)$ được tính theo công thức sau:

$$SED_C(I,Q) = 1.0 - \sqrt{\frac{\sum_r (I_R(r) - Q_R(r))^2 + \sum_g (I_G(g) - Q_G(g))^2 + \sum_b (I_B(b) - Q_B(b))^2}{2 * 3}}$$

Lưu ý rằng giá trị đo độ tương quan trên cũng nằm trong khoảng [0,1] và nếu các histogram I và Q là như nhau thì rõ ràng giá trị trên sẽ là 1.

2.1.1.2. So sánh chéo các bin histogram (cross-bin)

Độ tương quan qua khoảng cách toàn phương (QF)

Khoảng cách Minkowski coi tất cả các bin màu trong histogram hoàn toàn độc lập và thực tế không tính đến các cặp bin nhất định tương ứng với các đặc trưng mà được xác nhận là có tính tương quan hơn cặp bin khác. Khoảng cách toàn phương là cách để giải quyết vấn đề này, nó được định nghĩa như sau:

$$D(I,J) = \sqrt{(F_I - F_J)^T A (F_I - F_J)}$$

trong đó $A = [a_{i,j}]$ là ma trận độ tương quan, và $a_{i,j}$ biểu thị độ tương quan giữa bin i và bin j. F_I, F_J là các vector mà liệt kê tất cả các đầu vào dạng $f_I(I)$ và $f_i(J)$.

Khoảng cách toàn phương được dùng trong rất nhiều hệ tra cứu ảnh dựa trên việc tra cứu theo màu sắc, histogram. Nó có thể đưa ra kết quả về trực quan tốt hơn so với khoảng cách Euclidean và histogram intersection khi chúng ta tính đến độ tương quan về màu sắc.

Độ tương quan qua khoảng cách Mahalanobis

Tính toán đo khoảng cách Mahalanobis sẽ là thích hợp khi mỗi chiều trong vector đặc trưng của ảnh phụ thuộc vào các chiều khác và chiều đó là khá khác biệt. Nó được định nghĩa như sau:

$$D(I,J) = \sqrt{(F_I - F_J)^T C^{-1} (F_I - F_J)}$$

trong đó C là ma trận đồng khả năng của các vector đặc trưng.

Khoảng cách Mahalanobis có thể được làm đơn giản hơn nếu các chiều đặc trưng là độc lập. Trong trường hợp này, chỉ có một sự thay đổi của một thành phần đặc trưng C_i được tính đến:

$$D(I,J) = \sum_{i=1}^{N} (F_I - F_J)^2 / c_i$$

Độ tương quan qua khoảng cách Earth Movers EMD

Phương pháp EMD là phương pháp ước lượng độ khác biệt giữa hai phân bố đa chiều thuộc không gian các đặc trưng trong đó độ đo khoảng cách giữa các đặc

trưng đơn lẻ đã biết trước (còn gọi là khoảng cách nền). EMD "nâng" khoảng cách này từ các đặc trưng riêng biệt thành các phân bố đầy đủ.

Về mặt trực quan, cho trước hai phân bố, một có thể coi như một khối đất dàn trải đều trong không gian, phần kia như một loạt các lỗ hổng và chúng cùng trong một không gian. Gọi EMD là số đo khối lượng tối thiểu công việc để lấp đầy đất vào các lỗ hổng. Khi đó một đơn vị khối lượng công việc tương đương với việc dịch chuyển một đơn vị đất trên một đơn vị khoảng cách "nền".

Một phân bố có thể được coi là một tập các cụm mà mỗi cụm được đặc trưng bởi các tính chất (phương thức) của chính nó và bởi các phần tử của phân bố cùng cụm. Ta gọi đó là ký số của phân bố. Hai ký số có thể có kích cỡ khác nhau, các phân bố đơn giản có ký số ngắn hơn các phân bố phức tạp.

Tính toán EMD dựa trên cách giải quyết nổi tiếng cho bài toán vận chuyển. Giả sử có một số nhà cung cấp, mỗi nhà cung cấp có một lượng hàng nhất định, họ có nhiệm vụ phải cung cấp hàng cho các nhà tiêu dùng. Mỗi nhà tiêu dùng chỉ có thể chứa một lượng hàng nhất định. Mỗi cặp nhà cung cấp – tiêu dùng xác định được trước giá thành vận chuyển của một đơn vị hàng. Yêu cầu của bài toán ở đây là tìm ra luồng phân phối hàng có chi phí rẻ nhất từ các nhà cung cấp tới các nhà tiêu dùng đảm bảo đáp ứng được hết các nhu cầu tiêu dùng. Việc khớp đúng các ký số có ý nghĩa tương quan như bài toán vận chuyển hàng bằng cách coi một ký số là phía nhà cung cấp và một ký số khác là phía nhà tiêu dùng, sau đó đặt giá thành vận chuyển của một cặp cung cấp - tiêu dùng này bằng khoảng cách nền giữa một phần tử trong ký số thứ nhất và một phần tử trong ký số thứ hai. Rõ ràng rằng lời giải sẽ đưa ra kết quả khối lượng "việc" tối thiểu đòi hỏi đề vận chuyển hàng từ ký số vào ký số kia.

Ta có thể công thức hoá như sau:

Gọi $P = (p_1, w_{p1}),...,(p_m, w_{pm})$ là ký số thứ nhất với m cụm, trong đó p_i, w_{pi} tương ứng là cụm và trọng số của cụm i.

$Q = (q_1, w_{q1}),...,(q_n, w_{qn})$ là ký số thứ hai với n cụm và $D = [d_{ij}]$ là ma trận khoảng cách nền với d_{ij} là khoảng cách nền giữa cụm p_i và q_j. Ta cần xác định luồng $F = [f_{ij}]$ với f_{ij} là luồng giữa p_i và q_j có giá thành tối thiểu:

$$WORK(P,Q,F) = \sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} d_{ij}$$

thoả mãn các ràng buộc sau:

$f_{ij} \geq 0$ với $1 \leq i \leq m, 1 \leq j \leq n$

$\sum_{j=1}^{n} f_{ij} \leq w_{pi}$ với $1 \leq i \leq m$

$\sum_{i=1}^{m} f_{ij} \leq w_{qi}$ với $1 \leq j \leq n$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} fij = \min\left(\sum_{i=1}^{m} w_{pi}, \sum_{j=1}^{n} w_{qj}\right)$$

Ràng buộc thứ nhất cho phép chỉ có dịch chuyển "cung" từ *P* sang *Q* và không có chiều ngược lại. Hai ràng buộc tiếp theo giới hạn số lượng hàng có thể gửi từ các cụm trong *P* không thể lớn hơn trọng số của nó và các cụm trong *Q* cũng không thể nhận số hàng nhiều hơn trọng số của nó. Ràng buộc cuối cùng bắt buộc phải vận chuyển số hàng lớn nhất có thể. Ta gọi lượng hàng này là *tổng luồng* vận chuyển. Khi giải được bài toán ta sẽ tìm ra được luồng tối ưu *F*, khoảng cách EMD được xác định theo chuẩn hoá khối lượng công việc và theo tổng luồng:

$$EMD(P,Q) = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} d_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{n} f_{ij}}$$

2.1.1.3. Phép so sánh qua giá trị điểm ảnh

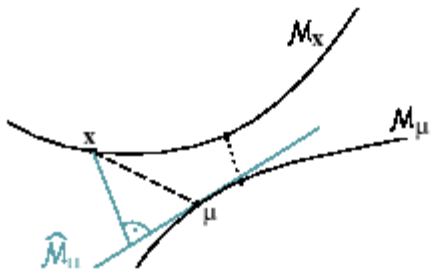
Trong nhận dạng đối tượng, đôi khi người ta dùng so sánh trực tiếp giá trị điểm ảnh trên các ảnh với nhau. Hiện nay phương pháp này đã được sử dụng trong một số hệ tra cứu ảnh.

Độ tương quan qua khoảng cách Euclidean

Hầu hết các tiếp cận theo hướng so sánh ảnh trực tiếp đều sử dụng khoảng cách Euclidean hoặc các khoảng cách thuộc nhóm khoảng cách Minkowski (như ta đã biết, khoảng cách Euclidean là trường hợp đặc biệt của khoảng cách Minkowski). Để có thể so sánh ảnh dùng khoảng cách Euclidean, trước hết sử dụng các thuật toán co dãn để chuẩn hoá kích cỡ cho ảnh. Có thể nhận thấy rằng đây chính là thao tác trừ ảnh có trong xử lý ảnh.

Độ tương quan qua khoảng cách đường tiếp tuyến

Các đối tượng trong ảnh thường bị tác động bởi một số các phép toán biến đổi như biến dạng, dịch chuyển, co dãn, phép xoay. Phương pháp khoảng cách Euclidean không thể dùng được khi tính tới các biến đổi này. Phương pháp khoảng cách đường tiếp tuyến được Keysers và Macherey đề xuất nhằm tổng hợp các bất biến đối với các phép biến đổi trên vào trong một hệ thống phân loại. Ở đây tính bất biến mang ý nghĩa là nó không làm thay đổi phân loại lớp của ảnh, tức là khi xét độ đo khoảng cách thì nó không được lớn.



Hình 2.2. Khoảng cách đường tiếp tuyến

Một ảnh $x \in R^{IxJ}$ bị biến đổi (chẳng hạn co dãn hoặc xoay) bởi một hàm biến đổi $t(x,\alpha)$ tuỳ thuộc vào không gian tham số L: $\alpha \in R^L$ (chẳng hạn như độ co dãn, hoặc góc xoay). Tập của tất cả các mẫu biến đổi theo không gian L bây giờ bao gồm các bản sao $\hat{M}_x = t(x,\alpha): \alpha \in R^L \subset R^{IxJ}$ trong không gian mẫu. Khoảng khác biệt giữa hai mẫu có thể được xác định là khoảng cách tối thiểu giữa các bản sao M_x của mẫu x và các bản sao M_μ của lớp mẫu μ, nó thực sự mang tính bất biến trên không gian biến đổi L:

$$d(x,\mu) = \min_{\alpha,\beta \in R^L} |t(x,\alpha) - t(\mu,\beta)|^2$$

Tuy nhiên việc tính toán khoảng cách này có độ phức tạp phi tuyến tính, hơn nữa các bản sao xét đến nói chung không có biểu thức giải tích mô tả. Để giải quyết vấn đề này, các bản sao có thể được xấp xỉ bởi một không gian con tiệm cận $\hat{M}$. Vector tiệm cận x_l tiệm cận không gian con và là đạo hàm từng phần của phép biến đổi $t(x,\alpha)$ theo tham số α_l, tức là $x_l = \partial t(x,\alpha)/\partial \alpha_l$. Do đó phép biến đổi $t(x,\alpha)$ có thể được xấp xỉ bởi hàm mở rộng Taylor xung quanh $\alpha = 0$:

$$t(x,\alpha) = x + \sum_{l=1}^{L} \alpha_l x_l + \sum_{l=1}^{L} \mathrm{O}(\alpha_l^2)$$

Chúng ta thu được thành phần xấp xỉ thứ nhất của Mx:

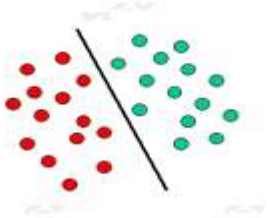
$$\hat{M}_x = \left\{ x + \sum_{l=1}^{L} \alpha_l x_l : \alpha \in R^L \right\} \in R^{\mathrm{IxJ}}$$

Việc tính toán trên xấp xỉ $\hat{M}$ có lợi là tính được khoảng cách theo phương pháp bình phương tối thiểu. Độ phức tạp tính toán của phép tính là chấp nhận được. Phép xấp xỉ là thoả với giá trị α là nhỏ.

2.2. Mô hình Vector

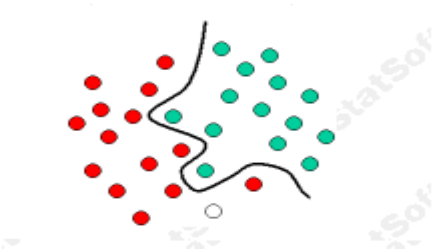
2.2.1. SVM

SVM (Support Vector Machine) được xây dựng dựa trên khái niệm mặt phẳng quyết định hay đường biên quyết định. Một mặt phẳng quyết định sẽ phân tách một tập các đối tượng bao gồm nhiều phần tử thành các lớp khác nhau. Trong ví dụ 0 các đối tượng sẽ thuộc hoặc lớp màu xanh (nhạt), hoặc lớp màu đỏ (đậm). Đường thẳng phân chia các đối tượng màu đỏ nằm bên trái, màu xanh nằm bên phải gọi là đường quyết định.



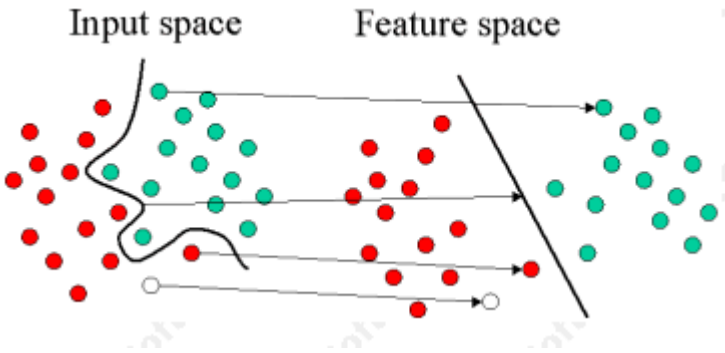
Hình 2.3. Phân lớp tuyến tính

Trên đây là một ví dụ điển hình cho việc phân lớp tuyến tính, tức là đường phân lớp có dạng đường thẳng. Tuy nhiên, phần lớn các bài toán lại không đơn giản như vậy, cấu trúc tập đối tượng phức tạp hơn (0) dẫn đến đường phân lớp phức tạp hơn và rõ ràng rằng trong trường hợp này đòi hỏi phải dùng đường có nhiều đoạn cong để phân lớp. Việc phân lớp bằng cách vẽ ra các đường phân chia tập đối tượng thành các lớp được hình dung như là các siêu phẳng phân lớp, khi đó SVM đặc biệt phù hợp cho việc giải quyết công việc này.



Hình 2.4. Phân lớp phi tuyến tính

Hình 2.4. diễn tả cho thấy ý tưởng của SVM. Phần bên trái là mô tả sơ đồ tập các đối tượng ban đầu (phía bên trái - không gian đầu vào). Sử dụng một số các hàm toán học, được biết như là các hàm nhân, để sắp xếp lại các đối tượng giống như phép ánh xạ hay còn gọi là phép biến đổi. Trong tập đối tượng mới, các đối tượng (phía bên phải) có thể được phân tách bằng phân lớp thẳng thay vì phải dùng phân lớp cong phức tạp (như phía bên trái). Phần công việc còn lại chỉ là tìm ra đường thẳng tối ưu để phân chia các đối tượng thành 2 lớp màu đỏ và màu xanh.



Hình 2.5. Ý tưởng cho mô hình SVM

Mô tả toán học của mô hình SVM như sau:

Ban đầu một tập dữ liệu không thể phân lớp tuyến tính được, biểu diễn dưới dạng vector x trong không gian R^n, giả sử tìm được ánh xạ phi tuyến tính ϕ từ không gian R^n vào không gian R^m, với $m>n$:

$$\phi : R^n \rightarrow R^m$$

Khi đó vector x_i trong không gian R^n sẽ tương ứng với vector $\phi(x_i)$ trong không gian R^m và điều cơ bản là trong không gian R^m này, tập các vector $\phi(x_i)$ có thể phân lớp tuyến tính được.

Thay các giá trị của x_i bởi $\phi(x_i)$ trong không gian R^m ta được bài toán OP2 (bài toán đối ngẫu), các tích vô hướng $x_i.x_j$ sẽ được thay thế bởi $\phi(x_i).\phi(x_j)$. Tuy

nhiên việc tính toán trực tiếp $\phi(x_i)$ là rất phức tạp, nhưng tích vô hướng $\phi(x_i).\phi(x_j)$ trong không gian R^m có thể tính được nếu tìm được hàm nhân (Kernel) $K(x_i, x_j)$:

$$K(x_i, x_j) = \phi(x_i).\phi(x_j)$$

Việc xác định hàm nhân *K* có một số điều kiện ràng buộc và việc lựa chọn nó như thế nào tất nhiên sẽ ảnh hưởng đến kết quả vector siêu phẳng thu được.

2.2.2. SVM trong kỹ thuật tra cứu ảnh

Xét bài toán phân lớp một tập các vector đặc trưng ảnh (mỗi vector đặc trưng là một ảnh) thành hai lớp:

$$(\vec{x}_i, y_i)_{i=1}^{N}, y_i = +1/-1$$

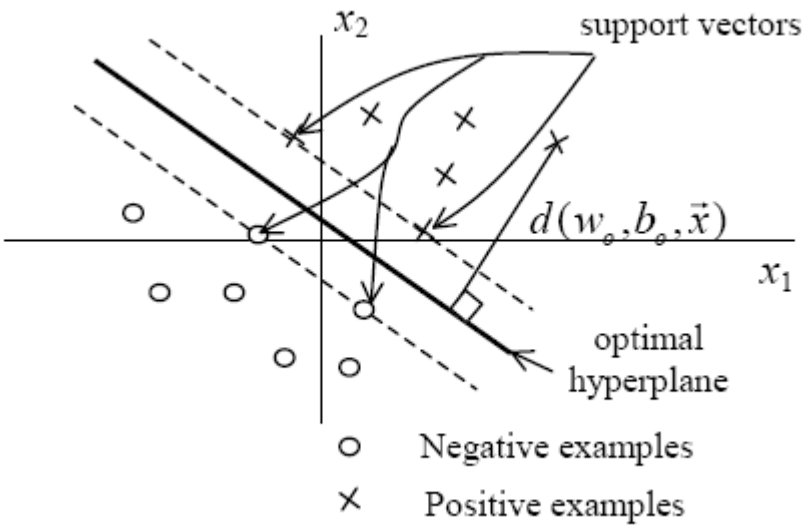
trong đó $\vec{x}_i$ là ảnh đầu vào thứ *i*, y_i là nhãn; +1, -1 tương ứng đại diện cho phân lớp mẫu dương và phân lớp mẫu âm. Nếu tập các vector mẫu đầu vào này có thể phân lớp tuyến tính thì siêu phẳng có thể xác định từ phương trình:

$$\vec{w}^T \vec{x} + b = 0$$

trong đó $\vec{w}$ là vector trọng số, *b* là hệ số nghiêng. Mục đích của SVM là xác định tham số w_0 và b_0 cho siêu phẳng để cực đại hóa khoảng cách giữa siêu phẳng với các phần tử gần nhất:

$$\vec{w}^T \vec{x} + b \geq 1 \text{ với } y_i = +1$$

$$\vec{w}^T \vec{x} + b < -1 \text{ với } y_i = -1$$



Hình 2.6. Tối ưu hoá siêu phẳng cho phân lớp tuyến tính

Hình 2.6. là một ví dụ về phân lớp tuyến tính trong không gian 2D (vector đặc trưng ảnh chỉ gồm hai thành phần). Đây là một bài toán tìm siêu phẳng tối ưu nên nó thường được giải quyết theo các mô hình mạng nơron. Dựa trên chức năng phân lớp, ta có thể ứng dụng SVM trong việc phân loại ảnh, điều đó hàm nghĩa đánh giá độ tương quan của ảnh. Khi áp dụng cho các ứng dụng ảnh, vector phần tử đầu vào cũng như các vector huấn luyện chính là các vector đặc trưng nội dung của mỗi bức ảnh.

Khoảng cách từ phần tử $\vec{x}$ đến siêu phẳng tối ưu được định nghĩa:

$$d(w_0, b_0, \vec{x}) = \frac{| w_0^T \vec{x} + b_0 |}{\| w_0 \|}$$

Khoảng cách trên là định lượng để đánh giá phần tử *x* thuộc về một phân lớp. Về mặt trực quan thấy rằng phần tử thuộc một phân lớp mà càng cách xa siêu phẳng thì chúng càng có độ khác biệt lớn so với phân lớp còn lại. Vì vậy, đối với các phần tử như thế thì khi huấn luyện ta nên chọn giá trị trọng số lớn hơn (miền giá trị của trọng số thường được chuẩn hoá, chẳng hạn từ 10-100). Thường thì quan hệ giữa khoảng cách trên đối với vector trọng số là hàm tuyến tính nhưng nó cũng có thể mở rộng ra thành quan hệ thành hàm phi tuyến.

Thực tế trong lĩnh vực về ảnh đã có nhiều ứng dụng dùng SVM trong việc đánh giá độ tương quan. Dưới đây là một ứng dụng tra cứu ảnh trong đó nội dung được trích chọn dựa trên các đặc tính màu sắc và kết cấu.

Hình 2.7. Tra cứu ảnh chỉ dựa trên phản hồi mẫu dương

Hình 2.7. chỉ ra 20 kết quả từ truy vấn một ảnh mang nội dung là bông hoa trên nền cây xanh, các kết quả được sắp xếp theo thứ tự từ trái sang phải, từ trên xuống dưới. Phương pháp này chỉ sử dụng cách truy cứu phản hồi mẫu dương, số lượng mẫu dương là 4 bao gồm các ảnh (1, 2, 3 và 6).

Hình 2.8. đưa ra kết quả của phép tra cứu dựa trên phương pháp sử dụng SVM. Vẫn sử dụng 4 mẫu dương trên, ngoài ra chọn thêm 4 mẫu âm mà một trong các mẫu đó là mẫu (19) ở 0, 3 mẫu còn lại không nhìn thấy trên hình. 2 mẫu dương (3 và 6) và 2 mẫu âm được lựa chọn là vector hỗ trợ (support vector), các trọng số sẽ được chọn dựa trên tính toán khoảng cách đầu ra của máy huấn luyện SVM. Mẫu ảnh (1) trong 0 có khoảng cách tới siêu phẳng lớn nhất, chúng được gán giá trị trọng số lớn nhất là 100. Các mẫu (3 và 6) trong 0 được lựa chọn là vector hỗ trợ nên được gán giá trị trọng số nhỏ nhất là 10. Kết quả cho thấy số lượng các ảnh có bông hoa trên nền cây xanh được tìm thấy là nhiều hơn.

Hình 2.8. Tra cứu ảnh dựa trên SVM

2.3. Mô hình k-phần tử kề cận (k-NN)

2.3.1. Thuật toán k-NN

Xét không gian *n* chiều R^n. Khái niệm *các phần tử kề cận* được mô tả trong chuẩn hình học Euclidean (khoảng cách giữa các điểm trong không gian *n* chiều). Cụ thể hơn, nếu *x* thuộc R^n thì nó sẽ được mô tả bởi vector đặc trưng $a_1(x), a_2(x), ..., a_n(x)$ trong đó $a_r(x)$ là giá trị của đặc trưng thứ *r* của *x*. Khi đó khoảng cách giữa hai điểm x_i và x_j được qua công thức:

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{n} a_r(x_i) - a_r(x_j)^2}$$

Trong cách học huấn luyện *các phần tử kề cận,* hàm mục tiêu có thể là rời rạc hoặc cũng có thể là liên tục. Hàm rời rạc sẽ có dạng $f: R^n \to V$ trong đó $V = v_1, v_2, ..., v_s$ là tập hữu hạn và R^n là không gian thực *n* chiều. Khi đó thuật toán *k-phần tử kề cận* hay viết tắt là thuật toán *k-NN* được mô tả như sau:

Huấn luyện:

Với mỗi mẫu $< x, f(x) >$, đưa mẫu vào tập ví dụ huấn luyện.

Phân lớp:

Cho phần tử x_q cần được phân lớp.

Bước 1: Xét danh sách $x_1.x_2,..., x_k$ là *k* phần tử thuộc tập ví dụ huấn luyện mà gần nhất với x_q.

Bước 2: Thực hiện

$$\hat{f}(x_q) \leftarrow \arg\max_{v \in V} \sum_{i=1}^{k} \delta(v, f(x_i))$$

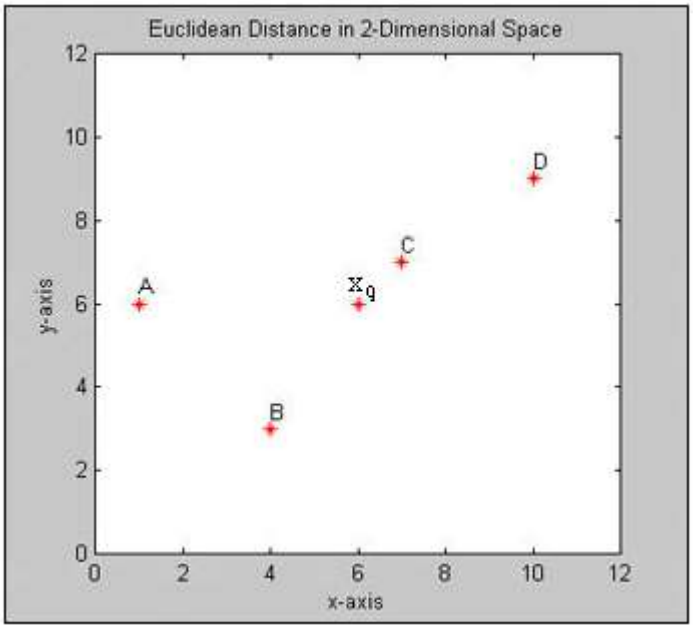
trong đó $\delta = \begin{cases} 1 & if(a = b) \\ 0 & if(a \neq b) \end{cases}$, (*argmax* là hàm cực đại)

Đối với hàm liên tục, thuật toán tương quan như vậy ngoại trừ ở bước 2 sẽ được thay thế bằng biểu thức:

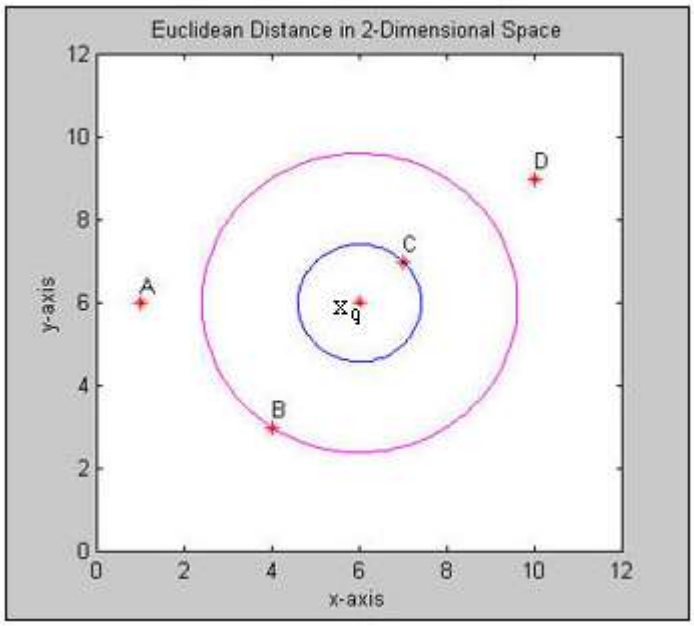
$$\hat{f}(x_q) \leftarrow \frac{\sum_{i=1}^{k} f(x_i)}{k}$$

0 là một ví dụ về tập không gian Euclidean 2 chiều, trong đó: *A=(1,6)*, *B=(4,3)* khi đó khoảng cách giữa *A* và *B* được xác định:

$$d(A,B) = \sqrt{(4-1)^2 + (3-6)^2} = 4{,}24$$



Hình 2.9 Khoảng cách Euclidean trong không gian 2D



Hình 2.10. *1-NN* và *2-NN*

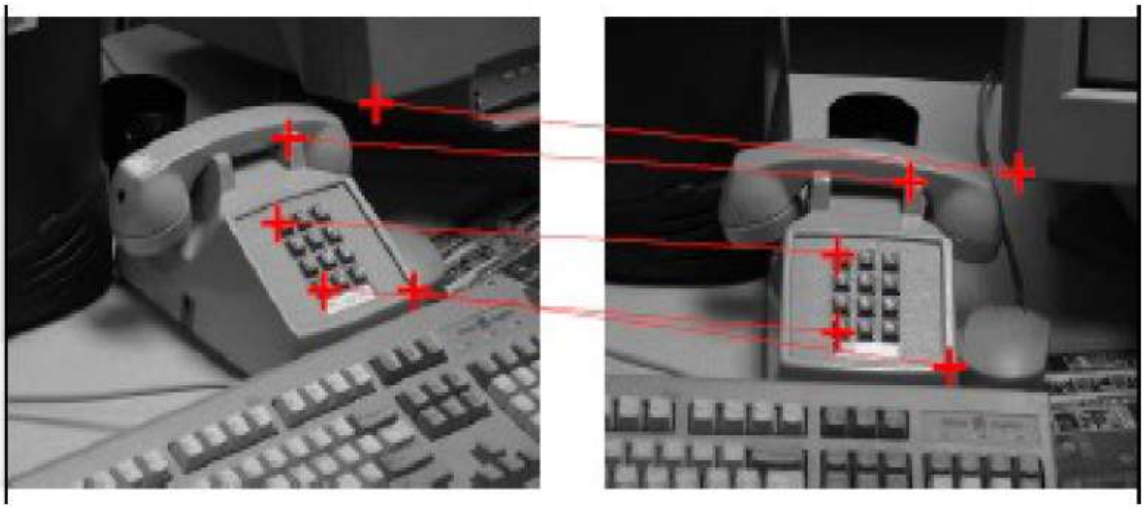
Dựa vào cách tính khoảng cách giữa 2 điểm thì việc phân lớp (tra cứu) phần tử x_q trong tập {*A, B, C, D*} theo thuật toán *k-NN* được mô tả trong 0. Với *k=1* ta có *1-NN* được biểu diễn bởi hình tròn nhỏ; *k=2* ta có *2-NN* được biểu diễn bằng hình tròn lớn (2 vòng tròn này đồng tâm tại x_q).

2.3.2. k-NN trong so khớp điểm ảnh

Sử dụng khoảng cách trọng số $k-NN$. Để xếp hạng cho ảnh i đã chúng ta xác định các ảnh này từ 2 bộ ảnh dương (P) và âm (N) mà theo đó k phần tử láng giềng kề cận của i (kề cận nhất được xác định theo chuẩn l_1). Sử dụng các phần tử láng giềng này chúng ta xác định độ khác biệt:

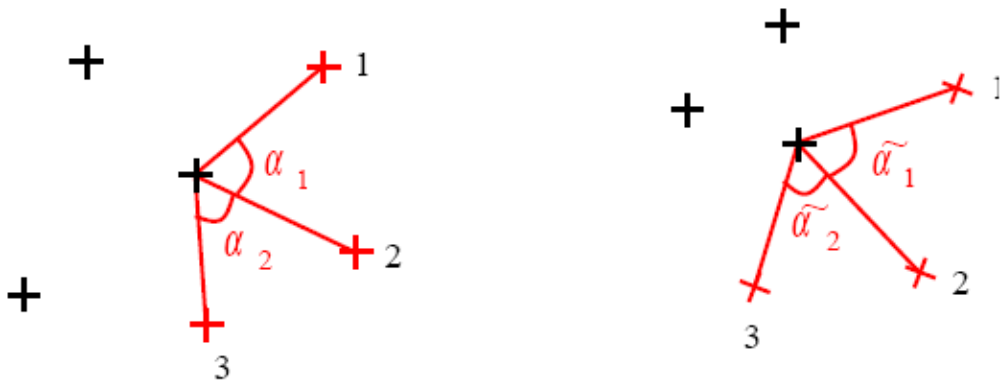
$$d(i) = \frac{\sum_{n \in N} dist^{-1}(i,n)}{\sum_{p \in P} dist^{-1}(i,P)}$$

Trong thực tế, người ta cộng một lượng nhỏ vào khoảng cách có trong công thức để tránh phép chia cho 0. Trong tất cả thực nghiệm số k được chọn là 5, số các ảnh tích cực và không tích cực tương ứng được chọn là 4 và 10.



Hình 2.11. So khớp các điểm "quan trọng" (matching)

Hãy xem Hình 2.11, chúng ta đã có được tập các điểm ảnh quan trọng, ý tưởng nảy sinh ở đây là cần thiết phải "so khớp các đỉnh góc tương quan từ một cặp ảnh cùng cảnh nhằm đạt được một thể hiện 3D của khung cảnh". Công việc này được thực hiện được bằng cách dựa vào tính ràng buộc cục bộ của các điểm ảnh lân cận. Cặp điểm sẽ "khớp" được với nhau nếu tất cả các điểm lân cận xung quanh nó từng cặp cũng khớp với nhau, đồng thời phải thoả thêm bộ điều kiện là có sự tương quan nhau từng cặp về khoảng cách và tỉ lệ khoảng cách (Hình 2.12).



Hình 2.12. Mô tả điều kiện có thể "ghép đúng" được theo 3 phần tử kề cận

Chương 3:

CHƯƠNG TRÌNH THỬ NGHIỆM

3.1. Bài toán

Trong bối cảnh kinh tế phát triển hiện nay, số lượng các doanh nghiệp, công ty ngày càng nhiều. Hàng ngày luôn có các hồ sơ xin đăng ký cấp giấy phép mở mới công ty và kèm theo đó là việc đăng ký lôgô cho hoạt động thương mại của công ty. Bài toán đặt ra là phải kiểm tra xem mẫu lôgô mà công ty đưa ra đăng ký có phải là một mẫu mới hay không, hay nó có gần giống (tương quan) với mẫu lôgô của một doanh nghiệp hay công ty nào đó đã đăng ký trước rồi hay không? Giả thiết là trước khi đi đến công đoạn kiểm tra tính tương quan này, lôgô đó đã được qua các khâu kiểm tra đảm bảo không vi phạm tính pháp lý, nói cách khác phạm vi của bài toán chỉ trong giới hạn đi sâu về mặt phân tích nội dung dựa trên các đặc trưng thuần tuý có trên bức ảnh lấy có trong các thuộc tính màu sắc, hình dạng, cấu trúc...

Mô tả một cách đầy đủ cho bài toán ở đây sẽ là "xây dựng một chương trình phần mềm quản lý việc đăng ký mới cho một mẫu lôgô thương mại bằng cách liệt kê và hiển thị ra màn hình khoảng 20 mẫu lôgô trong cơ sở dữ liệu về lôgô đã đăng ký trước, được sắp xếp theo thứ tự giảm dần về độ tương quan tính dựa trên nội dung trực quan của ảnh lôgô để người dùng kiểm tra lại lần cuối trước khi đi đến kết luận có cho đăng ký hay từ chối đối với mẫu lôgô đưa ra đó. Nếu việc đăng ký cho phép, lôgô mới được nhập vào cơ sở dữ liệu của các lôgô đã đăng ký."

3.2. Xây dựng chương trình

3.2.1. Lựa chọn môi trường

OS: Windows XP

Ngôn ngữ: Visual C++ 6.0

3.2.2. Phân tích về lôgô

Theo nhìn nhận trực quan về ảnh của lôgô nói chung thì 2 yếu tố thuộc tính là màu sắc và hình dạng mang nhiều thông tin hơn so với các thuộc tính khác như cấu

trúc bề mặt, tương quan không gian... Cả 2 thuộc tính này đề mang các thông tin dàn trải trên toàn bộ lôgô (tính tổng thể). Do đó việc trích chọn vector đặc trưng sẽ được ấn định là dùng bộ mô tả vector đặc trưng tổng thể (descriptor of global feature) dựa trên màu sắc và hình dạng.

Nội dung phong phú và đa dạng của lôgô đòi hỏi chương trình phải ổn định trước những biến đổi về góc xoay, tỉ lệ co giãn (các bất biến về hình dạng). Để đảm bảo cho các bất biến này thì phương pháp chung trong việc trích chọn màu sắc và hình dạng là dùng histogram. Ngoài ra kích cỡ của lôgô cũng cần được qui chuẩn trước, chẳng hạn sử dụng kích thước chuẩn chung là 256x256 điểm ảnh. Hệ màu sử dụng lựa chọn hệ RGB.

3.2.3. Đánh giá độ tương quan về lôgô

3.2.3.1. Trích chọn các đặc trưng cho lôgô

Trích chọn màu sắc theo histogram

Với kích thước của ảnh là 256x256 và bộ màu RGB (3 kênh màu) thì có thể lựa chọn số bin mỗi kênh màu là 16, tức là ta sẽ chia giải giá trị từ 0 – 255 thành 16 phân đoạn tương ứng với 16 bin, giá trị kênh màu thuộc phân đoạn nào thì sẽ thuộc bin đó.

Trích chọn hình dạng theo histogram

Để tính được histogram hình dạng của lôgô, trước tiên phải xác định biên cho các đối tượng trên ảnh lôgô. Hiện nay có rất nhiều các thuật toán dò biên mà mỗi thuật toán dựa trên các phương pháp riêng. Trước khi ảnh đưa vào dò biên người ta đều biến đổi ảnh về ảnh mức xám. Ở đây ta lựa chọn thuật toán Canny edge.

3.2.3.2. So sánh độ đo tương quan giữa các cặp lôgô

Có thể dùng một trong các phương pháp tính toán khoảng cách như Euclidean, Mahalanobis, đặc biệt là ở bước xây dựng vector đặc trưng chúng ta đã sử dụng phương pháp histogram trên các thuộc tính màu sắc và hình dạng thì cách tính toán độ tương quan theo histogram Euclidean đã giới thiệu trong chương trước sẽ là sự lựa chọn đầu tiên.

Lựa chọn 1- độ đo tương quan qua histogram intersection

Như đã giới thiệu trong chương trước, giả sử ta đã có các thành phần I_{R}, I_G, I_B là các histogram màu chuẩn hoá của một ảnh ảnh lôgô có trong cơ sở dữ liệu và

tương ứng ta phải tính các thành phần Q_R, Q_G, Q_B là các histogram màu chuẩn hoá của ảnh lôgô đưa vào kiểm tra. Độ tương quan giữa 2 ảnh lôgô là $SHI_C(I,Q)$ được tính theo công thức sau:

$$SHI_C(I,Q) = \frac{\sum_r \min(I_R(r), Q_R(r)) + \sum_g \min(I_G(g), Q_G(g)) + \sum_b \min(I_B(b), Q_B(b))}{\min(|I|, |Q|) * 3}$$

Các giá trị đo độ tương quan nằm trong khoảng [0,1].

Lựa chọn 2 - độ đo tương quan qua khoảng cách Euclidean

Tương quan theo phương pháp Euclidean độ tương quan giữa 2 ảnh lôgô là $SED_C(I,Q)$ được tính theo công thức sau:

$$SED_C(I,Q) = 1.0 - \sqrt{\frac{\sum_r (I_R(r) - Q_R(r))^2 + \sum_g (I_G(g) - Q_G(g))^2 + \sum_b (I_B(b) - Q_B(b))^2}{2 * 3}}$$

Và giá trị đo độ tương quan $SED_C(I,Q)$ cũng nằm trong khoảng [0,1]

3.2.4. Một số kết quả

Lựa chọn các tham số

Vào menu Image Retrieval -> Method's Option (0)

Các tham số lựa chọn bao gồm:

+ Similarity Method: Lựa chọn phương pháp đo độ tương quan cho màu sắc và hình dạng.

+ Các lựa chọn cho tra cứu theo: By Color (màu sắc), By Shape (hình dạng), By Both (màu sắc và hình dạng)

+ Top N Result: N số kết quả hiển thị (có độ tương quan lớn nhất)

+ Kích cỡ ảnh chuẩn: Widht (rộng), Height (cao)

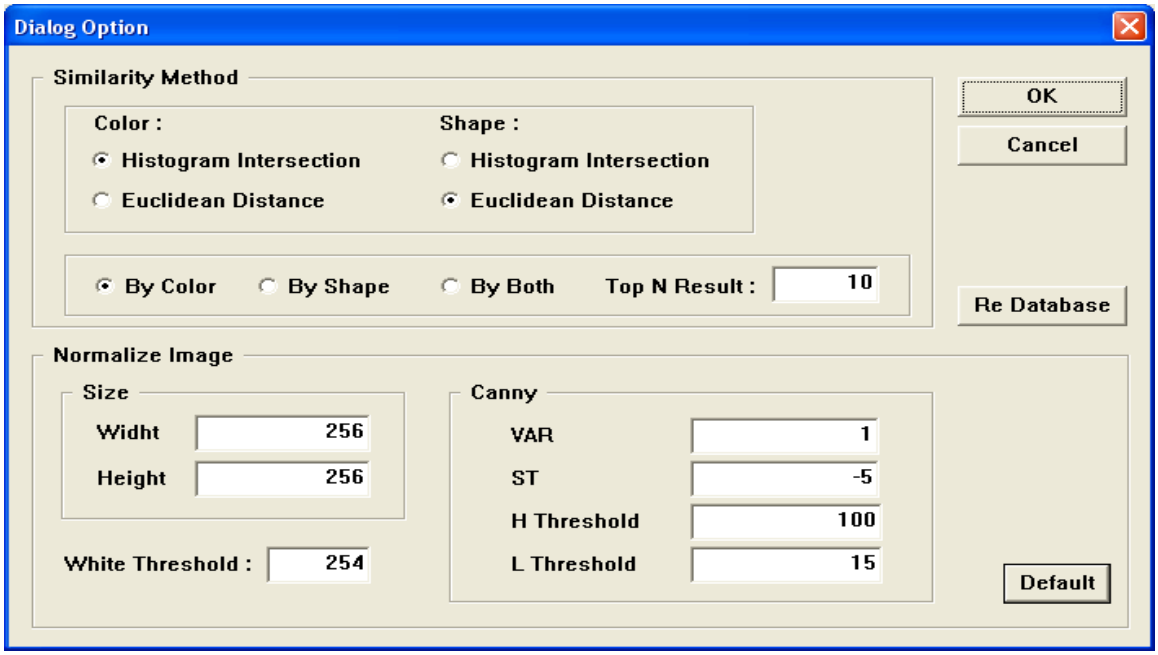
+ Threshold: giá trị ngưỡng cho phép biến đổi ảnh xám

+ Canny: các tham số cho việc thực hiện dò biên theo thuật toán Canny (xem phần thuật toán Canny)

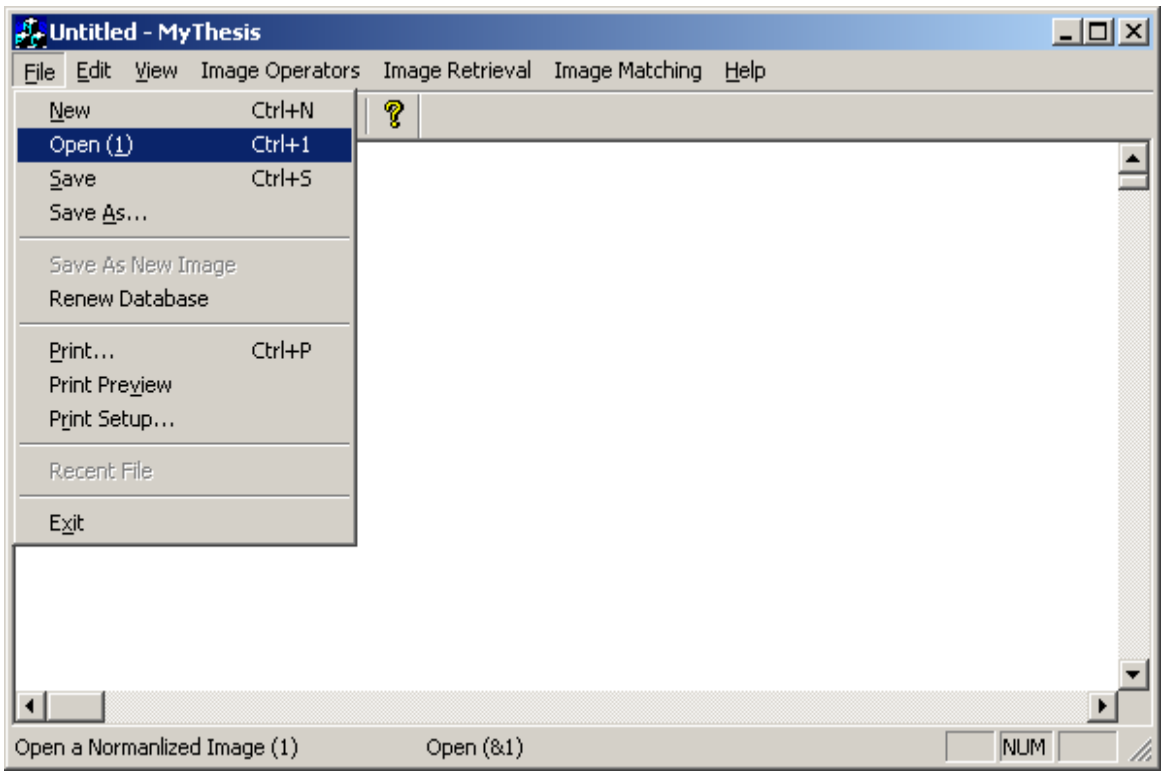
Cách sử dụng chương trình và một số kết quả

Thứ tự trình bày các màn hình dưới đây chỉ rõ các bước sử dụng chương trình và một số màn hình kết quả.

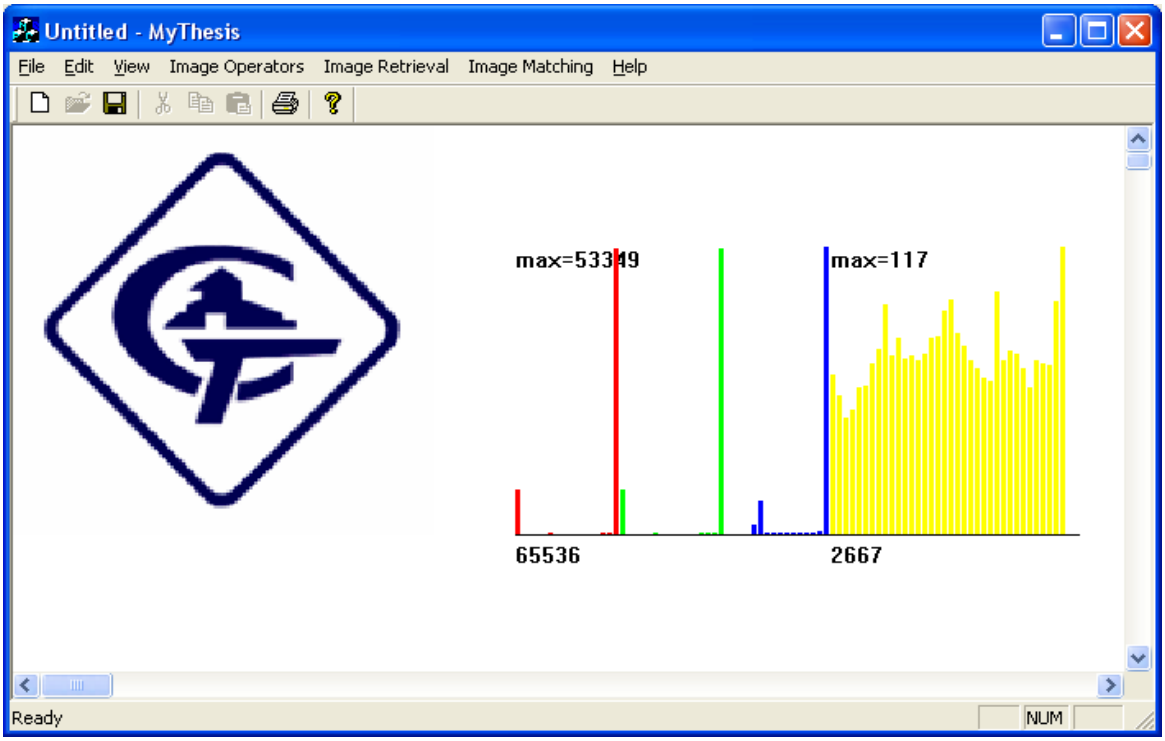
Vào menu Image Retrieval -> Method's Option



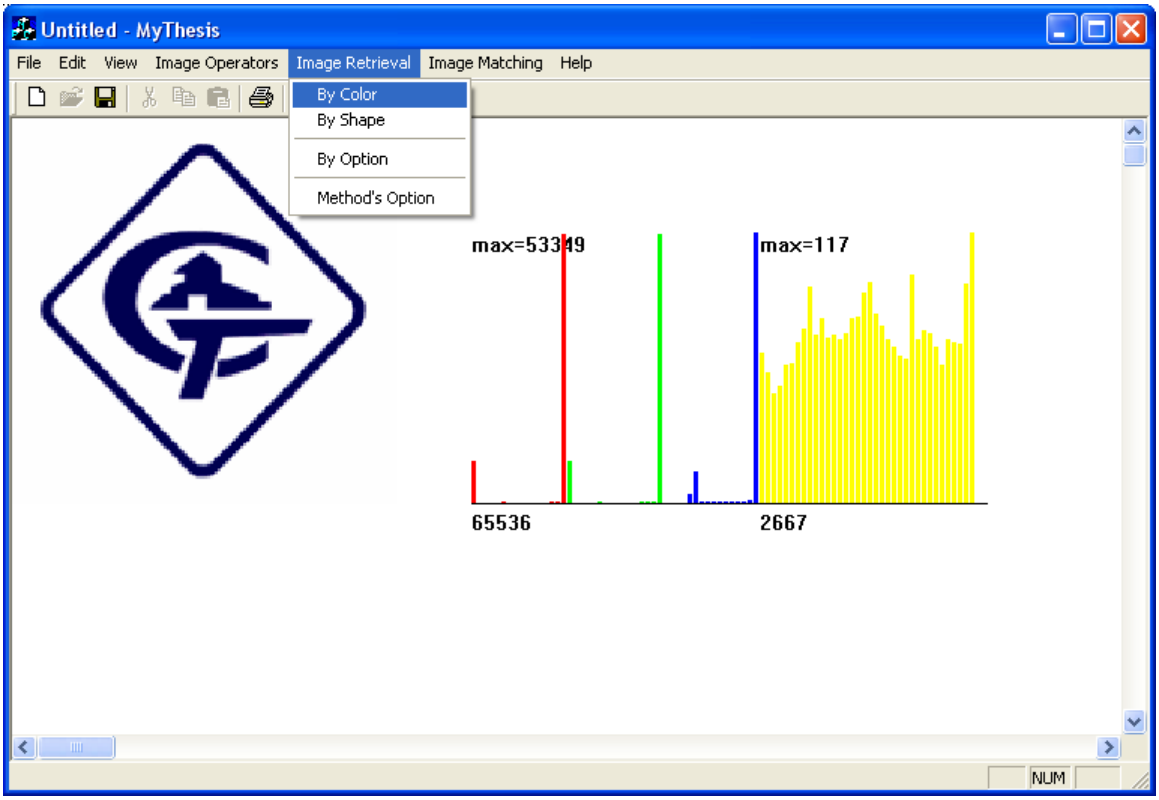
Hình 3.1. Màn hình lựa chọn các tham số cho chương trình



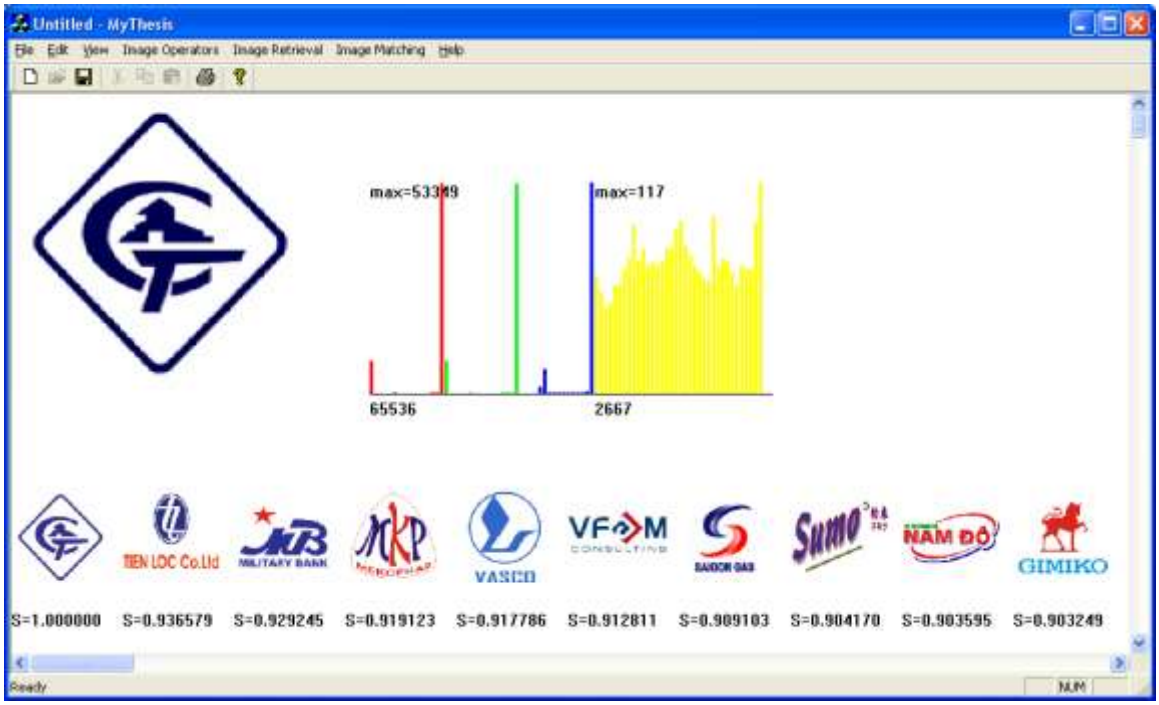
Hình 3.2. Màn hình lựa chọn menu mở file ảnh lôgô đưa vào tra cứu



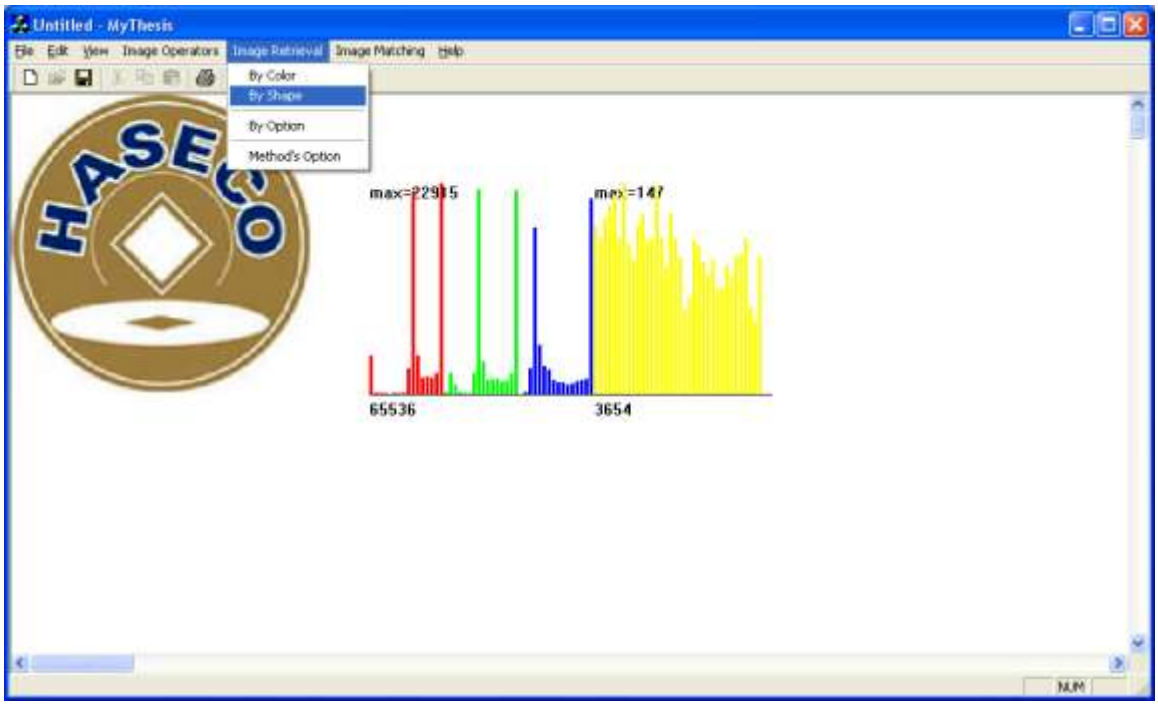
Hình 3.3. Màn hình hiển thị ảnh lôgô đưa vào cùng histogram màu và hình dạng



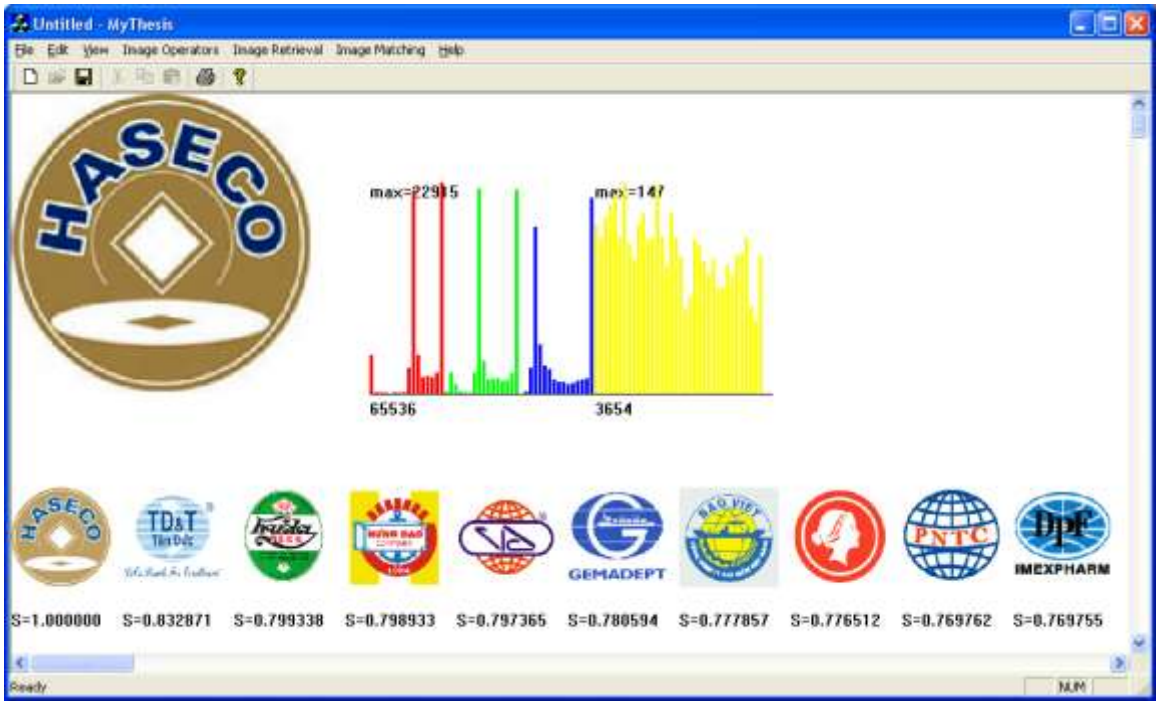
Hình 3.4. Lựa chọn menu Image Retrieval -> By Color (tra cứu theo màu sắc)



Hình 3.5. Màn hình hiển thị kết quả 10 ảnh lôgô gần giống nhất theo màu sắc



Hình 3..6. Lựa chọn menu Image Retrieval -> By Shape để tra cứu theo hình dạng



Hình 3.7. Màn hình hiển thị kết quả 10 ảnh lôgô gần giống nhất theo hình dạng

PHẦN KẾT LUẬN

Mô tả nội dung của ảnh đúng như trực quan cảm nhận của con người thông qua tri giác thực sự là mong muốn của ngành khoa học máy tính hiện nay. Đây là một chủ đề lớn và tương đối rộng, tất cả các khía cạnh của nó đều đã được quan tâm nghiên cứu trong các dự án, các công trình khoa học. Với cùng mối quan tâm như trên, nội dung bản đồ án đã nhấn mạnh đến 2 vấn đề chính như sau:

Một là: tìm hiểu về các phương pháp trích chọn đặc tính dựa trên các thuộc tính của ảnh bao gồm màu sắc, hình dạng, cấu trúc… làm cơ sở để mô tả nội dung của ảnh. Tiêu chí chung trong cách trích chọn các đặc tính là bám sát vào thực tế nhận thức tri gác của con người đối với đối tượng ảnh. Kết quả của phép trích chọn này được mô tả dưới dạng các vector đặc trưng ngắn gọn phù hợp với việc lưu trữ - khai thác thông tin đứng trên góc độ quản trị một cơ sở dữ liệu về hình ảnh.

Hai là: tìm hiểu và giới thiệu một số mô hình đánh giá độ tương quan hiện đã và đang được giới khoa học nghiên cứu ứng dụng. Nhìn chung về bản chất các kỹ thuật đánh giá - so sánh độ tương quan hay độ khác biệt đều dựa trên các hàm đánh giá ước lượng về "khoảng cách" giữa 2 vector đặc trưng mô tả nội dung của 2 bức ảnh. Tuy nhiên, cũng như nhiều vấn đề khác, để đưa vào ứng dụng thực tế thì không có phương pháp nào là tuyệt đối.

Trên cơ sở các kiến thức tìm hiểu được em đã phát triển một chương trình ứng dụng để tra cứu quản lý ảnh lôgô thương mại, tuy rằng hiện mới chỉ dừng lại ở góc độ trình diễn cho các nghiên cứu đã tiến hành, song cũng mong muốn nếu có điều kiện phát triển, có thể trở thành sản phẩm thực sự mang tính thương mại.

TÀI LIỆU THAM KHẢO

[1]. Đỗ Năng Toàn, Phạm Việt Bình (2008), *Giáo trình xử lý ảnh* – ĐH Thái Nguyên, Nhà xuất bản KH&KT, 2008.

[2]. Mari Partio, "*Content-based Image Retrieval using Shape and Texture Attributes*", Master of Science Thesis,Tampere University of Technology, April 2002.

[3]. S. M. Lee, J. H. Xin, S. Westland, "*Evaluation of Image Similarity by Histogram Intersection*", Hong Kong Polytechnic University, January 2005.

[4]. Konstantinos G. Derpanis, "*The Harris Corner Detector*", kosta@cs.yorku.ca, October 2004.

[5]. Anil K. Jain, Aditaya Vailaya, "*Image retrieval using color and shape*", Department of Computer Science Michigan State University East Lansing, May 1995.

[6]. V. Gouet N. Boujemaa, "*Object-based queries using color points of interest*", Valerie.Gouet@inria.fr, Nozha.Boujemaa@inria.fr, April 2001.

[7]. Chaur-Chin Chen, Hsueh-Ting Chu, "*Similarity Measurement Between Images*", Department of Computer Science National, Tsing Hua University, Taiwan, 2003.

[8]. Norio Katayama and Shin'ichi Satoh, "*Similarity Image Retrieval with Significane-Sensitive Nearest-Neighbor Search*", National Institute of Informatics, 2004.

[9]. Pengyu Hong, Qi Tian, Thomas S. Huang, "*Incorporate Support Vector Machine to Content-Based Image Retrieval With Relevant Feedback*", IFP Group, Beckman Institute - University of Illinois at Urbana-Champaign, Urbana, USA, 2000.

[10]. Kerry Rodden, Wojciech Basalaj,David Sinclair, Kenneth Wood, "*A comparison of measures for visualising image similarity*", University of Cambridge Computer Laboratory.

[11]. Pinar Duygulu, "*Interest Points*", Computer Vision, Bilkent University, Spring 2006.