

LỜI CẢM ƠN

Trước hết em xin bày tỏ lòng biết ơn sâu sắc tới thầy giáo hướng dẫn Ths. Ngô Trường Giang đã tận tình giúp đỡ em rất nhiều trong suốt quá trình tìm hiểu nghiên cứu và hoàn thành báo cáo tốt nghiệp.

Em xin chân thành cảm ơn các thầy cô trong Khoa Công nghệ Thông tin cũng như các thầy cô trong trường đã trang bị cho em những kiến thức cơ bản cần thiết để em có thể hoàn thành đồ án.

Xin gửi lời cảm ơn đến bạn bè những người luôn bên em đã động viên và tạo điều kiện thuận lợi cho em, tận tình giúp đỡ chỉ bảo em những gì em còn thiếu sót trong quá trình làm báo cáo tốt nghiệp.

Cuối cùng em xin bày tỏ lòng biết ơn sâu sắc tới những người thân trong gia đình đã giành cho em sự quan tâm đặc biệt và luôn động viên em.

Vì thời gian có hạn, trình độ hiểu biết của bản thân còn nhiều hạn chế. Cho nên trong đồ án không tránh khỏi những thiếu sót, em rất mong nhận được sự đóng góp ý kiến của tất cả các thầy cô giáo cũng như các bạn bè để đồ án của em được hoàn thiện hơn.

Em xin chân thành cảm ơn!

MỤC LỤC

LỜI CẢM ƠN	1
MỞ ĐẦU	4
CHƯƠNG 1: Tổng quan về tra cứu ảnh dựa trên nội dung.....	5
1.1 Tra cứu thông tin trực quan	5
1.2 Những thành phần cơ bản của 1 hệ thống tra cứu ảnh dựa trên nội dung	5
1.3 Các phương pháp tra cứu ảnh dựa trên nội dung.....	6
1.3.1 Tra cứu ảnh dựa trên màu sắc	6
1.3.1.1 Biểu đồ màu toàn cục (Global Color Histogram)	7
1.3.1.2 Biểu đồ màu cục bộ (Local Color Histogram)	8
1.3.1.3 Biểu đồ màu tương quan (Color Correlogram Histogram).....	8
1.3.1.4 Vector liên kết màu (Color Cohenrence Vector)	8
1.3.1.5 Tương quan màu (Color Correlogram).....	9
1.3.1.6 Độ đo tương đồng về màu sắc	9
1.3.2 Tra cứu ảnh dựa trên kết cấu.....	10
1.3.2.1 Phương pháp ma trận đồng nhất mức xám (Gray-Level Co-occurrence Matrices).....	11
1.3.2.2 Phương pháp Gray-Level Difference (GLD).....	12
1.3.2.3 Độ đo tương đồng cho kết cấu ảnh.....	13
1.3.3 Tra cứu ảnh dựa trên hình dạng.....	13
1.3.3.1 Phương pháp trích chọn đặc trưng dựa trên đường biên.	13
1.3.3.2 Phương pháp trích chọn đặc trưng dựa trên vùng.	16
1.3.3.3 Các phương pháp đối sánh dựa trên hình dạng	17
1.3.4 Tra cứu ảnh dựa trên đặc trưng bất biến.....	19
1.4 Các hệ thống tra cứu ảnh dựa trên nội dung.....	19
1.4.1 Google Image Search.....	19
1.4.2 Bing Image Search.....	20

1.4.3	Yahoo Image Search.....	20
1.4.4	PicSearch.....	21
1.5	Các ứng dụng cơ bản của tra cứu ảnh dựa trên nội dung.....	21
CHƯƠNG 2: Đối sánh ảnh dựa trên đặc trưng SIFT.....		23
2.1	Giới thiệu	23
2.2	Trích chọn đặc trưng SIFT.....	23
2.2.1	Phát hiện các điểm cực trị	25
2.2.2	Định vị điểm hấp dẫn:.....	28
2.2.3	Xác định hướng cho các điểm hấp dẫn.....	31
2.2.4	Mô tả các điểm hấp dẫn	32
2.3	Đối sánh đặc trưng SIFT.....	33
2.3.1	Độ đo khoảng cách và độ đo tương tự.....	33
2.3.2	Đối sánh đặc trưng cục bộ bất biến	34
2.3.2.1	Đối sánh các vector đặc trưng	34
2.3.2.2	Một số độ đo tương đồng cho ảnh sử dụng đặc trưng SIFT	35
CHƯƠNG 3: Thực nghiệm.....		36
3.1	Môi trường và các công cụ sử dụng trong thực nghiệm	36
3.2	Xây dựng tập dữ liệu ảnh.....	36
3.3	Giao diện chương trình.....	38
3.4	Một số kết quả.....	39
KẾT LUẬN		41
TÀI LIỆU THAM KHẢO.....		42

MỞ ĐẦU

Sự phát triển mạnh mẽ của công nghệ ảnh số làm lượng ảnh lưu trữ trên web tăng lên một cách nhanh chóng. Mỗi ngày, có hàng triệu bức ảnh được đăng tải trên các trang ảnh trực tuyến như: Flickr, Photobucket, Facebook... Theo thống kê đã có 30 tỉ ảnh trên Facebook, 4 tỉ ảnh trên Flickr, 6.2 tỉ ảnh trên Photobucket.

Cùng với nhu cầu tìm kiếm văn bản, nhu cầu tìm kiếm ảnh cũng nhận được nhiều quan tâm của người sử dụng. Tuy nhiên, với một số lượng ảnh quá lớn trên Internet công việc tìm kiếm trở nên vô cùng khó khăn. Để giải quyết vấn đề này, các hệ thống tìm kiếm ảnh đã ra đời như: Yahoo, Google Image Search, Bing,... Các hệ thống này cho phép người sử dụng nhập truy vấn về các ảnh cần quan tâm. Thông qua việc phân tích các văn bản đi kèm ảnh, hệ thống gửi trả các ảnh tương ứng với truy vấn của người dùng. Ngoài ra một số hệ thống còn cho phép người dùng nhập câu hỏi dưới dạng ảnh như Google Image Search, TinEye, Tiltomo... Đây là một hướng nghiên cứu mới nhận được sự quan tâm của nhiều công trình khoa học trên thế giới.

Hiện nay trên thế giới việc tra cứu ảnh đã bước sang thời kỳ mới, thời kỳ tra cứu ảnh dựa vào nội dung. Tra cứu dữ liệu hình ảnh dựa vào nội dung ảnh ngày càng phát triển mạnh mẽ, nó khắc phục khuyết điểm của việc truy tìm ảnh dựa vào văn bản kí tự. Dữ liệu đầu vào được mô phỏng gần gũi với con người, kết quả ảnh trả về mang ngữ nghĩa gần đúng với ảnh truy vấn hơn.

Nằm trong xu thế đó, trong đồ án này em trình bày một mô hình tra cứu thông tin hình ảnh dựa trên các đặc trưng bất biến của ảnh.

Nội dung của đề tài bao gồm ba chương:

- Chương 1: Tổng quan về tra cứu ảnh dựa trên nội dung
- Chương 2: Đối sánh ảnh dựa trên đặc trưng SIFT
- Chương 3: Thực nghiệm

CHƯƠNG 1: Tổng quan về tra cứu ảnh dựa trên nội dung

1.1 Tra cứu thông tin trực quan

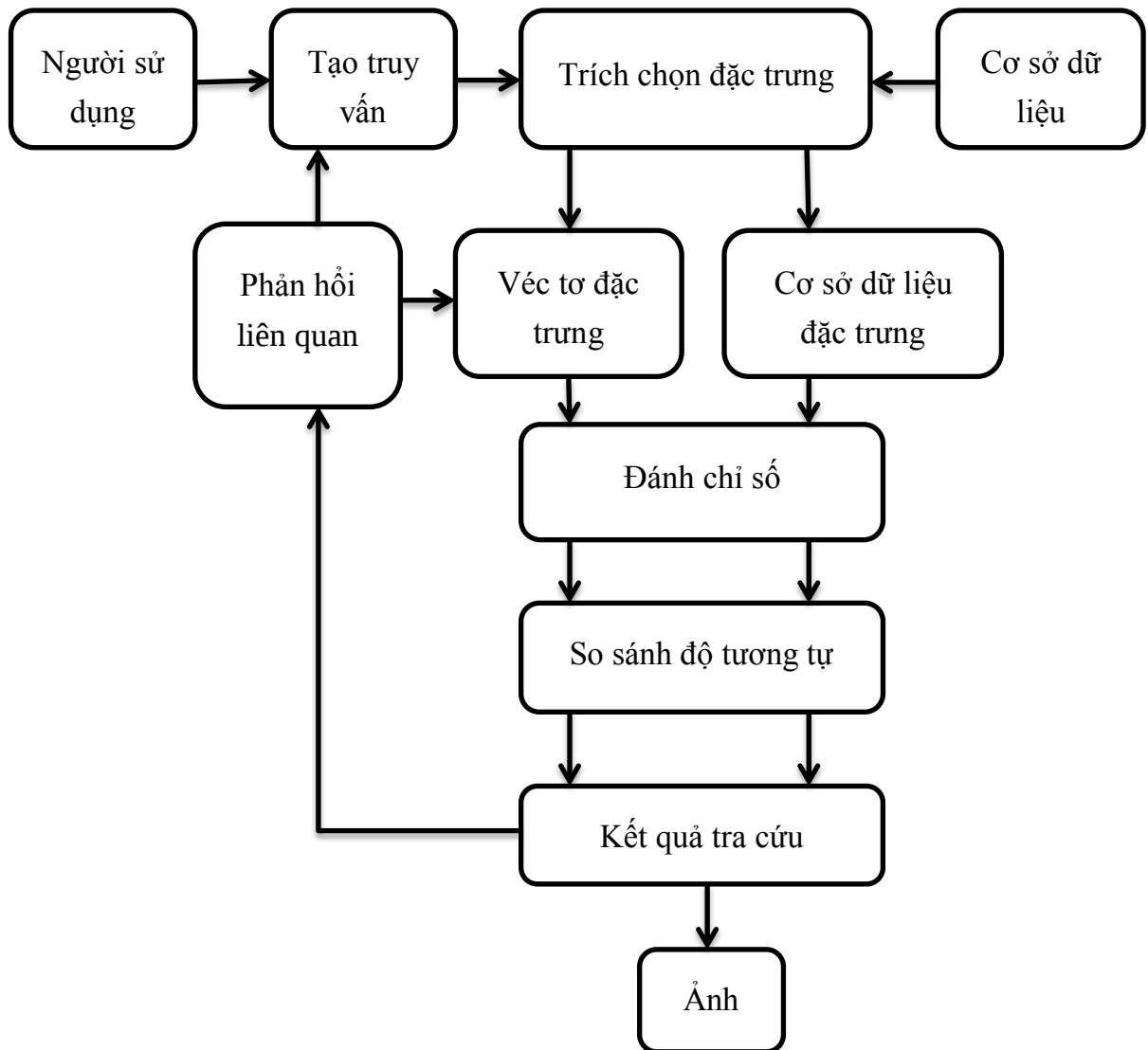
Tra cứu thông tin trực quan là chủ đề nghiên cứu mới trong lĩnh vực công nghệ thông tin. Tương tác với nội dung trực quan là cách thiết yếu nhất để truy tìm thông tin trực quan. Các yếu tố trực quan như màu sắc, kết cấu, hình dáng đối tượng và các yếu tố không gian trực tiếp liên quan đến khía cạnh của cảm nhận nội dung ảnh, cùng với các khái niệm ở mức cao như ý nghĩa đối tượng, khung cảnh trong ảnh, được dùng như là manh mối cho tìm kiếm hình ảnh với nội dung tương tự từ cơ sở dữ liệu.

Tra cứu ảnh dựa vào nội dung đòi hỏi phải có sự đóng góp từ các lĩnh vực nghiên cứu khác là rất lớn và đặt ra nhiều thử thách trong nghiên cứu đối với các nhà khoa học và kỹ sư. Các lĩnh vực nghiên cứu khác nhau, được phát triển một cách độc lập, đóng góp rất lớn cho chủ đề nghiên cứu mới mẻ này.

1.2 Những thành phần cơ bản của 1 hệ thống tra cứu ảnh dựa trên nội dung

Về cơ bản hệ thống phân tích cả các nội dung của nguồn thông tin cũng như các truy vấn sử dụng, và đem so sánh các nội dung này để tra cứu các mục tin liên quan. Một hệ thống tra cứu ảnh bao gồm các chức năng sau :

- Phân tích các nội dung của các nguồn thông tin, và biểu diễn các nội dung của các nguồn được phân tích theo cách thích hợp cho việc so sánh các truy vấn sử dụng.
- Phân tích các truy vấn người sử dụng và biểu diễn chúng ở dạng thích hợp cho so sánh với cơ sở dữ liệu nguồn
- Định nghĩa một chiến lược để so sánh các truy vấn tìm kiếm với thông tin trong cơ sở dữ liệu được lưu trữ.
- Thực hiện các điều chỉnh cần thiết trong hệ thống dựa trên phản hồi từ người sử dụng hoặc các ảnh được tra cứu.



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

1.3 Các phương pháp tra cứu ảnh dựa trên nội dung

1.3.1 Tra cứu ảnh dựa trên màu sắc

Tìm kiếm ảnh theo màu sắc là phương pháp phổ biến và được sử dụng nhiều nhất trong các hệ thống tìm kiếm ảnh theo nội dung. Đây là phương pháp đơn giản, tốc độ tìm kiếm tương đối nhanh tuy nhiên kết quả tìm kiếm có độ chính xác không cao.

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

Có ba kỹ thuật truyền thống được sử dụng trong tra cứu ảnh dựa trên màu sắc đó là biểu đồ màu toàn cục (Global Color Histogram), biểu đồ màu cục bộ

(Local Color Histogram) và biểu đồ màu tương quan (Color Correlogram Histogram). Những kỹ thuật này thích hợp với các kiểu truy vấn khác nhau.

1.3.1.1 Biểu đồ màu toàn cục (Global Color Histogram)

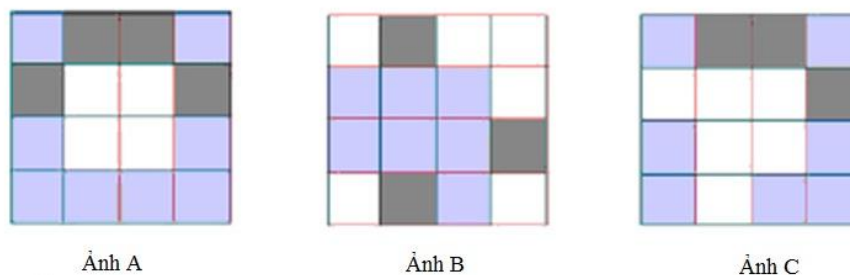
Biểu đồ màu loại này mô tả phân bố màu sử dụng tập các màu. Việc sử dụng biểu đồ màu toàn cục thì một ảnh sẽ được mã hóa với biểu đồ màu của nó và khoảng cách giữa hai ảnh sẽ được xác định bởi khoảng cách giữa những biểu đồ màu của chúng. Với kỹ thuật này chúng ta có thể sử dụng các thước đo khác nhau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

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

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

$$d_{GCH}(A, C) = d_{GCH}(A, B) = d_{GCH}(B, C) = 0 \quad [1.1]$$

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



Hình 1.2: Ba ảnh và biểu đồ màu của chúng
ảnh A: {25%, 25%, 50%}; ảnh B: {18.75%, 37.5%, 43.75%}; ảnh C: {18.75%, 37.5%, 43.75%}

1.3.1.2 Biểu đồ màu cục bộ (Local Color Histogram)

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

$$D(Q, I) = \sum_{k=1}^M \sqrt{\sum_{i=1}^N (H_0^k[i] - H_1^k[i])^2} \quad [1.2]$$

Ở đây M là số vùng được phân đoạn trong ảnh, N là số màu trong biểu đồ màu và $H[i]$ là giá trị của màu i trong biểu đồ màu đại diện cho vùng k của ảnh.

1.3.1.3 Biểu đồ màu tương quan (Color Correlogram Histogram)

Quan sát thấy rằng lược đồ màu thiếu thông tin về cách mà màu sắc được phân bố theo không gian, một đặc trưng mới được giới thiệu gọi là lược đồ tương quan màu. Lược đồ tương quan màu hứa hẹn mô tả không chỉ là phân phối màu của các điểm ảnh mà còn là sự tương quan về không gian giữa các cặp màu. Lược đồ này chỉ quan tâm đến sự tương quan về không gian giữa những màu giống nhau và do đó giảm được số chiều và chi phí tính toán.

Cách tính lược đồ tương quan màu :

Gọi $[D]$ là tập gồm D khoảng cách $d_1, d_2, \dots, d_D$ được đo bằng độ đo L_∞ .

Lược đồ tương quan màu của ảnh I được xác định với cặp màu c_i, c_j và khoảng cách d như sau:

$$\gamma_{c_i c_j}^{(d)}(I) = Pr_{P_1 \in I_{c_i}, P_2 \in I_{c_j}} [P_1 - P_2 | L_\infty = d] \quad [1.3]$$

Trong đó I là ảnh, kích thước $M \times N$ (Điểm ảnh), $I_{c_i} = \{p \in I \mid p = c_i\}$, lược đồ tương quan màu thể hiện xác suất cặp điểm ảnh bất kỳ p_1 và p_2 chịu sự ràng buộc về màu (p_1 có màu c_i , p_2 có màu c_j) và vị trí ($|p_1 - p_2| = d$).

1.3.1.4 Vector liên kết màu (Color Coherence Vector)

Vector liên kết màu đề xuất phân mỗi ngăn của lược đồ thành hai loại: liên kết nếu nó thuộc về một vùng màu đồng nhất lớn hoặc không liên kết nếu nó

không thuộc về một vùng màu đồng nhất lớn. Cho α_i biểu thị số các pixel gắn kết trong ngăn thứ i và β_i biểu thị số các pixel không gắn kết trong một ảnh thì vector liên kết màu của một ảnh được định nghĩa bằng vector $\langle(\alpha_1, \beta_1), (\alpha_2, \beta_2), (\alpha_3, \beta_3), \dots, (\alpha_N, \beta_N)\rangle$.

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

Việc thông tin không gian được kết hợp vào biểu đồ màu sắc làm cho Vector liên kết màu cung cấp các kết quả tra cứu tốt hơn lược đồ màu, đặc biệt với các ảnh có phần lớn màu đồng nhất hoặc có kết cấu theo khu vực.

1.3.1.5 Tương quan màu (Color Correlogram)

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

$$\gamma(k)_{i,j} = \Pr [p_2 \in I_{c(j)} | |p_1 - p_2| = k] \quad [1.4]$$

Trong đó: $p_1 \in I_{c(i)}, p_2 \in I$.

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

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

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

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

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

Giá trị d lớn thì cần nhiều chi phí tính toán và không gian lưu trữ.

Giá trị d nhỏ có thể giảm giá trị lưu trữ của đặc trưng.

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

1.3.1.6 Độ đo tương đồng về màu sắc

Một số độ đo tương đồng được sử dụng như: Độ đo khoảng cách Euclidean, độ đo Jensen-Shannon divergence (JSD). Gọi $h(I)$ và $h(M)$ tương ứng

là 2 lượt đồ màu của hai ảnh I và ảnh M. Khi đó các loại độ đo màu được định nghĩa là một số nguyên (hoặc số thực) theo các loại độ đo tương ứng như sau:

- Khoảng cách Euclidean:

Đây là khoảng cách Euclidean thông thường giữa các K bin:

$$Intersection(h(I), h(M)) = \sum_{j=1}^K \sqrt{(h(I) - h(M))^2} \quad [1.5]$$

Hoặc:

$$Intersection(h(I), h(M)) = \sum_{j=1}^K |h(I) - h(M)| \quad [1.6]$$

- Độ đo Jensen-Shannon divergence (JSD) : Độ đo Jensen-Shannon divergence sử dụng lược đồ màu RGB để tính toán độ tương đồng về màu sắc giữa 2 ảnh:

$$d_{JSD}(H, H') = \sum_{m=1}^M H_m \log \frac{2H_m}{H_m + H'_m} + H'_m \log \frac{2H'_m}{H'_m + H_m} \quad [1.7]$$

Trong đó : H và H' là 2 biểu đồ màu được so sánh, m H là bin thứ m của biểu đồ H.

1.3.2 Tra cứu ảnh dựa trên 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. Kết cấu được sử dụng rộng rãi và rất trực quan nhưng không có định nghĩa chính xác bởi tính biến thiên rộng của nó. Mặc dù không có một khái niệm chung cho kết cấu nhưng tất cả các nhà nghiên cứu đều tập trung thống nhất trên hai điểm chính:

- Trong phạm vi một kết cấu có sự biến đổi đáng kể về mức độ cường độ giữa các điểm ảnh liên kề, đó là giới hạn của độ phân giải, không có sự đồng nhất.
- Kết cấu là thuộc tính đồng nhất ở một vài không gian lớn hơn độ phân giải của ảnh, cái hàm ý trong những thuộc tính này của cấu trúc là ảnh có độ phân giải nhất định.

Khác với màu sắc, kết cấu diễn ra trên cả một vùng hơn là tại một điểm, nó thường được định nghĩa bằng những mức xám được hiểu như là màu sắc.

Các phương pháp phân tích kết cấu bao gồm : phương pháp Gray-Level Co-occurrence Matrices(GLC), phương pháp Gray-Level Difference (LGD).

Ma trận đồng mức xám có một số khó khăn đó là: không có một phương pháp được thiết lập hoàn hảo về sự lựa chọn véc tơ thay thế d và việc tính toán ma trận đối với một số giá trị khác nhau của d là không thể thực hiện được. Hơn nữa, với một giá trị của d có một số lượng lớn các đặc điểm có thể được tính toán. Điều này có nghĩa là một số phương pháp lựa chọn đặc điểm cần phải được sử dụng để lựa chọn những đặc điểm có liên quan nhất.

Bảng 1.1: Một số trích chọn đặc điểm kết cấu từ ma trận đồng nhất mức xám

Đặc điểm kết cấu	Công thức
Energy	$\sum_i \sum_j P_{d^2}(i, j)$
Entropy	$-\sum_i \sum_j P_d(i, j) \log \sum_i \sum_j P_d(i, j)$
Contrast	$\sum_i \sum_j (i - j)^2 P_d(i, j)$
Homogeneity	$\sum_i \sum_j P_{d^2}(i, j) / i + i - j $
Correlation	$\sum_i \sum_j (i - \mu_x)(j - \mu_y) P_d(i, j) / \sigma_x, \sigma_y$

Trong đó:

$P_d(i, j)$ Là phần tử thứ (i, j) của ma trận co-occurrence P_d

$\sum_i$ Nghĩa là : $\sum_{i=1}^M$ với M là số hàng

$\sum_j$ Nghĩa là: $\sum_{j=1}^N$ với N là số cột

$\sum_{i,j}$ Nghĩa là: $\sum_i, \sum_j$

- **Energry** (độ nhiễu) của kết cấu mô tả sự tương tự của kết cấu. Trong ảnh đồng nhất có rất ít chuyển đổi mức xám trọi, bởi vậy ma trận co-occurrence sẽ có ít vùng có cường độ lớn. Như vậy energry của ảnh là cao khi ảnh là đồng nhất.
- **Entropy** (năng lượng) đo sự ngẫu nhiên của những phần tử trong ma trận khi tất cả những phần tử của ma trận là ngẫu nhiên tối đa thì entropy có giá trị cao nhất. Bởi vậy một ảnh đồng nhất có entropy thấp hơn ảnh không đồng nhất.
- **Contrast** (độ tương phản) có giá trị cao tương đối khi những giá trị cao của ma trận gần với đường chéo chính.
- **Correlation** (độ tương quan) đo tương quan giữa các phần tử của ma trận, khi giá trị này cao thì ảnh phức tạp hơn.

1.3.2.2 Phương pháp Gray-Level Difference (GLD)

Phương pháp Gray-Level Difference (GLD) tương tự với các phương pháp Gray-Level Co-occurrence Matrices(GLC). Tuy nhiên điểm khác biệt chính giữa chúng là, trong khi phương pháp GLC tính toán ma trận của các cặp cường

độ thì phương pháp GLD lại tính toán một véc tơ của những chênh lệch cường độ. Điều này tương đương với việc tổng kết ma trận GLC với những đường chéo của nó.

Cụ thể, cho bất kỳ khoảng cách thay thế $d = (dx, dy)$ thì:

$$I_d(x, y) = ||I(x, y) - I(x + dx, y + dy)|| \quad [1.9]$$

Cho P_d là mật độ xác suất của $I_d(x, y)$. Nếu có m mức xám thì sẽ tạo thành một véc tơ m chiều trong đó thành phần thứ i chính là xác suất mà $I_d(x, y)$ sẽ có giá trị i . Nếu ảnh I là rời rạc thì dễ dàng tính toán P_d bằng việc đếm số lần mỗi giá trị $I_d(x, y)$ xảy ra.

1.3.2.3 Độ đo tương đồng cho kết cấu ảnh

Để đo độ tương đồng theo kết cấu giữa các ảnh, người ta thường sử dụng độ đo Euclidean. Kết cấu được trích xuất từ các bức ảnh sẽ được biểu diễn thành các vector nhiều chiều và khoảng cách Euclidean được dùng để đo độ tương đồng giữa các đặc trưng của ảnh truy vấn với đặc trưng của ảnh trong cơ sở dữ liệu.

1.3.3 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. Trong tra cứu ảnh theo nội dung, hình dạng là một cấp cao hơn so với màu sắc và kết cấu. Nó đòi hỏi sự phân biệt giữa các vùng để tiến hành xử lý về độ đo của hình dạng.

Trước đây, nghiên cứu hình dạng được thúc đẩy chủ yếu bởi sự nhận dạng đối tượng, các kỹ thuật mô tả và biểu diễn hình dạng này chủ yếu dựa vào các ứng dụng cụ thể. Trong đó, sự hiệu quả và chính xác là mối quan tâm chính của những kỹ thuật này.

Việc phân loại các phương pháp biểu diễn hình dạng phổ biến nhất là dựa trên việc sử dụng các điểm biên hình dạng và điểm vùng.

1.3.3.1 Phương pháp trích chọn đặc trưng dựa trên đường biên.

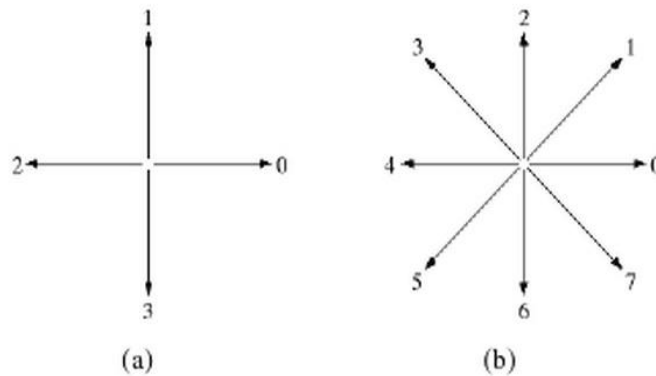
Trong phần này, chúng ta sẽ xem xét cụ thể phương pháp sử dụng mã xích cùng với shape number để biểu diễn và nhận dạng đối tượng.

a) Mã xích

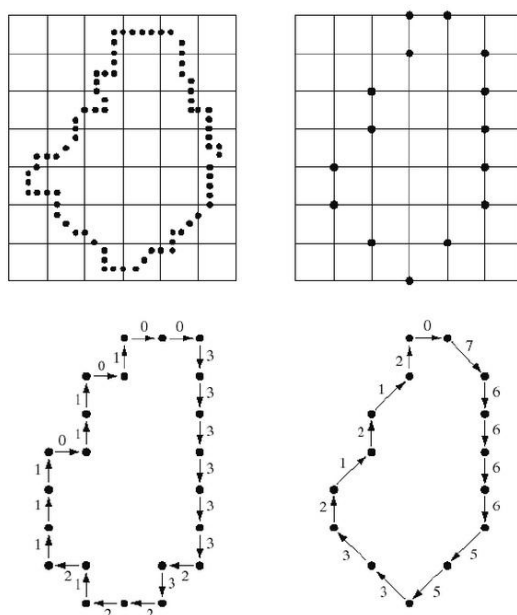
Mã xích biểu diễn đường biên đối tượng bằng một chuỗi kết nối của các phân đoạn đường thẳng có độ dài quy định và định hướng. Thông thường, biểu diễn này dựa trên 4 hoặc 8 hướng kết nối của các phân đoạn đường thẳng. Hướng của mỗi phân đoạn được mã hóa bằng cách sử dụng một lược đồ số như được hiển thị trong hình 1.4.

Những hình ảnh kỹ thuật số thường được xử lý với định dạng lưới với khoảng cách bình đẳng với các hướng x và y. Một chuỗi mã có thể tạo ra bằng cách định hướng các phân đoạn đường thẳng dọc theo biên theo chiều kim đồng hồ như minh họa trong hình 1.5.

Vấn đề đặt ra là một chuỗi mã phụ thuộc vào điểm bắt đầu và giải pháp được đưa ra là coi chuỗi mã như một chuỗi kín và xác định điểm bắt đầu để chuỗi kết quả không phụ thuộc vào sự lựa chọn điểm bắt đầu đó. Chúng ta có thể chuẩn hóa mã xích với phép quay bằng cách sử dụng sự khác biệt đầu tiên của mã xích thay vì bản thân mã. Sự khác biệt này thu được bằng cách đếm số lượng các hướng thay đổi giữa 2 yếu tố liên tiếp.



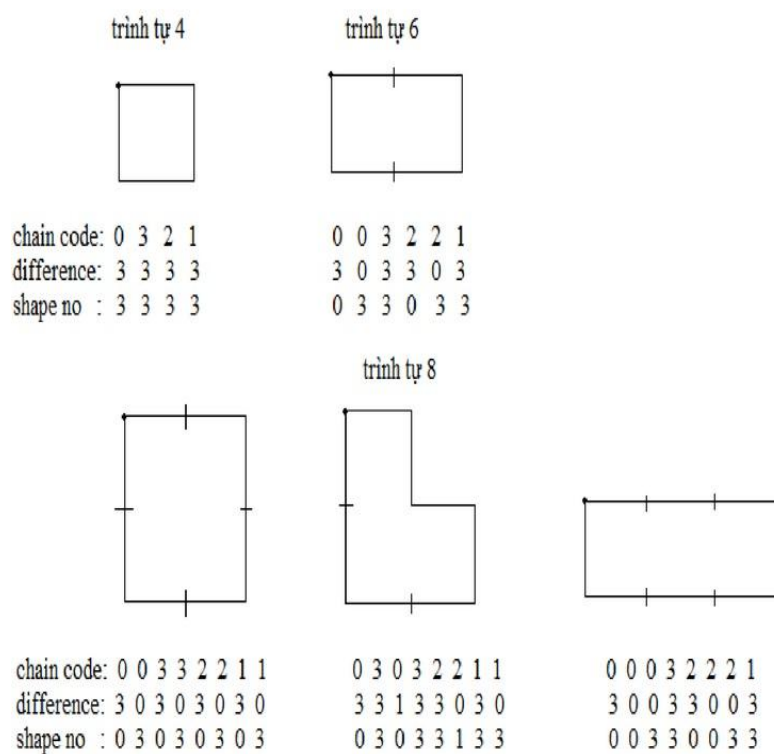
Hình 1.4: Các hướng của đoạn thẳng đơn vị : (a): 4 hướng, (b): 8 hướng



Hình 1.5: Hình ảnh của một chuỗi mã (theo 4 hướng và 8 hướng)

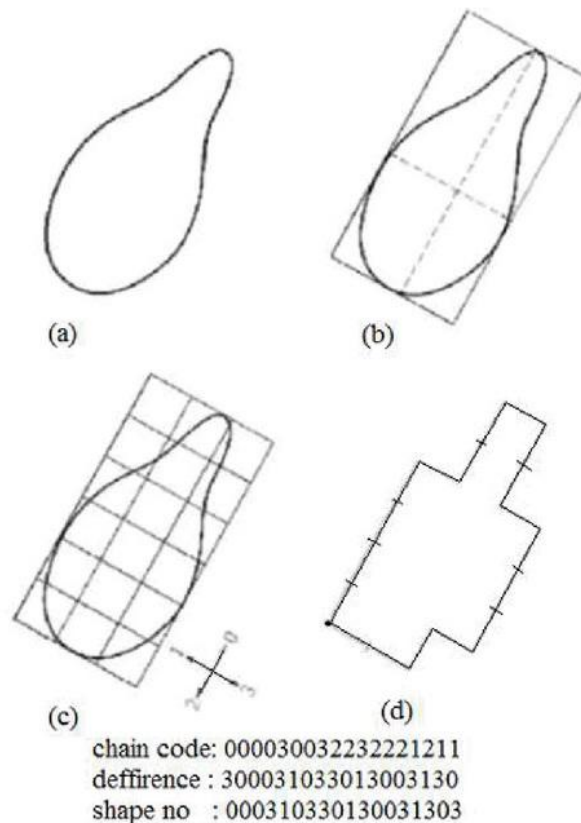
b) Shape number

Shap number của một biểu diễn đường biên được định nghĩa là sự khác biệt đầu tiên của cường độ nhỏ nhất. Trình tự n của một shape number là số lượng các chữ số được biểu diễn. Hình 1.6 minh họa hình dạng của trình tự 4,6,8.



Hình 1.6 : Biểu diễn hình dạng sử dụng shape number

Chúng ta xét một ví dụ cụ thể, giả sử $n=18$ được quy định cụ thể cho biên như hình 1.7(a). Để có được một shape number của trật tự này đòi hỏi phải làm theo các bước sau: Bước đầu tiên là tìm các hình chữ nhật cơ bản như trong hình 1.7(b). Hình chữ nhật gần nhất của trật tự 18 là hình chữ nhật 3×6 , yêu cầu phải chia nhỏ hình chữ nhật cơ bản như trong hình 1.7(c). Cuối cùng có được chuỗi mã và sử dụng điểm khác biệt đầu tiên để tính toán shape number.



Hình 1.7: Các bước tính toán shape number

1.3.3.2 Phương pháp trích chọn đặc trưng dựa trên vùng.

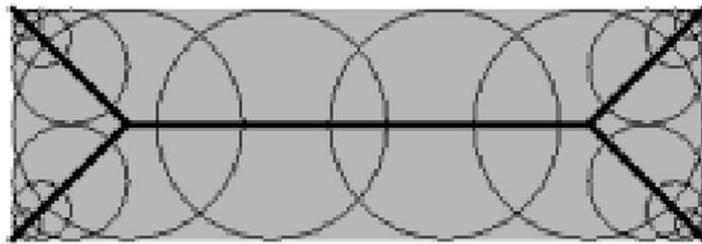
Trong phương pháp biểu diễn dựa trên vùng phải kể đến tất cả những pixel trong vùng hình dạng thu được trong biểu diễn hình dạng. Phương pháp biểu diễn vùng thường sử dụng các momen để mô tả hình dạng. Và một số phương pháp khác thường sử dụng gồm có: phương pháp lưới, bề mặt lồi và trục trung vị.

Biểu diễn hình dạng dựa trên vùng xem xét đến toàn bộ vùng hình dạng và sử dụng hiệu quả thông tin của toàn bộ pixel chứa trong vùng. Những phương pháp này đo sự phân phối pixel của vùng hình dạng, chúng ít có khả năng giả tạo bởi nhiễu và biến dạng. Phương pháp vùng phổ biến là những phương pháp moment. Ở mức thấp moment thứ tự hay momnet bất biến mang theo những ý

nghĩa vật lý kết hợp với sự phân phối pixel. Tuy nhiên nó rất khó khăn để kết hợp moment thứ tự cao hơn với sự giải thích vật lý. Phương pháp lưới là dựa trên khả năng trực quan quan sát hình dạng, nó không phản ánh sự thống kê phân bố của vùng hình dạng và bị ảnh hưởng bởi nhiễu và không cô đọng như moment bất biến.

1.3.3.2.1 Đồ thị xương

Xương (trục trung vị) là quỹ tích tâm của các đĩa cực đại của hình dạng như trong hình 1.8, đường in đậm là xương của hình chữ nhật.



Hình 1.8: Đồ thị xương của hình chữ nhật.

Ý tưởng cơ bản của việc sử dụng xương là loại bỏ các thông tin dư thừa trong khi vẫn giữ được các thông tin topo có liên quan đến cấu trúc của đối tượng để có thể nhận dạng đối tượng. Xương có thể được phân tách thành các đoạn và được biểu diễn dưới dạng các đồ thị theo một tiêu chí nhất định. Như vậy, việc đối sánh giữa các hình dạng sẽ trở thành việc đối sánh giữa các đồ thị. Tuy nhiên việc tính toán đối với xương khá phức tạp, hơn nữa xương rất nhạy cảm với nhiễu và các biến dạng.

1.3.3.3 Các phương pháp đối sánh dựa trên hình dạng

1.3.3.3.1 Đối sánh các shape number

Mức độ tương tự k giữa 2 hình dạng được định nghĩa là thứ tự lớn nhất mà shape number vẫn còn trùng khớp. Ví dụ, cho hai hình dạng a và b được biểu diễn bởi một chuỗi mã 4 hướng, hai hình dạng có độ tương tự k nếu:

$$S_j a = S_j b \text{ với } j=4, 6, 8, \dots, k$$

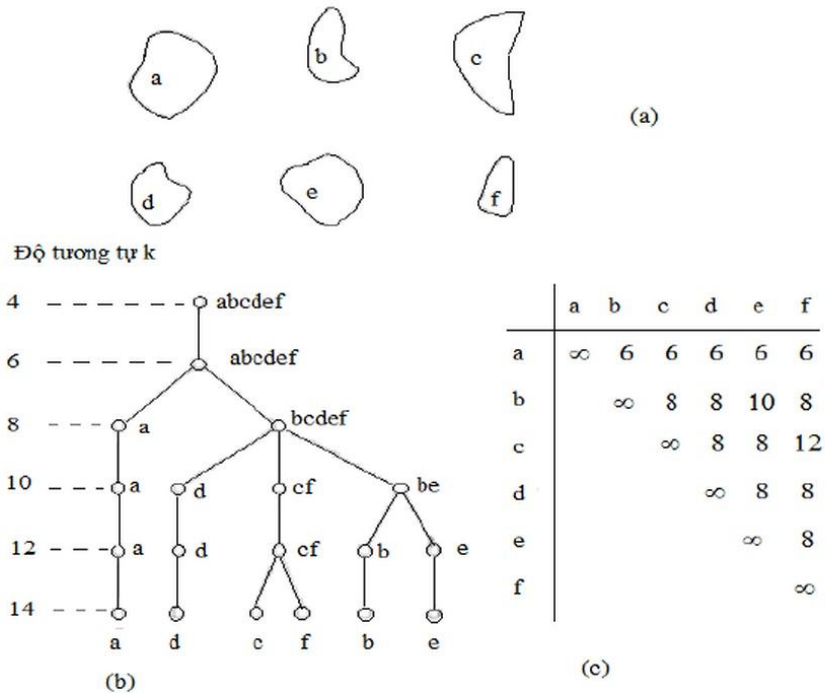
$$S_j a \neq S_j b \text{ với } j=k+2, k+4, \dots$$

Trong đó S cho biết shape number và chỉ số dưới là trình tự.

Khoảng cách giữa hai hình a và b được định nghĩa là nghịch đảo của mức độ tương tự: $D a, b = \frac{1}{k}$

Khoảng cách này có các thuộc tính sau:

- $D(a,b) \geq 0$
- $D(a,b) = 0$ nếu $a=b$
- $D(a,b) \leq \max[D(a,b), D(b,c)]$



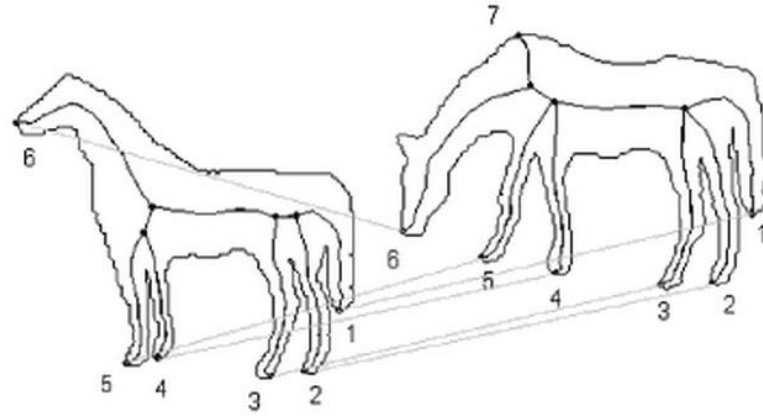
Hình 1.9 : Minh họa tìm kiếm hình dạng tương tự sử dụng shape number: (a) hình dạng; (b) cây tương tự; (c) ma trận tương tự.

1.3.3.3.2 Đối sánh đồ thị xương

Ý tưởng chính của phương pháp này là đối sánh đồ thị xương bằng cách so sánh các đường dẫn tới điểm cuối xương. Phương pháp đối sánh này không dựa trên cấu trúc topo hình học, bởi một thực tế trực quan là những bộ xương tương tự có thể có cấu trúc topo hình học khác nhau. Việc so sánh các đường dẫn giữa các điểm cuối của đồ thị xương mang lại kết quả chính xác phù hợp với mọi trường hợp. Thông thường dùng cho nhận dạng là các nhánh xương đã được cắt tia. Các xương được cắt tia bởi phân chia đường biên có điểm cuối của nhánh xương tương ứng với phần trực quan của đối tượng. Kết quả thực nghiệm cho thấy rằng phương pháp này có thể tạo ra kết quả chính xác với sự có mặt của sự khớp xương, sự kéo dài xương và biến dạng đường biên.

Để đối sánh đồ thị xương, độ tương tự của các đường đi ngắn nhất giữa mỗi cặp điểm cuối xương được sử dụng để thiết lập mối quan hệ tương ứng với điểm cuối trong đồ thị khác. Cuối cùng giá trị không giống nhau giữa các đồ thị

là tính ước lượng khoảng cách giữa các điểm cuối tương ứng. Vì vậy ý tưởng cơ bản của phương pháp này là xác định sự giống nhau của các cấu trúc phức tạp của đồ thị hoặc cây bằng cách kiểm tra đường đi ngắn nhất giữa các điểm cuối của chúng.



Hình 1.10: Sự tương ứng giữa các điểm cuối của hai đồ thị xương.

1.3.4 Tra cứu ảnh dựa trên đặc trưng bất biến

Phương pháp tra cứu này có tên là Scale-Invariant Feature Transform (SIFT) và đặc trưng trích rút được gọi là đặc trưng SIFT. Phương pháp này trích rút các đặc trưng cục bộ bất biến của ảnh. Các đặc trưng này bất biến với việc thay đổi tỉ lệ ảnh, quay ảnh, đôi khi là thay đổi điểm nhìn và thêm nhiễu ảnh hay thay đổi cường độ chiếu sáng của ảnh. Các đặc trưng này được trích rút ra từ các điểm đặc trưng cục bộ. Điểm đặc trưng: Là vị trí (điểm ảnh) "đặc trưng" trên ảnh. "Đặc trưng" ở đây có nghĩa là điểm đó có thể có các đặc trưng bất biến với việc quay ảnh, co giãn ảnh hay thay đổi cường độ chiếu sáng của ảnh.

1.4 Các hệ thống tra cứu ảnh dựa trên nội dung

Những năm gần đây, có nhiều hệ thống tra cứu ảnh, các hệ thống nghiên cứu và hệ thống thương mại đã được xây dựng. Dưới đây, là một số hệ thống của tra cứu ảnh phổ biến hiện nay.

1.4.1 Google Image Search

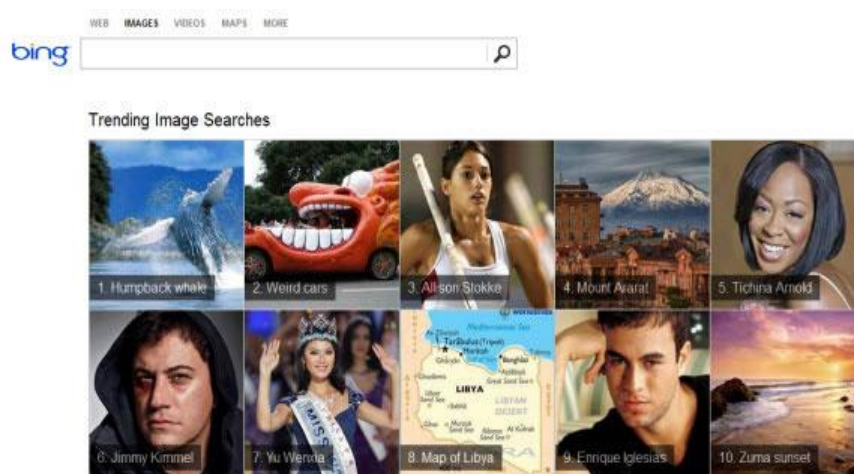
Khi nhắc đến việc tìm kiếm, trước tiên ai cũng sẽ nghĩ đến Google, điều đó là hiển nhiên vì công cụ này có thể cung cấp cho bạn một số lượng lớn các ảnh thuộc nhiều chủ đề khác nhau đã được lập chỉ mục cụ thể. Google Image Search là một công cụ tuyệt vời để tìm kiếm hình ảnh với từ khoá. Nó cũng hỗ trợ một số thuộc tính cơ bản như kích thước, hình dạng,...



Hình 1.11 : Công cụ tìm kiếm hình ảnh của Google

1.4.2 Bing Image Search

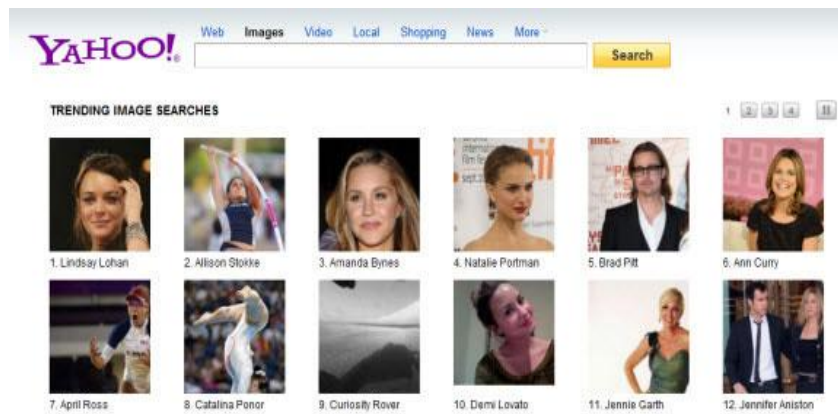
Là bộ máy tìm kiếm do hãng khổng lồ phần mềm Microsoft phát triển, đây có thể không phải là công cụ phổ biến như Google, tuy nhiên trong một số trường hợp, nó hoạt động tốt hơn cả Google. Tìm kiếm hình ảnh thông qua công cụ tìm kiếm của Bing là việc khá cơ bản, thế nhưng nó lại có thể mang lại một số lượng tốt các kết quả có liên quan.



Hình 1.12 : Công cụ tìm kiếm hình ảnh Bing

1.4.3 Yahoo Image Search

Công cụ tìm kiếm này do yahoo phát triển, nó cũng có thể cung cấp cho bạn các kết quả tốt đẹp khi bạn cần tìm các bức ảnh trong một chủ đề xu hướng nhất định. Tìm kiếm hình ảnh được trang bị với tất cả các tiêu chuẩn cơ bản của một công cụ tìm kiếm hình ảnh.



Hình 1.13 : Công cụ tìm kiếm hình ảnh của Yahoo

1.4.4 PicSearch

Được phát triển bởi công ty Picsearch Services, Picsearch là một công cụ tìm kiếm chỉ dành riêng cho mục đích duy nhất là tìm kiếm hình ảnh. Picsearch tự hào có một kho lưu trữ được lập chỉ mục với hơn 3 tỷ hình ảnh.



Hình 1.14 : Công cụ tìm kiếm Picsearch

1.5 Các ứng dụng cơ bản của tra cứu ảnh dựa trên nội dung

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

- Quân sự
- Quản lý tài sản trí tuệ
- Ngăn chặn tội phạm
- 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

CHƯƠNG 2: Đối sánh ảnh dựa trên đặc trưng SIFT

2.1 Giới thiệu

Trong đối sánh ảnh dựa trên nội dung, việc trích chọn dấu hiệu đặc trưng có ý nghĩa hết sức quan trọng. Với ảnh tự nhiên việc đối sánh gặp nhiều khó khăn do phải đối mặt với các thách thức do các thay đổi của ảnh như độ chiếu sáng, co giãn, chồng lấp, thay đổi góc nhìn... nên ảnh hưởng nhiều đến độ chính xác. Phương pháp đối sánh ảnh dựa trên đặc trưng bất biến SIFT được David Lowe[7] đề xuất có thể khắc phục một số khó khăn trên.

2.2 Trích chọn đặc trưng SIFT

Một thuật toán tiêu biểu và có hiệu quả khá cao dựa theo các đặc trưng cục bộ bất biến trong ảnh: SIFT (Scale-invariant Feature Transform) do David Lowe[7] đưa ra từ năm 2004 và đến nay đã có nhiều cải tiến trong thuật toán. Đặc trưng được trích chọn trong SIFT là các điểm đặc trưng (keypoint), các điểm này kèm theo các mô tả về nó và một véc tơ có lấy keypoint làm điểm gốc.

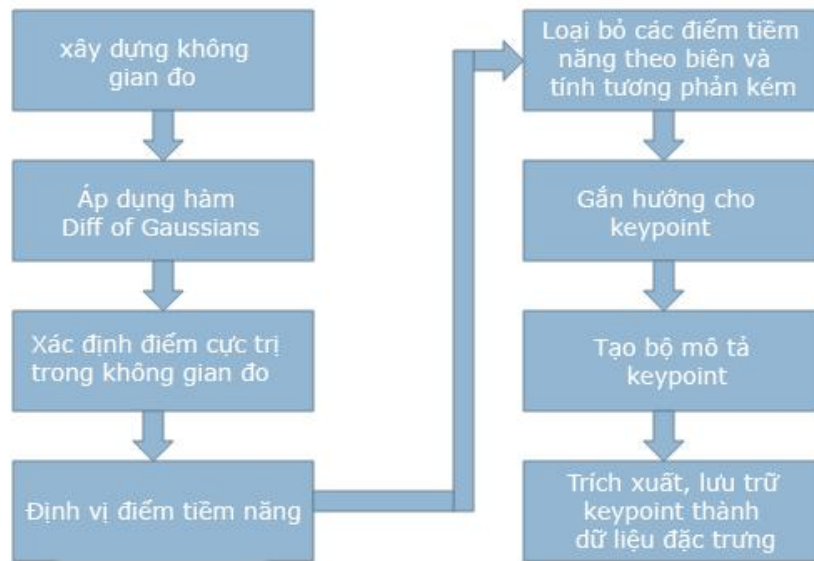
Phương pháp trích rút các đặc trưng bất biến SIFT được tiếp cận theo phương pháp lọc kim tự tháp, theo đó phương pháp được thực hiện lần lượt theo các bước sau:

- *Phát hiện các điểm cực trị (Scale-Space extrema detection)*: Bước đầu tiên này sẽ áp dụng đạo hàm của hàm Gaussian (DoG - Difference of Gaussians) để tìm ra các điểm có khả năng làm điểm đặc trưng tiềm năng, đó là những điểm rất ít phụ thuộc (bất biến) vào sự thu phóng ảnh và xoay ảnh.
- *Định vị các điểm đặc trưng (keypoint localization)*: Từ những điểm tiềm năng ở trên sẽ lọc và lấy ra tập các điểm đặc trưng tốt nhất (keypoints).
- *Xác định hướng cho các điểm đặc trưng (Orientation assignment)*: Mỗi điểm đặc trưng sẽ được gán cho một hoặc nhiều hướng dựa trên hướng gradient của ảnh. Mọi phép toán xử lý ở các bước sau này sẽ được thực hiện trên những dữ liệu ảnh mà đã được biến đổi tương đối so với hướng đã gán, kích cỡ và vị trí của mỗi điểm đặc trưng. Nhờ đó, tạo ra một sự bất biến trong các phép xử lý này.
- *Mô tả các điểm đặc trưng (Keypoint descriptor)*: Các hướng gradient cục bộ được đo trong ảnh có kích cỡ cụ thể nào đó trong vùng lân cận với mỗi điểm đặc trưng. Sau đó, chúng sẽ được biểu diễn thành một dạng mà cho

phép mô tả các tầng quan trọng của quá trình bóp méo hình dạng cục bộ và sự thay đổi về độ sáng.

Tập các điểm đặc biệt thu được thường phụ thuộc rất ít vào các phép biến đổi cơ bản như xoay, phóng to, thu nhỏ, tăng giảm cường độ sáng, vì vậy có thể xem đây là các đặc trưng mang tính cục bộ của ảnh. Để đối sánh và nhận dạng hai ảnh thì ta tìm tập keypoint giống nhau trong hai ảnh, dựa vào hướng và tỉ lệ để có thể biết đối tượng trong ảnh gốc đã xoay, thu phóng bao nhiêu so với ảnh đem đối sánh. Cách tiếp cận của thuật toán này dựa vào điểm bất biến cục bộ của ảnh, chúng được trích xuất ra, được định hướng và mô tả sao cho hai keypoint ở hai vùng khác nhau thì khác nhau. Tuy nhiên một yếu tố ảnh hưởng không nhỏ đến tốc độ thuật toán là số lượng các keypoint được lấy ra là không nhỏ. Trung bình một ảnh kích thước 500 x 500 pixels thì sẽ trích xuất được khoảng 1000 điểm (số lượng điểm này phụ thuộc vào tùy từng ảnh và tham số lọc khác nhau). Số lượng các điểm đặc trưng có một tầm quan trọng trong vấn đề nhận dạng đối tượng, để nhận dạng một đối tượng nhỏ trong một ảnh chứa tập hợp các đối tượng hỗn độn thì cần ít nhất 3 điểm đặc trưng giống nhau để phát hiện và bóc tách đối tượng.

Đối với vấn đề xây dựng một cơ sở dữ liệu ảnh và thực hiện nhận dạng đối tượng bất kì thì ban đầu thường sử dụng SIFT để tạo một hệ dữ liệu các đặc trưng (keypoints) được trích xuất từ dữ liệu ảnh gốc. Sau đó với mỗi ảnh đối tượng đem nhận dạng ta dùng giải thuật SIFT trích xuất tập đặc trưng từ ảnh và đem đối sánh với hệ dữ liệu đặc trưng để tìm ra tập keypoint giống nhau, từ đó nhận dạng đối tượng trong cơ sở dữ liệu ảnh ban đầu. Tuy nhiên việc đối sánh này cần chi phí đối sánh rất lớn đối với cơ sở dữ liệu ảnh có số lượng lớn do số lượng các đặc trưng ở mỗi ảnh là lớn.



Hình 2.1: Minh họa các bước chính trong giải thuật SIFT

2.2.1 Phát hiện các điểm cực trị

Như đã nêu ở trên, bước đầu tiên sẽ tìm các điểm tiềm năng có thể trở thành điểm đặc trưng bằng phương pháp lọc kim tự tháp dựa vào việc thay đổi tham số bộ lọc Gaussian. Trong bước này, ta cần dò tìm các vị trí và các độ đo mà chúng bất biến trong các khung nhìn khác nhau của cùng một đối tượng. Các vị trí đó bất biến về độ đo có thể được dò tìm bằng cách tìm kiếm các đặc trưng ổn định trên toàn bộ các độ đo có thể, sử dụng một hàm liên tục về số đo vốn rất nổi tiếng có tên là hàm độ đo không gian (Witkin 1983).

Theo các công bố của Koenderink (1984) và Lindeberg(1994) thì hàm Gaussian là hàm tốt nhất để biểu diễn độ đo không gian của ảnh 2 chiều. Vì vậy, độ đo không gian của một ảnh sẽ được định nghĩa như là một hàm $L(x,y,\sigma)$ được tạo ra bằng cách nhân chập ảnh gốc $I(x,y)$ với một hàm Gaussian $G(x,y,\sigma)$ có tham số về độ đo σ thay đổi.

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \quad [2.1]$$

Với:

$L(x, y, \sigma)$: Hàm không gian tỷ lệ của ảnh I

$G(x, y, \sigma)$: biến tỉ lệ Gaussian (variable scale Gaussian)

$I(x, y)$: Ảnh đầu vào

$*$ là phép nhân chập giữa x và y

$$\text{Và } G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad [2.2]$$

Để tìm những điểm đặc trưng có tính bất biến cao, thuật toán được sử dụng là tìm cực trị cục bộ của đạo hàm của hàm Gaussian viết tắt là DoG (Difference-of-Gaussian), kí hiệu là $D(x,y,\sigma)$. Hàm này được tính toán từ sự sai khác giữa 2 độ đo không gian cạnh nhau của một ảnh với tham số đo lệch nhau một hằng số k .

$$D(x,y,\sigma) = L(x,y,k\sigma) - L(x,y,\sigma) = (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y) \quad [2.3]$$

Các lý do lựa chọn hàm Gaussian là vì nó là kỹ thuật rất hiệu quả để tính toán L (cũng như làm tăng độ mịn của ảnh), mà L thì luôn phải được tính rất nhiều để mô tả đặc trưng trong không gian đo, và sau đó, D sẽ được tính một cách đơn giản chỉ với phép trừ ma trận điểm ảnh với chi phí thực hiện thấp.

Hơn nữa, đạo hàm của hàm Gaussian (DoG) có thể được sử dụng để tạo ra một sự xấp xỉ gần với đạo hàm bậc hai Laplace có kích thước chuẩn của hàm Gaussian ($\sigma^2 \nabla^2 G$) do tác giả Lindeberg đề xuất năm 1994. Ông đã chỉ ra rằng việc chuẩn hóa đạo hàm bậc hai với hệ số σ^2 là cần thiết cho bất biến đo trở nên đúng. Cụ thể, ông đã công bố rằng các giá trị cực đại và cực tiểu của $\sigma^2 \nabla^2 G$ chính là những giá trị có tính ổn định nhất (bất biến cao) so với một loạt các hàm đánh giá khác như : Gradient, Hessian hay Harris.

Mối quan hệ giữa D và $\sigma^2 \nabla^2 G$ được biểu diễn như sau:

$$\frac{\partial G}{\partial \sigma} = \sigma \nabla^2 G \quad [2.4]$$

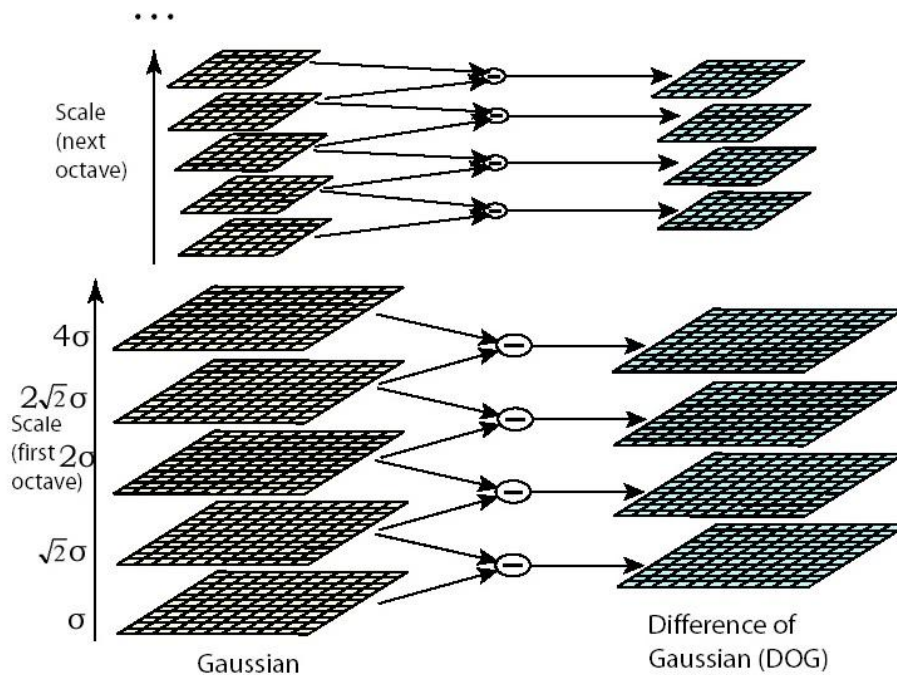
Như vậy, $\nabla^2 G$ có thể được tính thông qua việc xấp xỉ đạo hàm riêng $\partial G / \partial \sigma$ tại các tham số đo gần nhau $k\sigma$ và σ :

$$\sigma \nabla^2 G = \frac{\partial G}{\partial \sigma} \approx \frac{G(x,y,k\sigma) - G(x,y,\sigma)}{k\sigma - \sigma} \quad [2.5]$$

$$\text{Do đó : } G(x,y,k\sigma) - G(x,y,\sigma) \approx (k-1)\sigma^2 \nabla^2 G$$

Từ công thức này, ta thấy khi mà đạo hàm của Gaussian (DoG) được tính toán tại các tham số đo lệch nhau một hằng số k , thì ta có thể sử dụng DoG để xấp xỉ đạo hàm bậc hai Laplace của Gaussian. Vì hệ số $(k-1)$ trong phương trình trên là hằng số trong mọi không gian đo nên nó sẽ không ảnh hưởng đến việc tìm các vị trí cực trị. Sai số trong việc xấp xỉ đạo hàm bậc 2 tiến về 0 khi k gần với 1. Tuy nhiên, các kết quả thử nghiệm của tác giả cho thấy quá trình xấp xỉ đạo hàm không ảnh hưởng đến việc dò tìm các vị trí cực trị thậm chí ngay cả khi chọn k khá xa, ví dụ $k = \sqrt{2}$.

Như vậy, bước đầu tiên của giải thuật SIFT là phát hiện các điểm hấp dẫn với bộ lọc Gaussian ở các tỉ lệ khác nhau và các ảnh DoG từ sự khác nhau của các ảnh mờ.

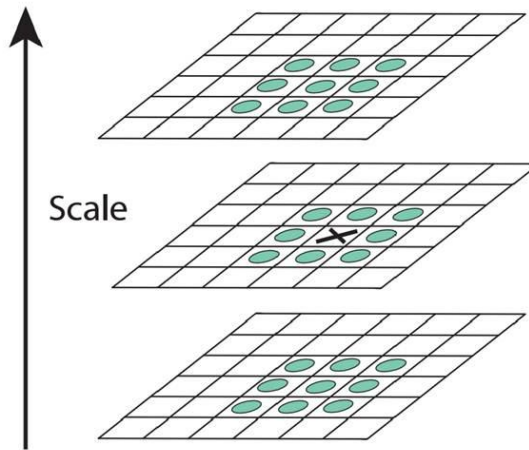


Hình 2.2 : Biểu đồ mô phỏng việc tính toán các ảnh DoG từ các ảnh mờ

Các ảnh cuộn được nhóm thành các nhóm tám. Giá trị của k được chọn sao cho số lượng ảnh mờ cho mỗi nhóm tám là cố định. Điều này đảm bảo cho số lượng các ảnh DoG cho mỗi nhóm tám không thay đổi.

Hai hình ảnh liên tục trong một nhóm tám được chọn để tính toán. Sau đó, cặp đôi tiếp theo được thực hiện, và lặp đi lặp lại quá trình. Điều này được thực hiện cho tất cả các nhóm tám. Các hình ảnh kết quả là một xấp xỉ bất biến tỷ lệ của Laplacian of Gaussian.

Sau khi áp dụng hàm DoG ta thu được các lớp kết quả khác nhau (scale) từ ảnh gốc, bước tiếp theo là tìm các cực trị trong các lớp kết quả theo từng miền cục bộ. Cụ thể là tại mỗi điểm trên các lớp kết quả sẽ được so sánh với 8 điểm lân cận trên cùng lớp và 9 điểm lân cận trên mỗi lớp khác (hình 2.3).



Hình 2.3 : Quá trình tìm điểm cực trị trong các hàm sai khác DoG

(X là điểm hiện tại, các vòng tròn màu xanh là các láng giềng của nó)

X sẽ được đánh dấu là điểm hấp dẫn nếu nó là cực đại hoặc cực tiểu của tất cả 26 láng giềng.

Thông thường, một vị trí không cực đại hoặc không cực tiểu sẽ không phải đi qua tất cả 26 kiểm tra. Một vài kiểm tra ban đầu thường là đủ để loại bỏ nó.

Vì số lượng các cực trị là rất lớn, vì vậy để tăng sự hiệu quả khi dò tìm các điểm cực trị (dò các điểm cực trị tốt nhất thay vì phải dò hết), ta cần xác định tần số lấy mẫu trong không gian đo và tần số lấy mẫu trong không gian quan sát (không gian ảnh). Thật không may là ta không thể xác định cả 2 loại tần số này một cách động trong mỗi tiến trình dò tìm. Thay vì vậy, các tần số này sẽ được xác định thông qua phương pháp thử nghiệm. Sau khi thử nghiệm với nhiều nguồn dữ liệu ảnh khác nhau, tác giả đã chỉ ra tần số lấy mẫu trong không gian đo tốt nhất là 3 (giữ lại 3 lớp trong mỗi bộ 8 lớp), và tần số lấy mẫu $\sigma = 1.6$.

2.2.2 Định vị điểm hấp dẫn:

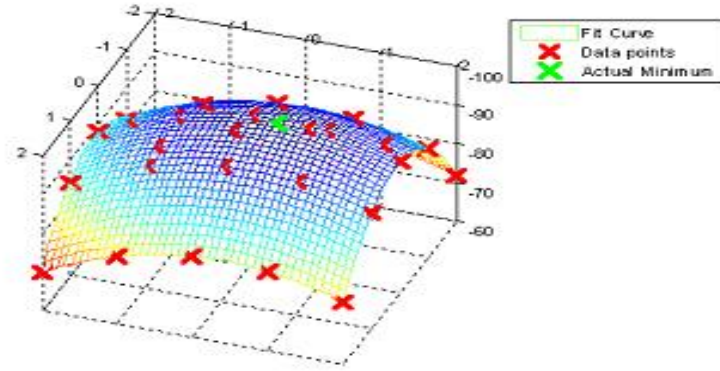
Sau bước 1 sẽ thu được rất nhiều điểm tiềm năng có thể làm điểm đặc biệt, tuy nhiên một số trong chúng là không cần thiết. ở bước tiếp theo này sẽ loại bỏ các điểm có độ tương phản kém (nhảy cảm với nhiễu) hoặc tính đặc trưng cục bộ ít hơn các điểm khác hoặc có xu hướng là đường biên đối tượng. Bước thực hiện này gồm 3 công đoạn :

a. Phép nội suy lân cận cho vị trí đúng của điểm tiềm năng:

Phép nội suy lân cận sử dụng mở rộng Taylor cho hàm Difference-of-Gaussian $D(x,y,\sigma)$:

$$D(X) = D + \frac{\partial D^T}{\partial X} X + \frac{1}{2} x^T \frac{\partial^2 D}{\partial^2 X} X \quad [2.6]$$

Trong đó: D và đạo hàm của nó được tính tại một điểm tiềm năng và $X = (x, y, \sigma)$ là khoảng cách từ điểm đó. Vị trí của điểm cực trị $\hat{X}$ được xác định bằng cách lấy đạo hàm của hàm trên với đối số X và tiến dần đến 0:



Hình 2.4 : Mô phỏng sử dụng công thức mở rộng của Taylor cho hàm DoG

$X = (x, y, \sigma)^T$: là độ dịch so với các điểm lân cận của điểm lấy mẫu

Vùng chứa điểm hấp dẫn được xác định qua $\hat{X}$

$$\hat{X} = - \frac{\partial^2 D^{-1}}{\partial X^2} \frac{\partial D}{\partial X} \quad [2.7]$$

Nếu $\hat{X} > 0.5$ theo một chiều nào đó thì nó có chỉ số cực trị không gần với các điểm tiềm năng khác, nó sẽ bị thay đổi và phép nội suy sẽ thay thế vai trò của nó bằng điểm khác gần nó. Thực hiện tiếp tục với các điểm lấy mẫu khác.

b. Loại trừ các điểm có tính tương phản kém :

Những điểm có $\hat{X}$ thỏa mãn (< 0.5) được thêm vào tập hợp mẫu tốt nhất, tiếp tục phân tích tiếp.

Dùng $D(\hat{X})$ để loại những điểm cực trị không ổn định (độ tương phản thấp).

Thay $\hat{X}$ vào $D(\hat{X})$ ta được:

$$D(\hat{X}) = D + \frac{1}{2} \frac{\partial D^T}{\partial X} \hat{X} \quad [2.8]$$

Nếu $D(\hat{X}) < 0.03$ thì điểm lấy mẫu đó sẽ bị loại.

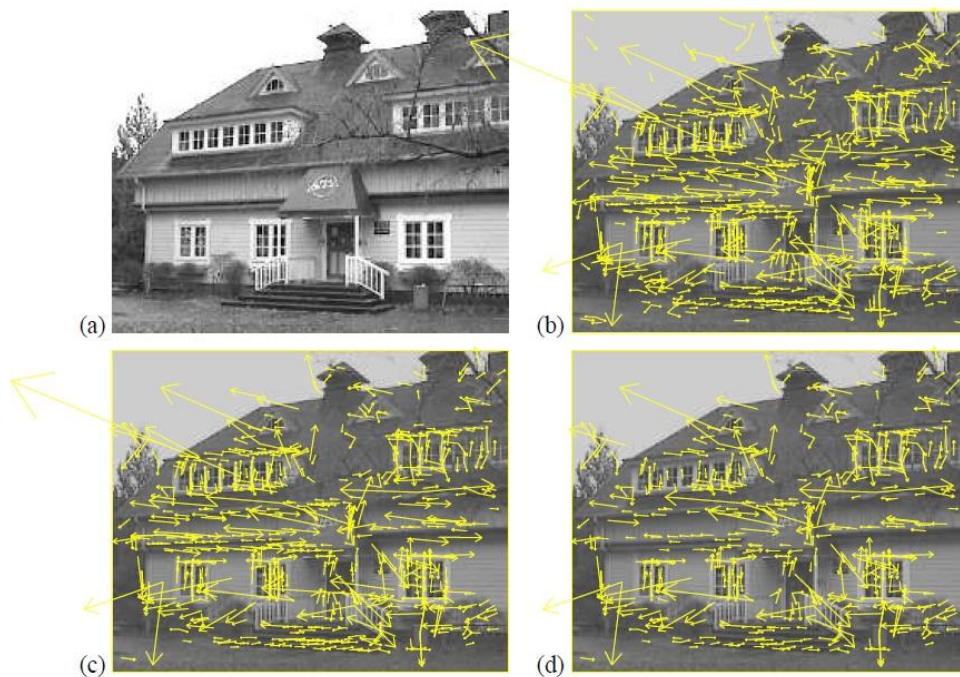
c. Loại bỏ các điểm dư thừa theo biên :

Sử dụng hàm DoG sẽ cho tác động mạnh đến biên khi vị trí của biên là khó xác định và vì vậy các điểm tiềm năng trên biên sẽ không bất biến và bị nhiễu. Và để tăng sự ổn định cho các điểm sẽ được chọn làm điểm đặc biệt ta sẽ loại trừ các điểm tiềm năng khó định vị (tức là vị trí dễ thay đổi khi có nhiễu do nằm ở biên).

Sau khi áp dụng hàm DoG sẽ làm đường biên ảnh không rõ ràng và độ cong chính sẽ có giá trị lớn hơn nhiều so với độ cong dọc theo biên vì vậy cần loại bỏ bớt các điểm đặc biệt dọc theo cùng một biên. Giải pháp cho việc này là sử dụng giá trị của ma trận Hessian cấp 2 :

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad [2.9]$$

Các giá trị riêng của H tỉ lệ thuận với độ cong của D, các phần tử của H là D_{xx} và D_{yy} .



Hình 2.5 : Quá trình lựa chọn các điểm hấp dẫn

a. Ảnh gốc, b. Các điểm hấp dẫn được phát hiện, c. Ảnh sau khi loại bỏ các điểm hấp dẫn có độ tương phản thấp, d. Ảnh sau loại bỏ các điểm hấp dẫn dọc theo cạnh.

Trong bước này số điểm hấp dẫn giảm, điều này giúp tăng hiệu quả và sự mạnh mẽ của thuật toán. Điểm hấp dẫn bị loại bỏ nếu nó có độ tương phản thấp hoặc nếu nó nằm trên một cạnh.

2.2.3 Xác định hướng cho các điểm hấp dẫn

Sau bước trên, chúng ta có những điểm hấp dẫn có độ ổn định. Chúng ta đã biết quy mô mà tại đó điểm hấp dẫn đã được phát hiện, vì vậy chúng ta có quy mô bất biến. Điều tiếp theo là xác định một định hướng cho mỗi điểm hấp dẫn.

Bằng cách chỉ định một hướng phù hợp cho mỗi điểm hấp dẫn dựa trên các thuộc tính hình ảnh cục bộ, dựa vào hướng của điểm hấp dẫn này ta có thể biết được điểm hấp dẫn bất biến với phép quay ảnh. Cách tiếp cận này trái ngược với các mô tả định hướng bất biến của Schmid and Mohr (1997). Bất lợi của cách tiếp cận đó là nó giới hạn mô tả có thể được sử dụng và loại bỏ thông tin hình ảnh bằng cách không yêu cầu tất cả các biện pháp cần phải dựa trên một quy trình phù hợp.

Sau khi thử nghiệm với một số phương pháp tiếp cận để xác định hướng, phương pháp sau đây đã được tìm thấy cho kết quả ổn định nhất. Tại mỗi điểm hấp dẫn người ta tính toán biểu đồ hướng Gradient trong vùng láng giềng của điểm hấp dẫn. Độ lớn và hướng của các điểm hấp dẫn được xác định theo công thức:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad [2.10]$$

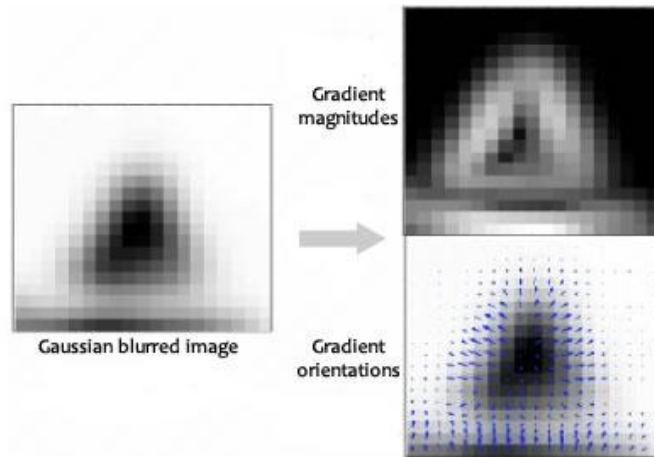
$$\theta(x, y) = \tan^{-1}((L(x, y+1) - L(x, y-1)) / (L(x+1, y) - L(x-1, y))) \quad [2.11]$$

Trong đó :

$m(x, y)$: Độ lớn của vector định hướng

$\theta(x, y)$: Hướng của vector định hướng (biểu diễn qua góc θ)

Một biểu đồ hướng được hình thành từ định hướng gradient của các điểm lấy mẫu trong một khu vực xung quanh các điểm hấp dẫn. Đỉnh trong biểu đồ hướng tương ứng với hướng chủ đạo của gradient. Đỉnh cao nhất trong biểu đồ được phát hiện, và sau đó bất kỳ điểm nào khác có cao điểm là 80% so với đỉnh cao nhất cũng được sử dụng cũng tạo ra một điểm hấp dẫn với định hướng đó. Vì vậy, đối với các địa điểm có nhiều đỉnh cường độ tương tự sẽ có nhiều điểm hấp dẫn tạo ra tại cùng một vị trí và tỷ lệ, nhưng có hướng khác nhau.



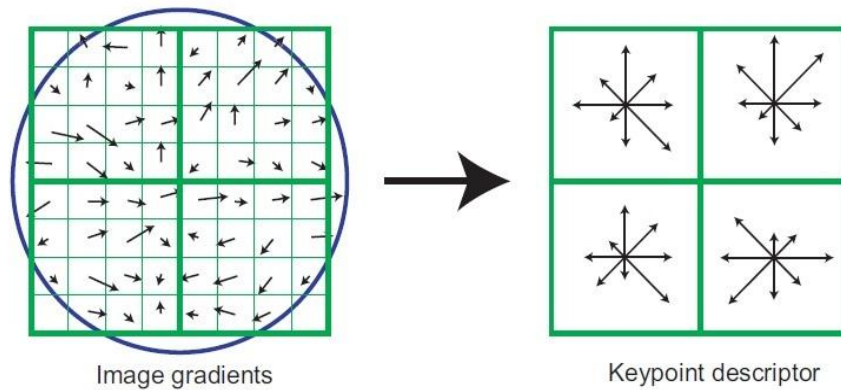
Hình 2.6: Tính độ lớn và hướng của Gradient

2.2.4 Mô tả các điểm hấp dẫn

Các phép xử lý trên đây đã thực hiện dò tìm và gán tọa độ, kích thước, và hướng cho mỗi điểm hấp dẫn. Các tham số đó yêu cầu một hệ thống tọa độ cục bộ 2 chiều có thể lặp lại được để mô tả vùng ảnh cục bộ và nhờ vậy tạo ra sự bất biến đối với các tham số đó. Bước tiếp theo đây sẽ tính toán một bộ mô tả cho một vùng ảnh cục bộ mà có tính đặc trưng cao (bất biến với các thay đổi khác nhau về độ sáng, thu - phóng ảnh, xoay).

Một cách tiếp cận đơn giản đó là lấy mẫu mật độ ảnh cục bộ lân cận điểm đặc trưng ở một độ đo thích hợp, và đối sánh các mật độ này sử dụng độ đo tương quan chuẩn. Tuy nhiên, hệ số tương quan đơn giản thì lại rất nhạy cảm với sự thay đổi mà gây ra sự đăng ký nhầm các mẫu, chẳng hạn như các biến đổi Affine, phối cảnh 3 chiều, hoặc bóp méo mềm.

Cách tiếp cận tốt hơn nhiều được đưa ra bởi Edelman, Intrator và Poggio (1997). Cách tiếp cận này dựa trên một mô hình thị giác sinh học, cụ thể là mô hình neuron phức tạp trong hệ thống não bộ. Các neuron sẽ tương ứng với một gradient tại một hướng và tần số không gian cụ thể, nhưng vị trí của gradient trên võng mạc được phép trượt trên một phạm vi nhỏ của khung nhìn.



Hình 2.7 : Tạo mạng lưu lược đồ định hướng

Điểm hấp dẫn sau khi được xác định hướng sẽ được biểu diễn dưới dạng các vector $4 \times 4 \times 8 = 128$ chiều (Số chiều = 8 hướng $\times$ (4×4) điểm hấp dẫn = 128 chiều) bằng cách tổng hợp các vector định hướng của các điểm trong khu vực, các vector này có đặc điểm :

- Chung gốc.
- Độ dài mỗi vector tương ứng độ lớn gradient m của nó.

2.3 Đối sánh đặc trưng SIFT

2.3.1 Độ đo khoảng cách và độ đo tương tự

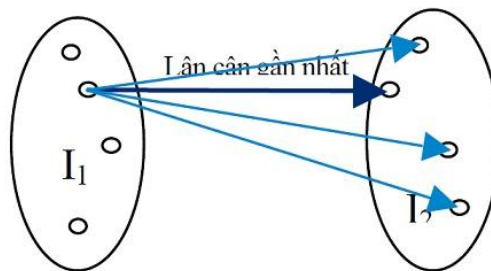
Độ đo tương tự là một trong những phương pháp tốt để máy tính phân biệt được các hình ảnh qua nội dung của chúng. Thông thường hệ thống tra cứu ảnh theo nội dung sẽ truy vấn hình ảnh bằng phương pháp đo tương tự dựa trên các đặc trưng, việc xác định nó có thể dưới nhiều hình thức như phát hiện biên,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

Do đó, độ đo có ý nghĩa quan trọng trong tra cứu ảnh dựa theo nội dung. Độ đo mang ý nghĩa quyết định kết quả tìm kiếm sẽ như thế nào, mức độ chính xác ra sao. Nhiều phép đo khoảng cách đã được khai thác trong việc tra cứu ảnh chúng bao gồm: khoảng cách Euclide, khoảng cách Cosin, khoảng cách giao nhau của biểu đồ histogram, khoảng cách Minkowski... Trong mục này, một vài phép đo khoảng cách sẽ được mô tả và ước lượng. Mục đích của việc ước lượng này để tìm ra một phép đo tương đồng cho các bộ mô tả ước lượng hình dạng khác nhau.

2.3.2 Đối sánh đặc trưng cục bộ bất biến

2.3.2.1 Đối sánh các vector đặc trưng

Trước hết để đối sánh các ảnh với nhau thì cần trích xuất tập keypoint tương ứng từ mỗi ảnh bằng các bước đã chỉ ra ở trên. Sau đó việc đối sánh sẽ thực hiện trên các tập keypoint này. Bước chính trong kĩ thuật đối sánh sẽ thực hiện tìm tập con keypoint so khớp nhau ở hai ảnh, để thực hiện việc này sẽ tìm các cặp keypoint trùng nhau lần lượt ở hai ảnh. Tập con các keypoint so khớp chính là vùng ảnh tương đồng. Việc đối sánh hai tập hợp điểm đặc trưng quy về bài toán tìm láng giềng gần nhất của mỗi điểm đặc trưng (hình 2.8).



Hình 2.8 : Đối sánh hai ảnh quay về đối sánh hai tập điểm đặc trưng trong không gian đặc trưng

Có 2 vấn đề cần được quan tâm :

- Tổ chức tập hợp điểm cho phép tìm kiếm láng giềng một cách hiệu quả
- Việc đối sánh phải đạt độ chính xác nhất định

Một phương pháp được đề xuất bởi D. Mount cho phép tìm kiếm nhanh các điểm lân cận được sử dụng[7], ANN là viết tắt của Approximative Nearest Neighbour. Nó cho phép tổ chức dữ liệu dưới dạng *kd-tree*, việc tìm kiếm láng giềng gần nhất mang tính xấp xỉ trên *kd-tree*. Cụ thể là hai điểm trong không gian đặc trưng được coi là giống nhau nếu khoảng cách Euclidean giữa hai điểm là nhỏ nhất và tỉ số giữa khoảng cách gần nhất với khoảng cách gần nhì phải nhỏ hơn 1 ngưỡng cho trước.

Giả sử cặp keypoint có bộ mô tả lần lượt là:

$$A = (a_1, a_2, a_3, \dots, a_{128}) \text{ và } B = (b_1, b_2, b_3, \dots, b_{128})$$

Thì khoảng cách Euclidean giữa A và B được tính bằng công thức:

$$D(A, B) = \sum (a_i - b_i)^2 \quad [2.12]$$

2.3.2.2 Một số độ đo tương đồng cho ảnh sử dụng đặc trưng SIFT

- Độ đo Cosin :

$$d(x, y) = \frac{x \cdot y}{\|x\| \cdot \|y\|} \quad [2.13]$$

- Khoảng cách góc :

$$d(x, y) = \cos^{-1}(x \cdot y) \quad [2.14]$$

- Độ đo Euclidean :

$$d(x, y) = \sqrt{\sum_{i=1}^n |x_i - y_i|^2} \quad [2.15]$$

- Độ đo Jensen-Shannon divergence :

$$d_{JSD}(H, H') = \sum_{m=1}^M H_m \log \frac{2H_m}{H_m + H'_m} + H'_m \log \frac{2H'_m}{H'_m + H_m} \quad [2.16]$$

Với H, H' là 2 biểu đồ biểu diễn các vector đặc trưng SIFT

CHƯƠNG 3: Thực nghiệm

3.1 Môi trường và các công cụ sử dụng trong thực nghiệm

- Cấu hình phần cứng

Bảng 3.1 : Cấu hình phần cứng sử dụng trong thực nghiệm

Thành phần	Thông số
CPU	E2180 2x2.0 GHz
RAM	2GB
Hệ điều hành	Windows7 32bit Service Pack 1
Bộ nhớ	160 GB

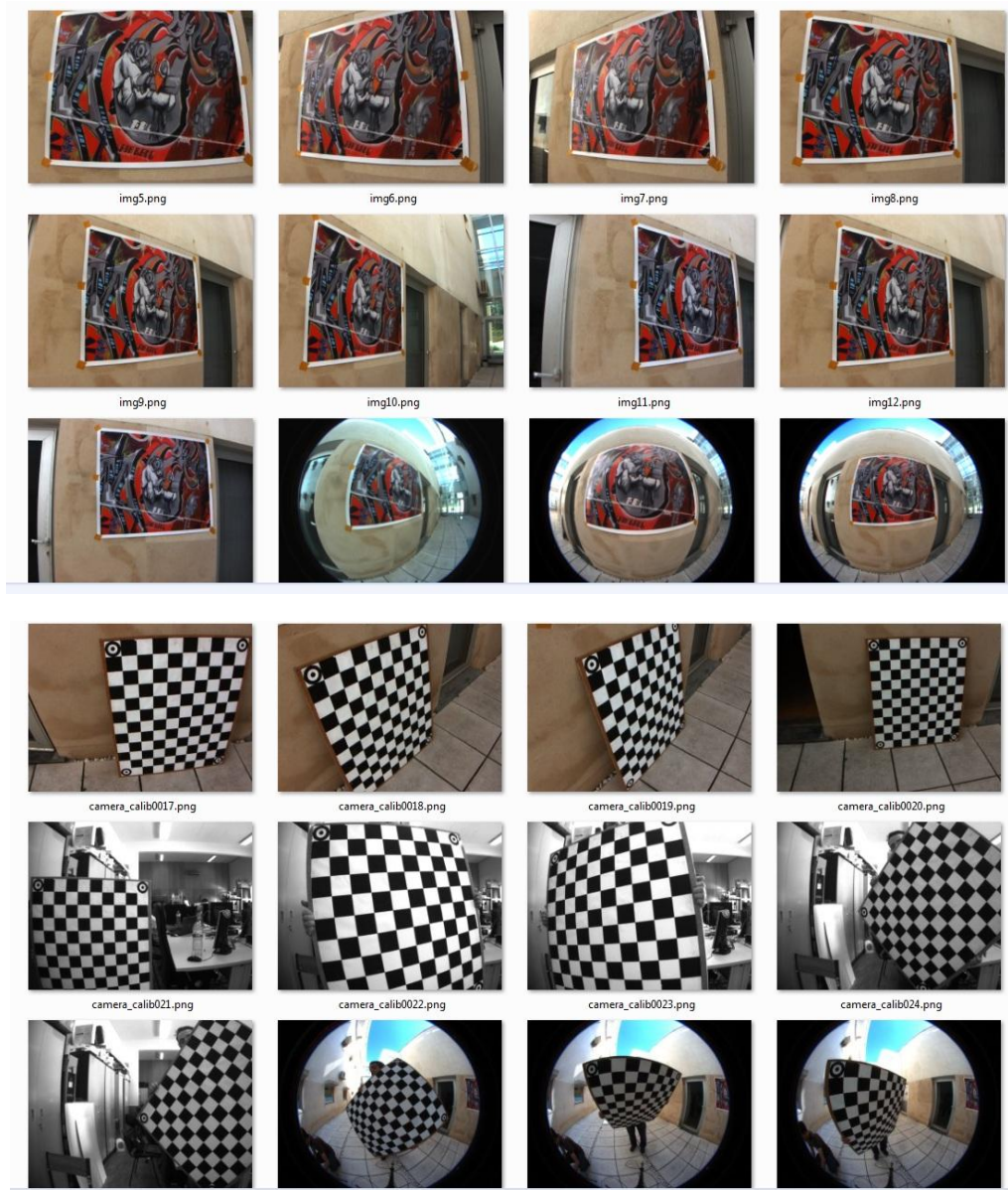
- Công cụ phần mềm sử dụng

Bảng 3.2 : Công cụ phần mềm sử dụng trong thực nghiệm

STT	Tên phần mềm	Nguồn
1	Matlab 7.0	http://www.mathworks.com/products/

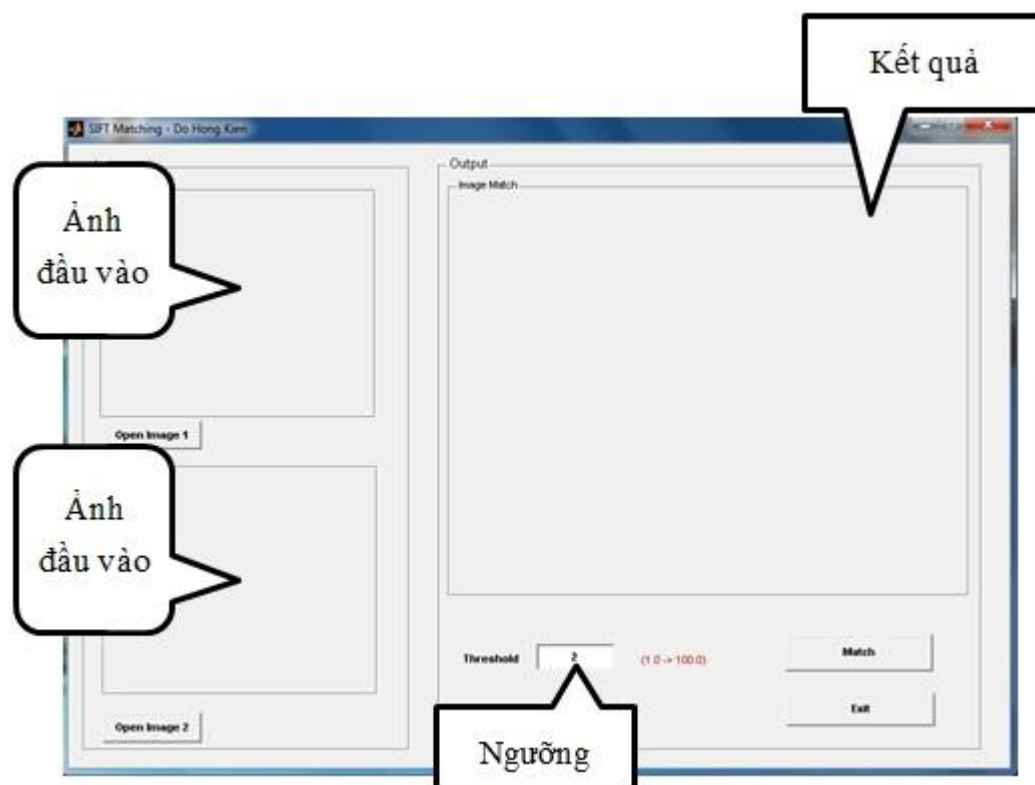
3.2 Xây dựng tập dữ liệu ảnh

Tập dữ liệu ảnh gồm 30 ảnh của đối tượng bị thay đổi về độ chiếu sáng, góc nhìn, co dãn, bị che lấp 1 phần...Các ảnh có kích thước từ 100x100 đến 800x600 (pixel) với nhiều chủ đề khác nhau.

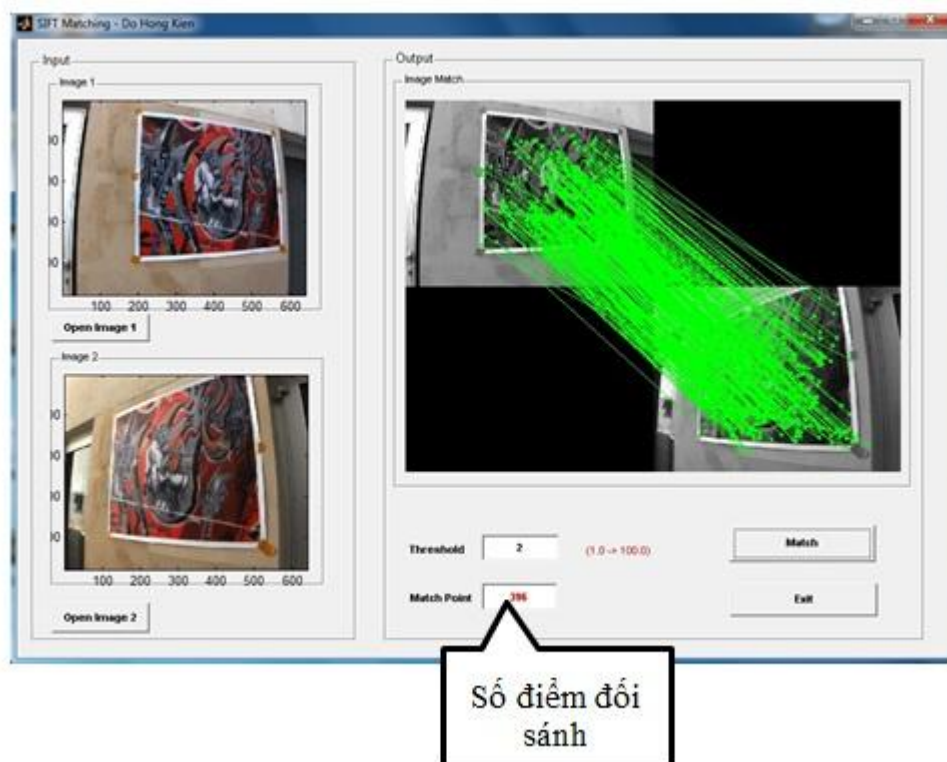


Hình 3.1 : Tập ảnh dữ liệu sử dụng thực nghiệm

3.3 Giao diện chương trình



Hình 3.2 : Giao diện chương trình

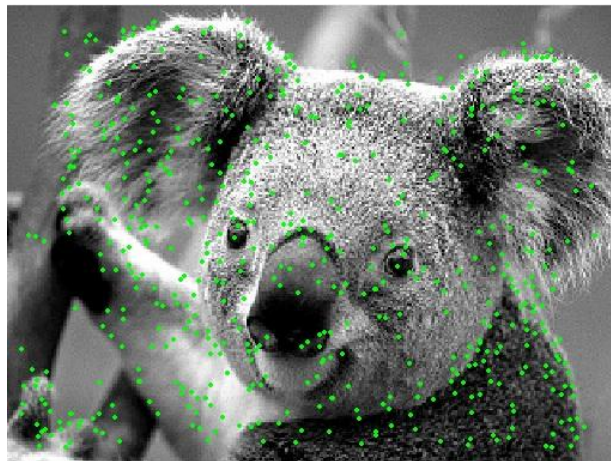


Hình 3.3 : Giao diện đối sánh

3.4 Một số kết quả



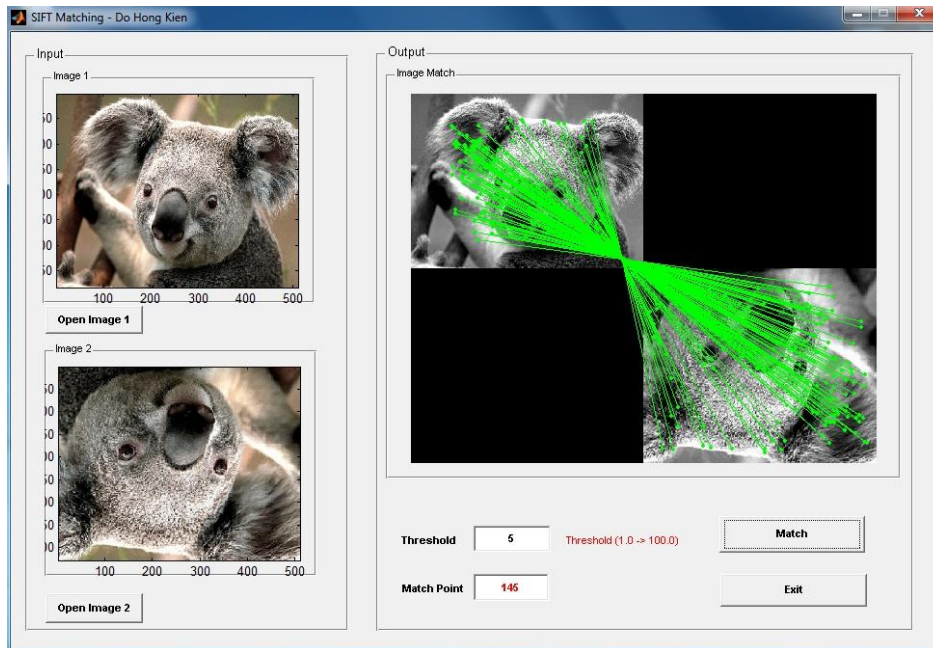
Hình 3.4 : Ảnh gốc



Hình 3.5 : Các điểm hấp dẫn đã được phát hiện

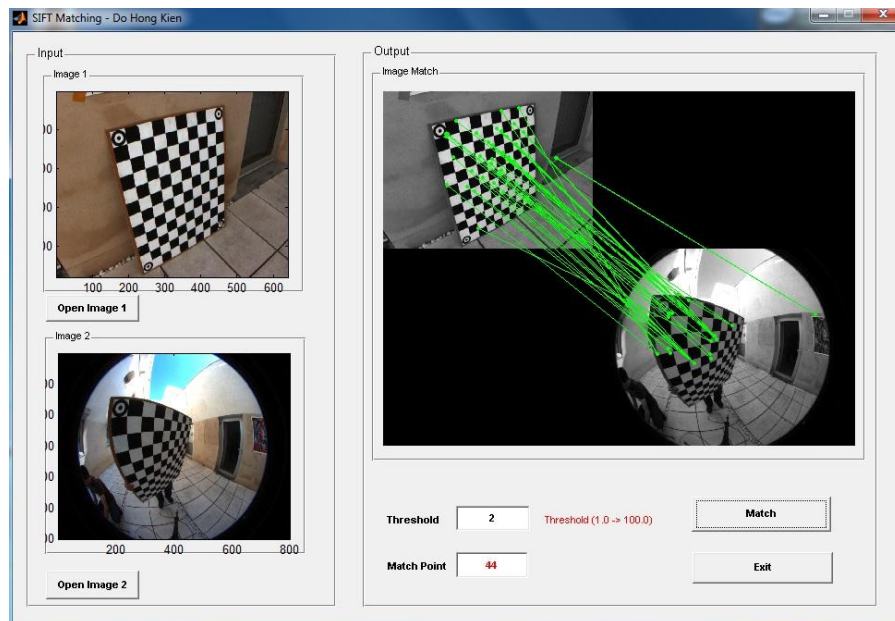


Hình 3.6: Các điểm hấp dẫn đã được gán hướng



Hình 3.7 : Thực hiện đối sánh 2 ảnh

Với ảnh bị biến đổi bởi phép xoay, co giãn và che lấp một phần, chương trình đã đối sánh chính xác 145 cặp điểm đặc trưng. Điều này cho thấy SIFT bất biến với phép xoay, thu phóng và không yêu cầu tính toàn vẹn của ảnh.



Hình 3.8 : Với ảnh chụp bằng ống kính Fisheye chương trình vẫn nhận ra chính xác đối tượng

KẾT LUẬN

Có thể nói SIFT là một thuật toán rất mạnh và phức tạp trong lớp các bài toán đối sánh ảnh. Trong đề án này em đã tìm hiểu và cài đặt thuật toán với đầy đủ các bước cơ bản của SIFT và xây dựng chương trình ứng dụng mô phỏng việc đối sánh ảnh tương tự sử dụng SIFT. Thuật toán SIFT được cài đặt bước đầu cho kết quả tốt với tập dữ liệu thử nghiệm.

Tuy nhiên, tra cứu ảnh là chủ đề phức tạp, cộng với khả năng và kinh nghiệm còn hạn chế nên em còn gặp một số khó khăn trong việc tìm hiểu nghiên cứu các kỹ thuật tra cứu ảnh.

Vì vậy em rất mong nhận được sự đóng góp ý kiến quý báu của các thầy cô giáo cũng như bạn bè để đề án của em được hoàn thiện hơn.

Em xin chân thành cảm ơn!

TÀI LIỆU THAM KHẢO

- [1] Bùi Thị Thúy Nga, *Tìm hiểu một số phương pháp trích chọn đặc trưng và ứng dụng tra cứu ảnh dựa trên nội dung*, đồ án tốt nghiệp CNTT 2011, ĐHDL Hải Phòng.
- [2] Đồng Nam Hà, *Tra cứu ảnh dựa trên đặc trưng kết cấu*, đồ án tốt nghiệp ngành CNTT 2010, ĐHDL Hải Phòng.
- [3] Nguyễn Thị Hoàn, *Phương pháp trích chọn đặc trưng ảnh trong thuật toán học máy tìm kiếm ảnh áp dụng vào bài toán tìm kiếm sản phẩm*, Đại học quốc gia Hà Nội.
- [4] Trần Thị Thanh Hải, Eric Marchand, *Một số phương pháp đối sánh ảnh thời gian thực*, Trung tâm MICA, Đại học Bách Khoa Hà Nội.
- [5] AI Shack, *Scale Invariant Feature Transform*, www.aishack.in
- [6] Andrea Vedaldi, *An implementation of SIFT detector and descriptor*, University of California at Los Angeles.
- [7] David G. Lowe, *Distinctive Image Features from Scale-Invariant Keypoints*, Computer Science Department, University of British Columbia.
- [8] Faraj Alhwarin, Chao Wang, Danijela Risti -Durrant, Axel Gräser, *Improved SIFT-Features Matching for Object Recognition*, Institute of Automation, University of Bremen.