

---

**MỤC LỤC**

<b>MỤC LỤC</b>	<b>1</b>
<b>MỞ ĐẦU</b>	<b>3</b>
<b>CHƯƠNG 1: TỔNG QUAN VỀ TRA CỨU ẢNH DỰA TRÊN NỘI DUNG..</b>	<b>4</b>
1.1 Tra cứu thông tin trực quan.....	4
1.2 Những thành phần cơ bản của một hệ thống tra cứu ảnh dựa trên nội dung 4	
1.3 Các phương pháp tra cứu ảnh dựa trên nội dung .....	5
1.3.1 Phương pháp trích chọn theo màu sắc .....	5
1.3.2 Kết cấu .....	7
1.3.3 Phương pháp trích chọn đặc trưng theo hình dạng.....	15
1.3.4 Độ đo khoảng cách và độ đo tương tự.....	18
1.4 Đánh giá hiệu năng hệ thống tra cứu ảnh .....	21
1.5 Các hệ thống tra cứu ảnh dựa trên nội dung .....	23
1.5.1 Hệ thống QBIC (Query By Image Content).....	23
1.5.2 Hệ thống Photobook .....	23
1.5.3 Hệ thống VisualSEEK và WebSEEK.....	24
1.5.4 Hệ thống RetrievalWare .....	24
1.5.5 Hệ thống Imatch .....	24
<b>CHƯƠNG 2: BIỂU DIỄN HÌNH DẠNG DỰA TRÊN VÙNG.....</b>	<b>25</b>
2.1 Giới thiệu .....	25
2.2 Biểu diễn hình dạng dựa trên vùng .....	25
2.2.1 Phương pháp toàn cục (Global Method) .....	26
2.2.2 Phương pháp cấu trúc (Structural methods) .....	30
2.3 Phương pháp tìm xương theo DCE.....	31
2.3.1 Ý tưởng chính của phương pháp .....	31
2.3.2 Rời rạc hóa đường cong.....	31
2.3.3 Phương pháp cắt tia xương với DCE.....	32
<b>CHƯƠNG 3: ĐỐI SÁNH SHAPE DỰA TRÊN ĐẶC TRƯNG VÙNG.....</b>	<b>34</b>
3.1 Đối sánh dựa trên sự tương tự của đồ thị xương.....	34
3.1.1 Đồ thị xương (Skeleton Graphs).....	36

---

3.1.2	Đối sánh các đồ thị xương (Matching the Skeleton Graphs) .....	36
3.1.3	Tối ưu dãy song ánh (Optimal subsequence bijection) .....	40
3.2	Đối sánh đồ thị xương dựa trên các điểm quan trọng bằng cách sử dụng đường dẫn tương tự.....	42
3.2.1	Giới thiệu .....	42
3.2.2	Sát nhập các nút giao nhau (Mergence of junction nodes).....	42
3.2.3	Đối sánh các nút quan trọng (Matching Critical Nodes).....	43
<b>CHƯƠNG 4: THỰC NGHIỆM.....</b>		<b>45</b>
4.1	Môi trường thực nghiệm .....	45
4.2	Một số kết quả.....	45
4.2.1	Thay đổi kích thước ảnh mẫu .....	45
4.2.2	Xoay ảnh mẫu một góc $\alpha$ .....	46
4.2.3	Đồng thời thay đổi kích thước và góc xoay của ảnh .....	46
4.3	Một số nhận xét về chương trình .....	47
<b>KẾT LUẬN .....</b>		<b>48</b>
<b>TÀI LIỆU THAM KHẢO .....</b>		<b>49</b>

---

## MỞ ĐẦU

Cùng với sự phát triển của công nghệ thông tin, lượng hình ảnh được số hóa là rất lớn và đang tăng lên nhanh chóng. Một số lượng lớn ảnh đang được sử dụng trong các thư viện ảnh số và trên Internet. Vì vậy nhu cầu tìm kiếm ảnh là một nhu cầu tất yếu.

Trước đây, người ta thường sử dụng cách tra cứu theo văn bản (Text Based Image Retrieval). Tuy nhiên việc tìm kiếm chỉ dựa vào văn bản đi kèm ảnh còn có nhiều kết quả không phù hợp với mong muốn giữa nội dung văn bản truy vấn và nội dung ảnh trả về. Những công cụ tìm kiếm ảnh theo nội dung của các bức ảnh ra đời và tỏ ra ưu thế vì hạn chế được những nhược điểm trên. Tìm kiếm ảnh dựa trên nội dung đang nhận được sự quan tâm của nhiều nhà khoa học, nhiều công trình nghiên cứu về phương pháp này được đăng trên các tạp chí như: International Journal of Computer Vision, IEEE conference...

Nội dung của đề tài này trình bày về “Một số phương pháp đối sánh Shape”, là một phần nằm trong phương pháp tìm kiếm ảnh dựa theo nội dung. Phương pháp này sử dụng đồ thị xương và các điểm quan trọng của ảnh (điểm cuối) để ứng dụng vào bài toán tìm kiếm ảnh.

Nội dung của đề tài bao gồm bốn chương:

- Chương 1: Tổng quan về tra cứu ảnh dựa trên nội dung
- Chương 2: Biểu diễn hình dạng dựa trên vùng
- Chương 3: Đối sánh Shape dựa trên đặc trưng vùng
- Chương 4: 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.

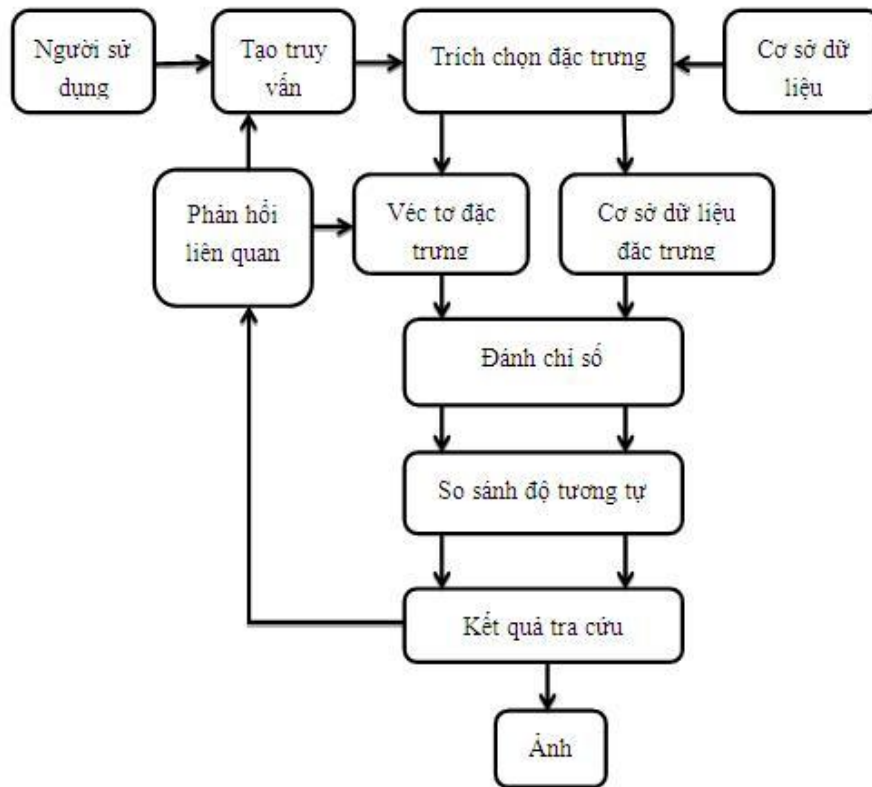
### 1.2 Những thành phần cơ bản của một hệ thống tra cứu ảnh dựa trên nội dung

Trong mô hình trong Hình 1.1 người sử dụng sẽ tạo truy vấn bằng cách chọn một ảnh mẫu hoặc phác thảo một hình vẽ mô tả đối tượng ảnh cần tìm.

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

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

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



Hình 1.1 Các thành phần của hệ thống tra cứu ảnh dựa trên nội dung

### 1.3 Các phương pháp tra cứu ảnh dựa trên nội dung

#### 1.3.1 Phương pháp trích chọn theo 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.

##### 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 mức. Việc sử dụng biểu đồ màu toàn cục một ảnh sẽ được mã hóa với biểu đồ màu của nó và khoảng cách giữa hai ảnh sẽ được xác định bởi khoảng cách giữa những biểu đồ

màu của chúng. Với kỹ thuật này chúng ta có thể sử dụng các độ đo khác nhau để tính toán khoảng cách giữa hai biểu đồ màu.

Đây 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 sắc của các vùng. Do đó, khoảng cách giữa các ảnh đôi khi không thể chỉ ra được sự khác nhau thực sự giữa chúng.

### 1.3.1.2 Biểu đồ màu cục bộ (Local Color Histogram)

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

$$d(Q, I) = \sum_{k=1}^M \sqrt{\sum_{i=1}^N (H_Q^k[i] - H_I^k[i])^2} \quad (1.1)$$

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

### 1.3.1.3 Vector liên kết màu

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  $\alpha_i$  biểu thị số các pixel gắn kết trong ngăn thứ  $i$  và  $\beta_i$  biểu thị số các pixel không gắn kết trong một ảnh thì vector liên kết màu của một ảnh được định nghĩa bằng vector  $\langle (\alpha_1, \beta_1), (\alpha_2, \beta_2), (\alpha_3, \beta_3), \dots, (\alpha_M, \beta_M) \rangle$ . Trong đó:  $\langle (\alpha_1 + \beta_1), (\alpha_2 + \beta_2), \dots, (\alpha_M + \beta_M) \rangle$  là lược đồ màu của ảnh.

Việc thông tin không gian được đối sánh 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.

#### 1.3.1.4 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:

$$\gamma_{i,j}^{(k)} = \Pr[p_2 \in I_{c(j)} | |p_1 - p_2| = k] \quad (1.2)$$

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.

#### 1.3.2 Kết cấu

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

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

### 1.3.2.1 Các đặc trưng Tamura

Các đặc trưng Tamura bao gồm độ thô, độ tương phản, hướng, giống nhất, tính chất đều và nhám, được thiết kế phù hợp với các nghiên cứu tâm lý về nhận thức của người đối với kết cấu. Trong đó, độ thô, độ tương phản, hướng được sử dụng trong một số hệ thống tra cứu ảnh nổi tiếng như QBIC và Photobook.

- Độ thô ( Coaseness):

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

$$A_k(x,y) = \frac{\sum_{i=x-2^{k-1}}^{x+2^{k-1}-1} \sum_{j=y-2^{k-1}}^{y+2^{k-1}-1} g(i,j)}{2^{2k}} \quad (1.3)$$

Trong đó,  $g(i,j)$  là cường độ pixel tại  $(i,j)$ .

Sự khác nhau giữa các cặp trung bình động không theo hướng ngang và đứng cho mỗi pixel được tính toán, đó là:

$$E_{k,h}(x,y) = |A_k(x+2^{k-1},y) - A_k(x-2^{k-1},y)| \quad (1.4)$$

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

Giá trị của  $k$  cực đại hóa  $E$  theo một trong hai hướng được sử dụng để đặt cỡ tốt nhất cho mỗi pixel, đó là:

$$S_{best} = \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n S_{best}(i,j) \quad (1.5)$$

Cải tiến của đặc trưng thô có thể thu được bởi sử dụng một lược đồ để mô tả phân bố của  $S_{best}$ . Đã làm tăng đáng kể hiệu năng tra cứu và làm cho đặc trưng



có khả năng xử lý với một ảnh hoặc vùng có đa đặc tính kết cấu. Do vậy, nó là hữu ích hơn đối với các ứng dụng tra cứu ảnh.

-Độ tương phản:

Công thức cho độ tương phản là:

$$F_{\text{con}} = \frac{\sigma}{\alpha_4^{1/4}} \quad (1.6)$$

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

$\mu_4$ : là moment thứ tư về trung bình.

$\sigma^4$ : là phương sai.

-Độ định hướng:

Để tính toán hướng ta sử dụng hai mảng 3x3 và một vector gradient tại mỗi điểm ảnh được tính toán.

Độ lớn và góc của vector được định nghĩa như sau:

$$|\Delta G| = (|\Delta_H| + |\Delta_V|) / 2 \quad (1.7)$$

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

Trong đó,  $\Delta_H$  và  $\Delta_V$  là các khác biệt ngang và dọc của chập.

Sau đó bằng lượng hóa  $\theta$  và đếm số các pixel với độ lớn tương ứng  $|\Delta G|$  lớn hơn một ngưỡng, một lược đồ của  $\theta$ , biểu thị bằng  $H_D$ , có thể được xây dựng. Lược đồ này sẽ cho biết các đỉnh bền vững cho các ảnh hướng cao và sẽ là tương đối phẳng với các ảnh không có hướng bền vững. Toàn bộ lược đồ được tóm lược để thu toàn bộ độ đo hướng dựa trên tính nhọn của các đỉnh:

$$F_{\text{dir}} = \sum_p^{n_p} \sum_{\phi \in w_p} (\phi - \phi_p)^2 H_D(\phi) \quad (1.8)$$

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

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

$\phi_p$  là bin màu nhận giá trị đỉnh.

### 1.3.2.2 Các đặc trưng Wold

Phân rã Wold cung cấp một cách tiếp cận khác để mô tả các kết cấu về mặt các đặc tính tri giác. Ba thành phần Wold, điều hòa, độ phai mờ và độ bất định tương ứng với chu kỳ, hướng và tính ngẫu nhiên của kết cấu tương ứng. Các kết cấu chu kỳ có thành phần độ điều hòa cao, các kết cấu có tính định hướng cao có kết cấu độ phai mờ lớn, và các kết cấu được cấu trúc kém hơn có thành phần độ bất định lớn hơn.

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

$$y(m,n) = u(m,n) + d(m,n) = u(m,n) + h(m,n) + e(m,n) \quad (1.9)$$

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

$d(m,n)$  là thành phần tiên định.

$e(m,n)$  là thành phần tạm thời.

$h(m,n)$  là thành phần điều hòa.

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

Trong miền tần số ta có:

$$F_y(\xi, \eta) = F_u(\xi, \eta) + F_d(\xi, \eta) = F_u(\xi, \eta) + F_h(\xi, \eta) + F_e(\xi, \eta) \quad (1.10)$$

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

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

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

### 1.3.2.3 Các đặc trưng lọc Gabor

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

Hàm Gabor hai chiều  $g(x,y)$  được định nghĩa:

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right) + 2\pi j W x\right] \quad (1.11)$$

Trong đó:  $\sigma_x$  là độ lệch chuẩn của các bao Gaussian dọc theo hướng x.

$\sigma_y$  là độ lệch chuẩn của các bao Gaussian dọc theo hướng y.

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

$$\begin{aligned} g_{mn}(x,y) &= a^{-m} g(x',y') \\ x' &= a^{-m}(-x\cos\theta + y\sin\theta) \\ y' &= a^{-m}(-x\sin\theta + y\cos\theta) \end{aligned} \quad (1.12)$$

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

$K$  và  $S$  là số các hướng và các tỷ lệ co giãn.

$a^{-m}$  là hệ số co giãn để đảm bảo rằng năng lượng là độc lập của  $m$ .

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

$$W_{mn}(x,y) = \int I(x,y) g_{mn}^*(x - x_1, y - y_1) dx_1 dy_1 \quad (1.13)$$

Trong đó: \* chỉ ra số liên hợp phức.

$\mu_{mn}$  là trung bình.

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

Nghĩa là:  $f = [\mu_{00}, \sigma_{00}, \dots, \mu_{mn}, \sigma_{mn}, \Delta, \mu_{s-1k-1}, \sigma_{s-1k-1}]$  có thể được sử dụng để biểu diễn đặc trưng kết cấu của một vùng kết cấu thuần nhất.

#### 1.3.2.4 Các đặc trưng biến đổi sóng

Tương tự với lọc Gabor, phương pháp biến đổi dạng sóng là cách phân tích và phân loại áp dụng cho các kết cấu nhiều chiều. Các biến đổi sóng phân tích một tín hiệu thành một họ các hàm cơ sở  $\psi_{mn}(x)$  thu được thông qua dịch chuyển và sự co giãn của sóng  $\psi(x)$ :

$$\psi_{mn}(x) = 2^{-m/2} \psi(2^{-m}x-n) \quad (1.14)$$

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

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

$$f(x) = \sum_{m,n} c_{mn} \psi_{mn}(x) \quad (1.15)$$

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

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

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

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

Khi thực hiện phân tích mức thì PWT cho kết quả là một vector đặc trưng có  $3 \times 4 \times 2$  thành phần. Đối với TWT, vector đặc trưng phụ thuộc vào thứ tự phân tích các dải tần số con. Có thể xây dựng được một cây phân tích cố định bằng

cách phân tích tuần tự các dải tần LL, LH và HH, kết quả cho ra sẽ là một vector đặc trưng có 52x2 thành phần.

Trong trường hợp này thì vector đặc trưng kết quả của phân tích PWT chỉ là tập con của vector do phân tích TWT sinh ra. Ngoài ra qua so sánh sự khác nhau của vector đặc trưng thu được khi sử dụng các phương pháp biến đổi dạng sóng khác nhau, người ta thấy rằng việc lựa chọn bộ lọc dạng sóng không ảnh hưởng lớn lắm đến các phân tích kết cấu ảnh.

### 1.3.2.5 Ma trận đồng khả năng (Co-occurrence matrix)

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

Giả sử có một toán tử tại vị trí  $P(i,j)$ ,

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

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

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

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

$$C_t(a,b) = \text{card}\{(s,s+t) \in R^2 \mid A[s] = a, A[s+t] = b\} \quad (1.16)$$

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

Ví dụ, với một ảnh có 8 mức xám và vector  $t$  là một dịch chuyển một vị trí thì với ảnh mẫu:

---

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

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

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

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

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

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

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

$$\text{Độ tương phản: } \sum_i \sum_j (i - j)^2 C(i, j) \quad (1.17)$$

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

Ngoài ra còn có thể xây dựng được nhiều đặc trưng khác như độ tương quan, phương sai, tổng trung bình, tổng phương sai, tổng entropy, trung vị cục bộ... Như vậy với mỗi đặc trưng kết cấu chúng ta thu được một ma trận đồng khả

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

### 1.3.3 Phương pháp trích chọn đặc trưng theo hình dạng

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

#### 1.3.3.1 Các bất biến moment

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

$$\mu_{p,q} = \sum_{(x,y) \in R} (x - x_c)^p (y - y_c)^q \quad (1.18)$$

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

Moment trung tâm này có thể được chuẩn hóa để bất biến tỉ lệ:

$$\eta_{p,q} = \frac{\mu_{p,q}}{\mu_{0,0}^{\frac{p+q+2}{2}}}, \quad \gamma = \frac{p+q+2}{2} \quad (1.19)$$

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

$$\phi_1 = \mu_{2,0} + \mu_{0,2} \quad (1.20)$$

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

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

$$\Phi_4 = (\mu_{3,0} + \mu_{1,2})^2 + (\mu_{0,3} + \mu_{2,1})^2$$

$$\Phi_5 = (\mu_{3,0} - 3\mu_{1,2})(\mu_{3,0} + \mu_{1,2})[(\mu_{3,0} + \mu_{1,2})^2 - 3(\mu_{0,3} + \mu_{2,1})^2] \\ + (\mu_{0,3} - 3\mu_{2,1})(\mu_{0,3} + \mu_{2,1})[(\mu_{0,3} + \mu_{2,1})^2 - 3(\mu_{3,0} + \mu_{1,2})^2]$$

$$\Phi_6 = (\mu_{2,0} - \mu_{0,2})[(\mu_{3,0} + \mu_{1,2})^2 - (\mu_{3,0} + \mu_{2,1})^2] \\ + 4\mu_{1,1}(\mu_{3,0} + \mu_{1,2})(\mu_{0,3} + \mu_{2,1})$$

$$\Phi_7 = (3\mu_{2,1} - \mu_{0,3})(\mu_{3,0} + \mu_{1,2})[(\mu_{3,0} + \mu_{1,2})^2 - 3(\mu_{0,3} + \mu_{2,1})^2]$$

### 1.3.3.2 Các góc uốn

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

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

$$y'_s = \frac{dy_s}{ds} \tag{1.21}$$

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

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

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

$$d_p(A, B) = (\min_{\omega \in R, t \in [0, 1]} \int_0^1 |\theta_A(s+t) - \theta_B(s) + \omega|^p ds)^{1/p} \tag{1.22}$$

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



### 1.3.3.3 Mô tả Fourier

Mô tả Fourier mô tả hình của một đối tượng với biến đổi Fourier của đường biên của đối tượng. Một lần nữa ta lại coi biên của một đối tượng ảnh 2D là dãy liên tiếp các điểm ảnh nằm bên cạnh nhau  $(x_s, y_s)$ . Trong đó,  $0 \leq s \leq N - 1$  và  $N$  là tổng số các pixel trên đường bao.

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

$$K(s) = \frac{d}{ds} \theta(s) \quad (1.23)$$

Với  $K(s)$  là độ cong,  $\theta(s)$  là hàm xoay của đường biên.

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

$$R(s) = \sqrt{(x_s - x_c)^2 + (y_s - y_c)^2} \quad (1.24)$$

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

$$Z(s) = (x_s - x_c) + j(y_s - y_c) \quad (1.25)$$

Các biến đổi Fourier của ba loại biểu diễn này của đường biên sinh ra ba tập hệ số phức hợp, biểu diễn hình của một đối tượng trong miền tần số. Các hệ số tần số thấp hơn mô tả đặc tính tổng quan của hình dạng trong khi các hệ số tần số cao phản ánh các chi tiết của hình. Để không bị phụ thuộc vào góc quay của đối tượng (Tức là việc mã hóa là không bị ảnh hưởng bởi sự lựa chọn của các điểm tham chiếu) thì ta chỉ sử dụng biên độ của các hệ số phức và bỏ qua thành phần pha.

Để không bị phụ thuộc vào tỷ lệ thì ta phải chia biên độ của các hệ số phức cho biên độ của thành phần một chiều DC hay là cho hệ số khác 0 đầu tiên. Bản thân các cách biểu diễn đường biên này đã không phụ thuộc vào sự tịnh tiến của đối tượng ảnh.

Mô tả Fourier của đường cong là:

$$f_K = [|\mathbf{F}_1|, |\mathbf{F}_2|, \dots, |\mathbf{F}_{M/2}|] \quad (1.26)$$

Mô tả Fourier của khoảng cách trọng tâm là:

$$f_R = \left[ \frac{|\mathbf{F}_1|}{|\mathbf{F}_0|}, \frac{|\mathbf{F}_2|}{|\mathbf{F}_0|}, \dots, \frac{|\mathbf{F}_{M/2}|}{|\mathbf{F}_0|} \right] \quad (1.27)$$

Trong đó,  $F_i$  biểu thị thành phần thứ  $i$  của các hệ số biến đổi Fourier.

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

Mô tả Fourier của tọa độ phức hợp là:

$$f_Z = \left[ \frac{|\mathbf{F}_{-(\frac{M-1}{2})}|}{|\mathbf{F}_1|}, \dots, \frac{|\mathbf{F}_{-1}|}{|\mathbf{F}_1|}, \frac{|\mathbf{F}_2|}{|\mathbf{F}_1|}, \dots, \frac{|\mathbf{F}_{M/2}|}{|\mathbf{F}_1|} \right] \quad (1.28)$$

Trong đó,  $F_1$  là thành phần tần số khác 0 đầu tiên được sử dụng để chuẩn hóa các hệ số biến đổi.

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

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

### 1.3.3.4 Hình tròn, độ lệch tâm và hướng trục chính

Hình tròn được định nghĩa là:

$$\alpha = \frac{4\pi S}{P^2} \quad (1.29)$$

Trong đó:  $S$  là cỡ,  $P$  là chu vi của một đối tượng.

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

## 1.3.4 Độ đo khoảng cách và độ đo tương tự

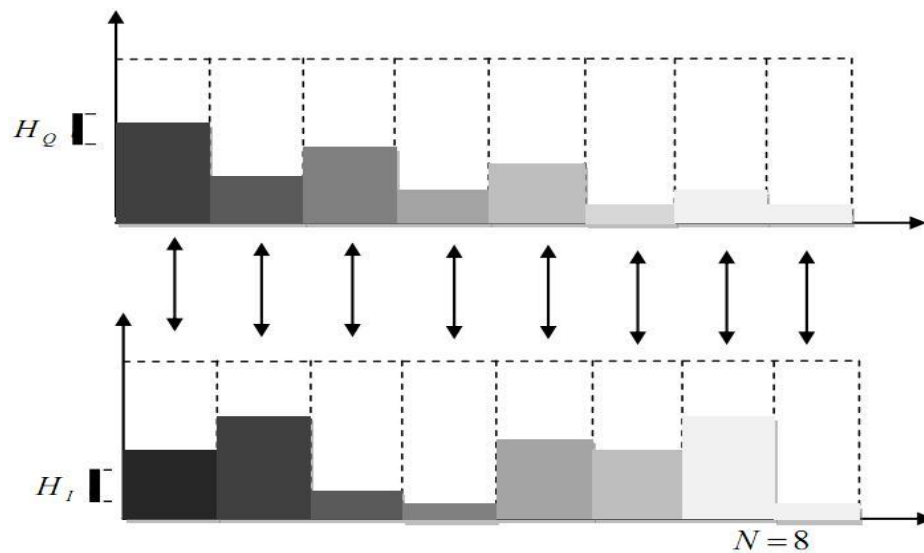
### 1.3.4.1 Độ đo khoảng cách Minkowski

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

$$d(Q,I) = \sum_{i=1}^N |H_Q[i] - H_I[i]|^r \quad (1.30)$$

Với Q và I là 2 ảnh, N là số mức trong biểu đồ màu (đối với mỗi ảnh, số lượng màu được giảm xuống N màu trong không gian màu RGB. Bởi vậy mỗi biểu đồ màu có N mức.).  $H_Q[i]$  là giá trị của mức I trong biểu đồ màu,  $H_Q$  biểu diễn cho ảnh Q và  $H_I[i]$  là giá trị mức i trong biểu đồ màu  $H_I$  biểu diễn cho ảnh I.

Khi  $r = 1$  thì khoảng cách Minkowski trở thành  $L_1$ . Khi  $r = 2$  thì khoảng cách đó trở thành khoảng cách Euclidean. Trong thực tế khoảng cách Euclidean có thể được xem như khoảng cách không gian trong không gian đa chiều.



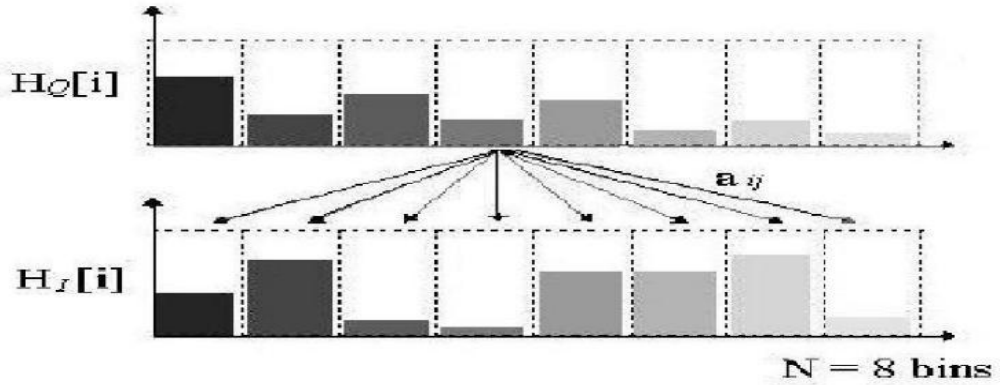
Hình 1.2 Độ đo khoảng cách Minkowski

### 1.3.4.2 Độ đo khoảng cách Quadratic

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

$$d(Q,I) = (H_Q - H_I)^t A (H_Q - H_I) \quad (1.31)$$

Với Q và I là hai ảnh,  $H_Q$  là biểu đồ màu của ảnh Q và  $H_I$  là biểu đồ màu của ảnh I,  $A = [a_{i,j}]$  là một ma trận cỡ  $N \times N$  với N là số mức trong các biểu đồ màu, và  $a_{i,j}$  biểu thị sự tương tự giữa màu i và màu j. Độ đo khoảng cách này khắc phục được sự thiếu sót của độ đo khoảng cách dạng Minkowski đó là các mức trong biểu đồ màu hoàn toàn không liên quan đến nhau.



Hình 1.3 Độ đo khoảng cách Quadratic

### 1.3.4.3 Độ đo khoảng cách Non-histogram

Stricker và Orengo đã đưa ra phương pháp Color Moments nhằm vượt qua hiệu quả của lượng tử hóa biểu đồ màu. Trong phương pháp này những đặc điểm phân bố màu của ảnh được biểu diễn bởi những đặc điểm chủ yếu của chúng gọi là các moment. Đó là độ trung bình, sự thay đổi và tính đối xứng. Moment đầu tiên là màu trung bình của ảnh, thứ hai là độ lệch chuẩn của mỗi kênh màu và thứ ba là mối liên hệ của mỗi kênh màu, và chúng được định nghĩa như sau:

$$\begin{aligned}
 E_i &= \left( \frac{1}{F} \sum_{j=1}^F (P_{ij} - E_i)^2 \right)^{\frac{1}{2}} \\
 \delta_i &= \left( \frac{1}{F} \sum_{j=1}^F (P_{ij} - E_i)^2 \right)^{\frac{1}{2}} \\
 S_i &= \left( \frac{1}{F} \sum_{j=1}^F (P_{ij} - E_i)^3 \right)^{\frac{1}{3}}
 \end{aligned} \tag{1.32}$$

Ở đây \$P\_{ij}\$ là giá trị của kênh màu thứ \$j\$ và điểm ảnh thứ \$i\$. \$E\_i\$ là màu trung bình của kênh màu thứ \$i\$. \$\delta\_i\$ là độ lệch tiêu chuẩn của kênh màu thứ \$i\$. \$S\_i\$ là giá trị của mối liên hệ thứ ba của ba kênh màu thứ \$i\$ và \$F\$ là tổng số điểm ảnh. Nếu \$Q\$ và \$I\$ là hai ảnh và đặc điểm màu của chúng được biểu diễn bởi \$r\$ kênh màu thì sự tương tự giữa hai ảnh này được định nghĩa như:

$$d(Q, I) = \sum_{i=1}^r (W_{i1} |E_i^Q - E_i^I| + W_{i2} |\delta_i^Q - \delta_i^I| + W_{i3} |S_i^Q - S_i^I|) \tag{1.33}$$

với \$W\_{i1}\$, \$W\_{i2}\$, \$W\_{i3}\$ là các trọng số được xác định bởi người dùng.

#### 1.3.4.4 Khoảng cách Mahalanobis

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

$$d(Q,I) = \sqrt{(H_Q - H_I)^T C^{-1} (H_Q - H_I)} \quad (1.34)$$

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

Trong trường hợp đó chỉ có sự tương quan của mỗi thành phần đặc trưng,  $C^t$  được cần đến:

$$d(Q,I) = \sum_{t=1}^n (H_Q - H_I)^2 / C_t \quad (1.35)$$

#### 1.3.4.5 Độ phân kỳ Kullback-Leibler

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

$$D(I,J) = \sum_i f_i(I) \log \frac{f_i(I)}{f_i(J)} \quad (1.36)$$

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

#### 1.3.4.6 Độ phân kỳ Jeffrey

Độ phân kỳ Jeffrey được định nghĩa là:

$$D(I,J) = \sum_i f_i(I) \log \frac{f_i(I)}{f_i^{\wedge}} + f_i(J) \log \frac{f_i(J)}{f_i^{\wedge}} \quad (1.37)$$

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

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

### 1.4 Đánh giá hiệu năng hệ thống tra cứu ảnh

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

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

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

$$\text{Precision} = \frac{|Q(q) \cap R(q)|}{|Q(q)|} \quad (1.38)$$

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

$$\text{Recall} = \frac{|Q(q) \cap R(q)|}{|R(q)|} \quad (1.39)$$

Trong các hệ thống tra cứu điển hình thì độ thu hồi có xu hướng tăng lên khi số lượng các kết quả thu được tăng lên trong khi đó thì độ chính xác dường như lại giảm đi. Ngoài ra việc lựa chọn  $R(q)$  lại rất không ổn định do sự đa dạng của cách hiểu về một bức ảnh. Hơn nữa, khi số lượng ảnh thích hợp lại lớn hơn số lượng ảnh hệ thống tìm được thì lúc đó khái niệm thu hồi trở thành vô nghĩa.

Do đó, độ chính xác và độ thu hồi chỉ là các mô tả ở dạng thô về hiệu năng của một hệ thống truy vấn mà thôi.

Gần đây MPEG7 có khuyến nghị một cách đánh giá mới về hiệu năng của các hệ thống tra cứu gọi là ANMRR (average normalized modified retrieval rank). Theo cách này độ chính xác và độ thu hồi được đối sánh thành một số đo duy nhất.

Ký hiệu số lượng ảnh hoàn toàn đúng với truy vấn  $q$  là  $N(q)$  và số lượng lớn nhất của các ảnh hoàn toàn đúng với tất cả  $Q$  truy vấn tức là  $\max\{N(q_1), N(q_2), \dots, N(q_Q)\}$  là  $M$ . Sau đó đối với mỗi truy vấn  $q$  thì mỗi bức ảnh hoàn toàn đúng  $k$  được gán một giá trị xếp hạng  $\text{rank}(k)$  giá trị này là thứ hạng của nó trong số những ảnh hoàn toàn đúng nếu ảnh đó nằm trong  $K$  kết quả truy vấn đầu tiên (ở đây  $K = \min\{4N(q), 2M\}$ ) hoặc có giá trị  $K+1$  nếu ảnh đó không nằm trong  $K$  kết quả truy vấn đầu tiên.

Thứ hạng trung bình  $\text{AVR}(q)$  đối với truy vấn  $q$  được tính như sau:

$$AVR(q) = \sum_{k=1}^{N(q)} \frac{rank(q)}{N(q)} \quad (1.40)$$

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

$$MRR(q) = AVR(q) - 0.5 - 0.5 * N(q) \quad (1.41)$$

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

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

$$NMRR(q) = \frac{MRR(q)}{K + 0.5 - 0.5 * N(q)} \quad (1.42)$$

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

$$ANMRR = \frac{1}{Q} \sum_{q=1}^Q NMRR(q) \quad (1.43)$$

## 1.5 Các hệ thống tra cứu ảnh dựa trên nội dung

### 1.5.1 Hệ thống QBIC (Query By Image Content)

Hệ thống QBIC của hãng IBM là một hệ thống tra cứu ảnh thương mại đầu tiên và nổi tiếng nhất trong số các hệ thống tra cứu ảnh dựa trên nội dung. Nó cho phép người sử dụng tra cứu ảnh dựa vào màu sắc, hình dạng và kết cấu. QBIC cung cấp một số phương pháp: Simple, Multi-feature và Multi-pass. Trong phương pháp truy vấn Simple chỉ sử dụng một đặc điểm. Truy vấn Multi-feature bao gồm nhiều hơn một đặc điểm và mọi đặc điểm đều có trọng số như nhau trong suốt quá trình tìm kiếm. Truy vấn Multi-pass sử dụng đầu ra của các truy vấn trước làm cơ sở cho bước tiếp theo. Người sử dụng có thể vẽ ra và chỉ định màu, kết cấu mẫu của hình ảnh yêu cầu. Trong hệ thống QBIC màu tương tự được tính toán bằng độ đo bình phương sử dụng biểu đồ màu k phần tử và màu trung bình được sử dụng như là bộ lọc để cải tiến hiệu quả của truy vấn.

### 1.5.2 Hệ thống Photobook

Hệ thống này được phát triển ở viện kỹ thuật Massachusetts. Nó cho phép người sử dụng tra cứu ảnh dựa trên màu sắc, hình dạng và kết cấu. hệ thống này

---

cung cấp một tập các thuật toán đối sánh gồm: Euclidean, Mahalanobis, Vector Space Angle, Histogram, Fourier Peak, và Wavelet Tree Distance như là những độ đo khoảng cách. Hệ thống như là một công cụ bán tự động và có thể sinh ra một mẫu truy vấn dựa vào những ảnh mẫu được cung cấp bởi người sử dụng. Điều này cho phép người sử dụng trực tiếp đưa những yêu cầu truy vấn của họ với những lĩnh vực khác nhau, và mỗi lĩnh vực họ có thể thu được những mẫu truy vấn tối ưu.

### **1.5.3 Hệ thống VisualSEEK và WebSEEK**

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

### **1.5.4 Hệ thống RetrievalWare**

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

### **1.5.5 Hệ thống Imatch**

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



---

## CHƯƠNG 2: BIỂU DIỄN HÌNH DẠNG DỰA TRÊN VÙNG

### 2.1 Giới thiệu

Ngày càng có nhiều hình ảnh được số hóa trên toàn thế giới, nên nhu cầu tìm kiếm ảnh trong bộ sưu tập lớn hoặc từ cơ sở dữ liệu từ xa. Để tìm ra một hình ảnh, hình ảnh đó phải được mô tả hoặc biểu diễn bởi một số tính năng. Hình dạng (Shape) là một tính năng quan trọng của một hình ảnh trực quan. Tìm kiếm hình ảnh bằng cách sử dụng tính năng hình dạng đang nhận được sự quan tâm của rất nhiều nhà phát triển.

Tuy nhiên biểu diễn và mô tả hình dạng là một nhiệm vụ khó khăn. Điều này bởi vì khi đối tượng ở thế giới thực được phản ánh theo không gian hình ảnh 2-D, một lượng thông tin kích thước của đối tượng bị mất. Kết quả là hình dạng từ hình ảnh chỉ biểu diễn một phần của đối tượng. Hình dạng thường bị hỏng vì nhiễu, khuyết, biến dạng...

Biểu diễn hình dạng cho biết đặc điểm đối tượng và hiệu quả về mặt nhận thức quan trọng dựa trên thông tin hình dạng ranh giới hoặc ranh giới cộng với nội dung cấu tạo bên trong. Các tính năng khác nhau đã được thiết kế, bao gồm: chữ ký hình dạng (shape signature), biểu đồ chữ ký (signature histogram), shape invariants (hình dạng bất biến), moments, độ cong (curvature), ngữ cảnh hình dạng (shape context), ma trận hình dạng (shape matrix)...

### 2.2 Biểu diễn hình dạng dựa trên vùng

Trong kỹ thuật biểu diễn dựa trên vùng, tất cả các điểm ảnh trong một khu vực hình dạng được xét đến để có được các biểu diễn hình. Các phương pháp dựa trên vùng phổ biến sử dụng mô tả moments để mô tả hình dạng. Phương pháp dựa trên các vùng khác bao gồm phương pháp lưới, ma trận hình dạng, trục trung vị, bề mặt lồi. Phương pháp dựa trên vùng cũng có thể được chia thành các phương pháp toàn cục và cấu trúc tùy thuộc vào việc hình dạng của chúng có tách biệt với các thành phần phụ hay không.

---

### 2.2.1 Phương pháp toàn cục (Global Method)

Phương pháp toàn cục xem xét đến toàn bộ hình dạng, kết quả của mô tả là vector số đặc trưng (numeric feature vector), nó được sử dụng để mô tả hình dạng.

#### 2.2.1.1 Moment bất biến hình học (Geometric Moment Invariant)

M.K. Hu đã công bố bài nghiên cứu đầu tiên về việc sử dụng các moment bất biến cho ứng dụng nhận dạng mẫu hai chiều. Phương pháp tiếp cận của ông dựa trên các nghiên cứu của các nhà toán học thế kỷ 19 và lý thuyết đại số:

$$m_{pq} = \sum_x \sum_y x^p y^q f(x, y) \quad (2.1)$$

$p, q = 0, 1, 2, \dots$

Sử dụng đối sánh phi tuyến các moment có thứ tự thấp, một tổ hợp các moment bất biến (thường được gọi là moment hình học), trong đó các thuộc tính bất biến với co giãn và phép quay được rút ra. Việc sử dụng các moment có thứ tự cao cho phân tích mẫu không được áp dụng. Vấn đề chính với moment hình học là chỉ có một số bất biến được rút ra từ các thứ tự thấp của moment, như vậy không đủ để mô tả chính xác hình dạng, nhưng cũng rất khó để lấy được những bất biến thứ tự cao hơn.

#### 2.2.1.2 Moment bất biến đại số (Algebraic Moment Invariant)

Moment bất biến đại số (AMI) được Taubin và Cooper giới thiệu và sử dụng trong QBIC. Các AMI được tính toán cho từ moment trung tâm đầu tiên và được đặt ra như là giá trị riêng của ma trận định trước  $M[j, k]$ , trong đó các phần tử tỉ lệ với các yếu tố của các moment trung tâm. Khác với phương pháp moment bất biến hình học của Hu, các moment bất biến đại số có thể được xây dựng từ các thứ tự bất kỳ. AMI có xu hướng làm việc tốt trên các đối tượng có điểm ảnh được phân bố và không phải là hình dạng phức tạp.

#### 2.2.1.3 Moment trực giao (Orthogonal Moments)

Moment đại số có thể biến đổi để trở nên tổng quát hơn bằng cách thay thế hạt nhân biến đổi thông thường  $x^p y^q$  bằng một hạt nhân tổng quát hơn của  $P_p(x)P_q(y)$ . Teague sử dụng ý tưởng này để mô tả moment trực giao-Moment

Legendre và moment Zernike—bằng cách thay thế  $x^p y^q$  trong phương pháp toàn cầu với đa thức Legendre và đa thức Zernike tương ứng.

Moment Legendre được mô tả:

$$\lambda_{mn} = \frac{(2m+1)(2n+1)}{4} \sum_x \sum_y P_m(x) P_n(y) f(x, y) \quad (2.2)$$

$$\text{Với } P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2 - 1)^n$$

Các moments Zernike được mô tả:

$$A_{nm} = \frac{n+1}{\pi} \sum_x \sum_y V_{nm}^*(x, y) f(x, y), \quad x^2 + y^2 \leq 1 \quad (2.3)$$

$$\text{Trong đó: } V_{nm}(x, y) = V_{nm}(\rho \cos \theta, \rho \sin \theta) = R_{nm}(\rho) \exp(jm\theta)$$

$$R_m(\rho) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \times \frac{(n-s)!}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)!} \rho^{n-2s}$$

$\rho$  và  $\theta$  tương ứng là bán kính và góc của điểm ảnh  $(x, y)$  liên quan đến trọng tâm của hình dạng.

Vì cả hai đa thức Legendre và Zernike đều có đủ cả hai yếu tố của một trực giao cơ sở nên moment Legendre và moment Zernike đều được gọi là moment trực giao. Moment trực giao cho phép tái tạo chính xác hình dạng mô tả và sử dụng một cách tối ưu các thông tin hình dạng.

#### 2.2.1.4 Mô tả Fourier tổng quát

Mô tả Fourier tổng quát được đề xuất bởi Zhang và Lu được tạo ra bằng cách áp dụng biến đổi Fourier 2-D lấy mẫu trên một hình dạng vùng raster.

*D. Zhang, G. Lu / Pattern Recognition 37 (2004) 1–19*



Hình 2.1 Mô tả Fourier

$$PF_2(\rho, \phi) = \sum_r \sum_t f(r, \theta_t) \exp \left[ j2\pi \left( \frac{r}{R} \rho + \frac{2\pi i}{T} \phi \right) \right] \quad (2.4)$$

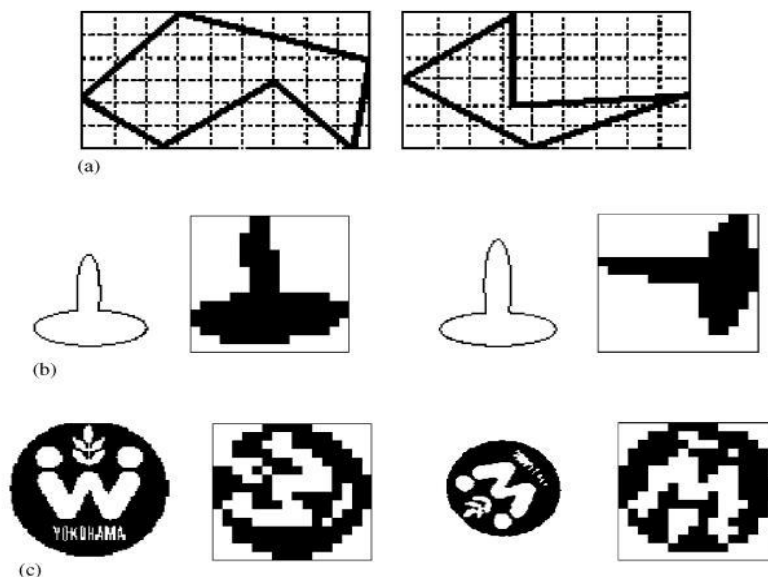
Trong đó:  $0 \leq r < R$  và  $\theta_t = i(2\pi/T)$  ( $0 \leq i < T$ );  $0 \leq \rho < R$ ,  $0 \leq \phi < T$ .

R và T là độ phân giải tần số xuyên tâm và độ phân giải tần số góc tương ứng. Các hệ số chuẩn hóa là các GFD. Sự tương tự giữa hai hình được đo bằng khoảng cách các khối giữa các GFD của hai hình.

So với những moment Zernike, GFD đơn giản hơn trong tính toán. Các tính năng là tính năng thuần túy quang phổ và có hiệu suất thu hồi tốt hơn nhờ phân tích đa độ phân giải theo hướng xuyên tâm và hình tròn của hình dạng.

### 2.2.1.5 Phương pháp lưới (Grid based method)

Phương pháp này được đề xuất bởi Lu và Sajjanhar. Về cơ bản một mạng lưới các ô được phủ lên trên hình dạng sau đó được quét từ trái sang phải và từ trên xuống dưới. Kết quả thu được là một ảnh bitmap. Các ô bao phủ hình dạng được gán giá trị 1, các ô còn lại được gán giá trị 0. Hình dạng sau đó có thể được biểu diễn bởi các vector đặc trưng nhị phân. Khoảng cách Hamming nhị phân được sử dụng để đo lường sự giống nhau giữa hai hình dạng.



Hình 2.2 Phương pháp lưới

Những lợi thế của phương pháp lưới là đơn giản hóa biểu diễn, sự phù hợp với trực giác và thống nhất với phương pháp mã hóa hình dạng MPEG-4. Vấn đề chính của phương pháp này là trực chỉnh dựa trên góc xoay. Trực chỉnh nhạy cảm với điểm nhiễu và không đáng tin cậy. Ví dụ hai hình dạng tương tự có hình dạng lưới rất khác nhau. Hai hình dạng tương tự nhau nhưng góc quay khác nhau có hình dạng lưới rất khác nhau. Phương pháp lưới không bất biến với quay.

### 2.2.1.6 Ma trận hình dạng (Shape Matrix)

Các phương pháp mô tả hình dạng thông thường sử dụng khung lưới hình chữ nhật để lấy thông tin hình. Mô tả hình dạng có được theo cách này thường là không bất biến dịch chuyển, quay và mở rộng. Goshtasby đề nghị sử dụng một ma trận hình dạng có nguồn gốc từ một kỹ thuật lấy mẫu raster hình tròn. Thay vì che phủ lưới vuông trên một hình ảnh, một raster các vòng tròn đồng tâm và đường xuyên tâm được đặt ở trung tâm của hình. Giá trị nhị phân được lấy tại các giao điểm của vòng tròn và đường xuyên tâm. Ma trận hình dạng được hình thành để các vòng tròn tương ứng với các cột của ma trận và các đường xuyên tâm tương ứng với các hàng của ma trận. Trước khi lấy mẫu, hình dạng được mô tả bằng cách sử dụng bán kính tối đa của hình dạng. Các kết quả biểu diễn ma trận là bất biến với dịch chuyển, xoay và biến dạng.

Khi một ma trận hình dạng là một hình dạng rải rác nó có thể dễ dàng bị ảnh hưởng bởi nhiễu. Bên cạnh đó đối sánh hình bằng cách sử dụng ma trận hình dạng là quá tốn kém. Perui et al. đề xuất một mô tả hình dạng dựa trên các khu vực tương đối của hình nằm trong các vòng tròn đồng tâm nằm ở trung tâm của hình dạng.  $L$  là bán kính tối đa của hình chữ  $S$  được mô tả,  $C_k$ , vòng thứ  $k$  của  $n$  vòng tròn thu được bằng cách phân đoạn bán kính tối đa  $L$  vào  $n$  phân đoạn tương đương.

Một vùng bất biến tỷ lệ được định nghĩa là:

$$x_i = \frac{A(S \cap C_i)}{A(C_i)} \quad (2.5)$$

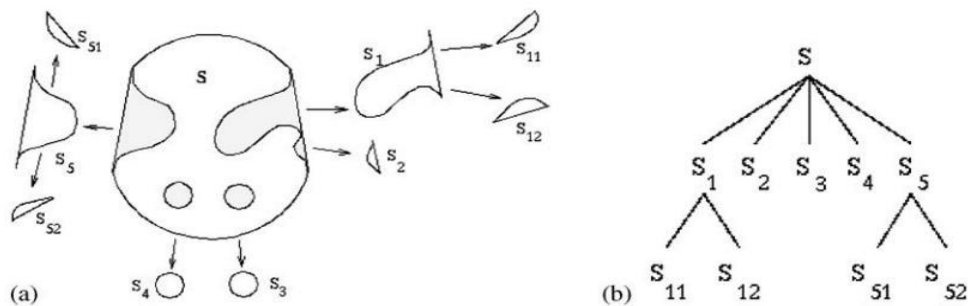
$A(\cdot)$  là hàm diện tích. Mô tả hình dạng là vector đặc trưng của  $x = [x_1, \dots, x_n]^T$ .

## 2.2.2 Phương pháp cấu trúc (Structural methods)

Phương pháp dựa trên cấu trúc vùng phân chia các khu vực của hình dạng thành các bộ phận mà sau đó được sử dụng để mô tả và biểu diễn hình dạng.

### 2.2.2.1 Bề mặt lồi (Convex hull)

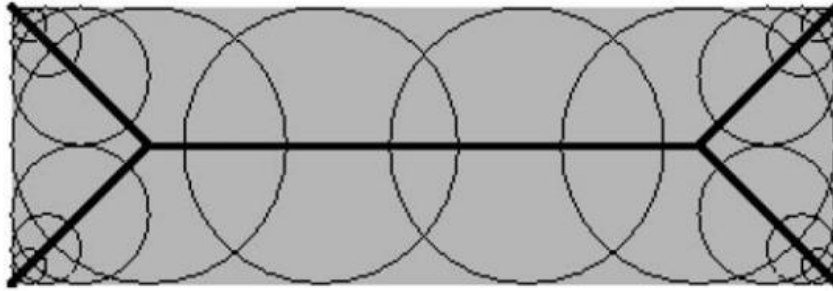
Một vùng  $R$  là lồi khi và chỉ khi với hai điểm bất kỳ  $x_1, x_2 \in R$  thì toàn bộ đoạn  $x_1x_2$  nằm bên trong vùng. Bề mặt lồi của một vùng lồi là vùng lồi  $H$  nhỏ nhất đáp ứng điều kiện  $R \subset H$ . Sự khác biệt của  $R-H$  được gọi là thiếu hụt lồi  $D$  của vùng  $R$  (convex deficiency). Khi đường bao hình dạng có xu hướng bất thường vì nhiễu, việc số hóa và các biến thể trong phân chia nhỏ thường dẫn đến sự thiếu hụt lồi nhỏ. Đầu tiên, bề mặt lồi của một đối tượng thu được với các thiếu hụt lồi của nó, sau đó lại tìm bề mặt lồi và thiếu hụt lồi của các thiếu hụt lồi đã tìm thấy ở bước trước, quá trình tiếp tục cho đến khi các thiếu hụt lồi đều là các vùng lồi. Hình 2.3 minh họa quá trình này. Hình dạng sau đó được mô tả như là một cây lổm. Việc đối sánh giữa các hình dạng trở thành đối sánh đồ thị hoặc chuỗi.



Hình 2.3 Minh họa phương pháp bề mặt lồi, (a): Bề mặt lồi và các thiếu hụt lồi của nó, (b): Cây lổm biểu diễn bề mặt lồi.

### 2.2.2.2 Trục trung vị (Medial Axis)

Cũng giống như bề mặt lồi, xương cũng có thể được sử dụng để mô tả và biểu diễn hình dạng. Xương (trục trung vị) là quỹ tích tâm của các đĩa cực đại của hình dạng như trong hình..., đường in đậm là xương của hình chữ nhật.



Hình 2.4 Minh họa phương pháp trục trung vị

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 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.

## 2.3 Phương pháp tìm xương theo DCE

### 2.3.1 Ý tưởng chính của phương pháp

Nhóm tác giả Xiang Bai, Login Jan Latecki, Wen-Yu Liu đã đề xuất một phương pháp loại bỏ hoàn toàn những điểm lỗi ra mà không loại bỏ những điểm biên, vì vậy không loại bỏ những điểm xương chính. Những điểm sai hoặc thừa ra hoàn toàn bị loại bỏ trong khi những nhánh xương chính không bị ngắt đi. Phương pháp này có thể cắt tia xương dựa trên việc phân chia đường biên thành những đoạn cong. Ý tưởng chính của phương pháp là di chuyển tất cả các điểm xương của điểm tăng trưởng nằm trên cùng đoạn đường biên. Từ đó đặt ra câu hỏi làm thế nào để tìm ra các đoạn phân chia đường biên tốt nhất. Nhóm tác giả đã tìm ra được sự phân chia như vậy nhờ quá trình DCE.

### 2.3.2 Rời rạc hóa đường cong

DCE được giới thiệu bởi nhóm các tác giả Xiang Bai và các cộng sự. Đường biên của đối tượng trong ảnh số bị thay đổi bởi nhiễu và các lỗi phân đoạn. DCE loại bỏ những thay đổi đó trong khi vẫn đảm bảo được hình dạng ban



đầu của đối tượng bởi đơn giản hóa hình dạng. Bất kỳ đường cong của ảnh số có thể được coi là một đa giác mà không bị mất thông tin, nhưng phải có số đỉnh lớn để nghiên cứu sự phát triển của hình dạng. Ý tưởng cơ bản của sự phát triển đề xuất là các đa giác đều đơn giản hơn.

Trước tiên tác giả Xiang Bai đưa ra phép đo liên quan K:

$$K(\mathbf{S}_1, \mathbf{S}_2) = \frac{B(s_1 s_2)l(s_1)l(s_2)}{l(s_1)+l(s_2)} \quad (2.6)$$

Trong đó  $s_1, s_2$  là những cạnh của đa giác liên quan tới đỉnh  $v$ ;  $\beta(s_1, s_2)$  là góc quay tại đỉnh chung của đoạn  $s_1, s_2$ ;  $l$  là tổng độ dài của đường cong đa giác  $C$ .

Đầu vào là đường biên đa giác  $P$  với  $n$  đỉnh, DCE tạo ra một chuỗi các đa giác đơn giản với  $K$  là nhỏ nhất.

Tác giả còn chỉ ra rằng một tính chất quan trọng của DCE là phân chia trình tự với đa giác đầu vào  $P$ .  $\{v_1, \dots, v_n\}$  là đỉnh của  $P$ ,  $\{u_1, \dots, u_n\} \subset \{v_1, \dots, v_n\}$  là đỉnh lồi của  $P^{n-k}$  sao cho  $m \leq n-k$ .

Nhận thấy DCE và phân chia trình tự có thể được định nghĩa cho một tập hữu hạn của đường cong đa giác. Trong mỗi bước DCE một vector đơn được loại bỏ từ một đa giác mà phép đo liên quan là nhỏ nhất. Phương pháp cắt tĩa đề xuất có thể được áp dụng cho mặt phẳng  $D$ , với đường biên bao gồm số các đa giác đóng đơn giản.

DCE có thể loại bỏ hiệu quả nhiều và từng phần không quan trọng của ảnh, nhưng vẫn cần phải có một giới hạn  $T$  dừng đúng cách để phù hợp với những ứng dụng cụ thể. Nói cách khác tìm kiếm  $k$  để đa giác đơn giản  $P^{n-k}$  miêu tả chi tiết những đường biên đầu vào.

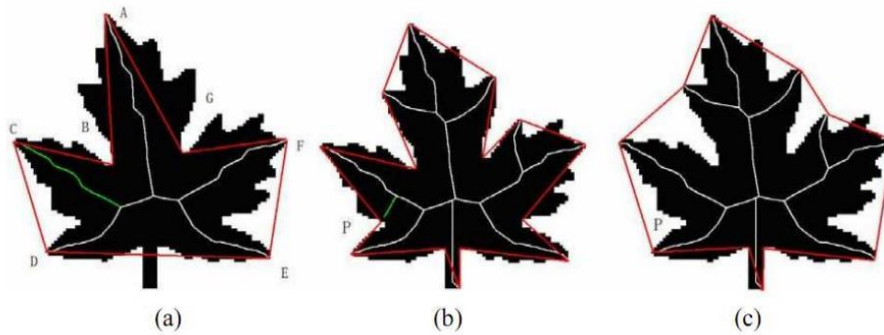
### 2.3.3 Phương pháp cắt tĩa xương với DCE

Cho một bộ xương  $S(D)$  của một mặt phẳng  $D$  và đưa ra một DCE đa giác đơn giản  $P^k$ , thể hiện cắt tĩa xương bằng cách di chuyển tất cả những điểm  $s \in S(D)$ , như vậy tạo ra những điểm tăng trưởng  $\tan(s)$  của  $s$  chứa trong cùng đoạn DCE mở. Mỗi điểm cắt tĩa  $s$  là kết quả từ phân đường biên cục bộ với sự



phân chia DCE, và do đó,  $s$  có thể coi như là điểm xương không quan trọng và có thể loại bỏ. Quá trình làm đơn giản đường biên với DCE đã hoàn thành cắt tỉa nhánh của xương. Đặc biệt, loại bỏ một đỉnh lồi  $v$  từ  $P^{n-k}$  tới  $P^{n-(k+1)}$  bởi DCE, tức là hoàn thành loại bỏ những nhánh xương mà kết thúc tại  $v$ .

Tác giả thực hiện phân tích đoạn DCE dựa trên các đỉnh lồi bởi đơn giản hóa DCE. Khi một đỉnh lồi trở thành một đỉnh lõm trong quá trình phát triển của DCE thì những nhánh xương kết thúc tại đỉnh đó bị loại bỏ. Cách tiếp cận này cho phép loại bỏ những nhánh xương nhỏ trong quá trình phát triển DCE.



Hình 2.5 Minh họa tìm xương theo DCE

Một thuộc tính quan trọng của DCE là gây ra phân chia đường biên và mỗi phân chia làm giảm các đỉnh của đường biên đa giác, kết quả có một nhánh xương kết thúc tại mỗi điểm phân chia. Theo các kết quả thực nghiệm mà tác giả đã nghiên cứu, trong một bước tiến hóa DCE nếu đỉnh  $u_i$  của đa giác bị xóa ( tức là  $u_i \in (P^{n-k} - P^{n-(k+1)})$ ) hoặc trở thành lõm ( do việc xóa đi một trong những đỉnh bên cạnh của nó) thì cung  $[u_{i-1}, u_{i+1}]$  thay thế cung  $[u_{i-1}, u_i]$ ,  $[u_i, u_{i+1}]$ . Khi đó cắt tỉa xương sẽ loại bỏ toàn bộ nhánh xương kết thúc tại  $u_i$ .

---

## CHƯƠNG 3: ĐỐI SÁNH SHAPE DỰA TRÊN ĐẶC TRƯNG VÙNG

### 3.1 Đối sánh dựa trên sự tương tự của đồ thị xương

Xương (hay còn gọi là trục trung vị) tích hợp các tính năng hình học và topo của đối tượng, là một mô tả hình dạng quan trọng đối với nhận dạng đối tượng. Sự tương đồng về hình dạng dựa trên đồ thị xương thường được thực hiện tốt hơn so với dựa trên đường biên hoặc các mô tả hình dạng khác khi có mặt của chồng lấp từng phần và khớp nối nhiều phần.

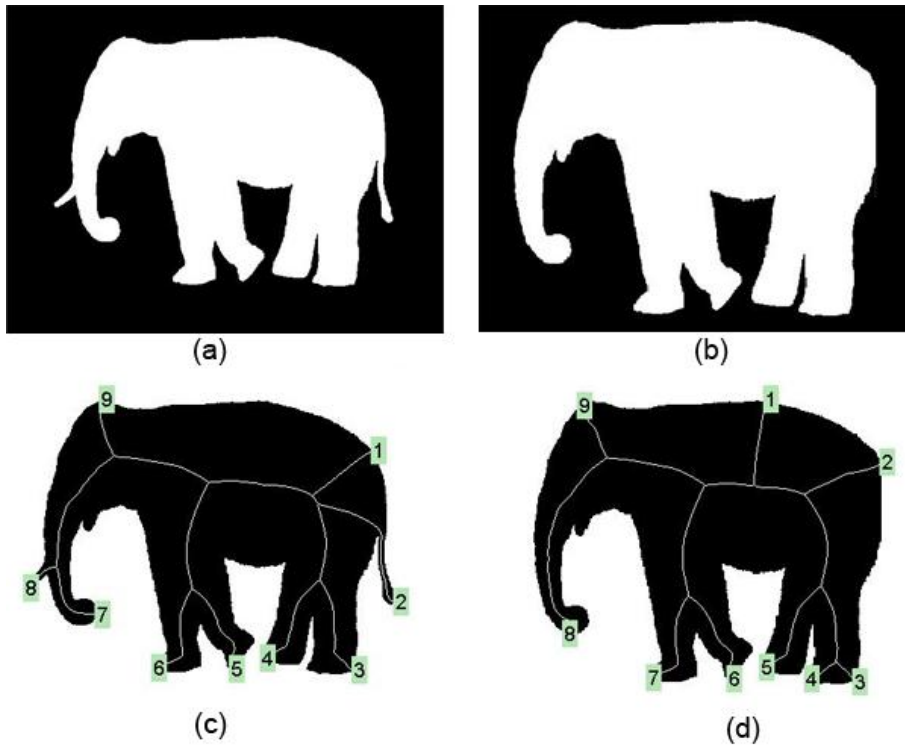
Tuy nhiên, để tự động nhận dạng các đối tượng bằng cách sử dụng xương là một nhiệm vụ đầy thách thức vì sự nhạy cảm của xương đối với biến dạng biên. Một hạn chế chính của phương pháp nhận dạng dựa trên xương là cấu trúc phức tạp của cây hoặc đồ thị biểu diễn của xương.

X.Bai và L.Jan Latecki [6] đã đưa ra một phương pháp thực hiện khá hiệu quả trong việc nhận dạng đối tượng dựa trên xương. Ý tưởng chính của phương pháp 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 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 đa số các 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 với DCE (Discrete Curve Evolution) có điểm cuối nhánh xương tương ứng với phần trực quan của đối tượng.

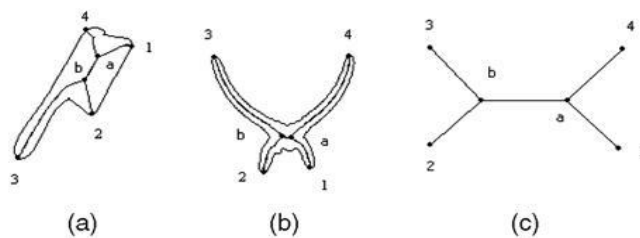
X.Bai và các cộng sự sử dụng sự tương tự của các đường đi ngắn nhất giữa mỗi cặp điểm cuối của xương để thiết lập mối quan hệ tương ứng với điểm cuối trong đồ thị. Vì vậy ý tưởng cơ bản của phương pháp này là xác định sự giống nhau của các cấu trúc phức tạp (complex structures) của đồ thị hoặc cây bằng cách kiểm tra đường đi ngắn nhất giữa các điểm cuối của chúng.

Thách thức khó khăn nhất cho đo độ tương tự xương là cấu trúc topo của cây xương hoặc đồ thị của các đối tượng tương tự nhau lại có thể hoàn toàn khác nhau. Thực tế này được minh họa trong hình 3.1, mặc dù bộ xương của hai con

voi (hình 3.1a và hình 3.1b) là tương tự, nhưng đồ thị xương (hình 3.1c và hình 3.1d) lại rất khác nhau. Ví dụ này minh họa những khó khăn phải đối mặt bởi cách tiếp cận dựa trên hoạt động chỉnh sửa đồ thị trong đối sánh xương.



Hình 3.1 Hình dạng (a) và (b) là tương tự nhưng khác đồ thị xương. Mặt khác, đồ thị xương của các đối tượng khác nhau có thể có cấu trúc topo giống nhau, như trong hình 3.2. Các xương của bàn chải trong hình 3.2(a) và kìm trong hình 3.2(b) có cùng topo như thể hiện trong hình 3.2(c).



Hình 3.2 Hình dạng (a) và (b) khác nhau nhưng có đồ thị xương (c) giống nhau.

Đề xuất đối sánh đồ thị xương của ảnh dựa trên giả định rằng sự tương tự xương có sự tương tự cấu trúc của các nút cuối (sự giống nhau của đường đi ngắn nhất đến các nút cuối khác).

### 3.1.1 Đồ thị xương (Skeleton Graphs)

Phần này mô tả những bước đầu tiên để xây dựng đồ thị xương. Các định nghĩa sau đây áp dụng cho xương liên tục, cũng như xương trong ảnh kỹ thuật số (bao gồm các điểm ảnh).

- Định nghĩa 1: Một điểm xương chỉ có một điểm liền kề là một điểm cuối (điểm cuối xương), một điểm xương có ba hoặc nhiều hơn các điểm lân cận là điểm giao. Nếu một điểm xương không phải là điểm cuối hoặc điểm giao nó được gọi là điểm kết nối (giả định rằng các đường cong của xương có chiều rộng là 1 pixel).

- Định nghĩa 2: Chuỗi các điểm kết nối trực tiếp giữa hai điểm xương được gọi là một nhánh xương. Một cách tiêu chuẩn để xây dựng một đồ thị xương như sau: các điểm cuối và các điểm giao nhau được lựa chọn là các nút cho các đồ thị, và tất cả các nhánh xương giữa các nút là các cạnh giữa các nút.

- Định nghĩa 3: Các điểm cuối trong đồ thị xương được gọi là một nút cuối và điểm giao nhau trong đồ thị xương được gọi là một nút giao nhau.

### 3.1.2 Đối sánh các đồ thị xương (Matching the Skeleton Graphs)

#### 3.1.2.1 Biểu diễn Shape-path

- Định nghĩa 4: Đường đi ngắn nhất giữa một cặp của các nút cuối cùng trên một đồ thị xương được gọi là một đường dẫn xương.

Giả sử có  $N$  nút kết thúc trong đồ thị xương  $G$ , và  $v_i$  chỉ nút thứ  $i$  dọc theo đường viền theo chiều kim đồng hồ. Cho  $p(v_m, v_n)$  biểu thị đường xương từ  $v_m$  tới  $v_n$ . Lấy mẫu  $p(v_m, v_n)$  với  $M$  điểm cách đều, đó là tất cả các điểm xương.  $R_{m,n}(t)$  là bán kính tối đa của đĩa tại điểm xương với chỉ số  $t$  trong  $p(v_m, v_n)$ . Một vector bán kính của đĩa tối đa có tâm tại điểm  $M$  trên  $p(v_m, v_n)$  được biểu thị như:

$$R_{m,n} = (R_{m,n}(t))_{t=1,2,\dots,M} = (r_1, r_2, \dots, r_M) \quad (3.1)$$

Bán kính  $R_{m,n}(t)$  là xấp xỉ với các giá trị của hàm biến đổi khoảng cách  $DT(t)$  (Distance Transform) tại mỗi điểm xương với chỉ số  $t$ . Giả sử có  $N_0$  điểm ảnh trong hình dạng ban đầu  $S$ . Để làm cho phương pháp đề nghị bất biến với co giãn, các tác giả đã chuẩn hóa  $R_{m,n}(t)$  theo cách sau:

$$R_{m,n} = \frac{DT(t)}{\frac{1}{N_0} \sum_{i=1}^{N_0} DT(s_i)} \quad (3.2)$$

Với  $S_i$  ( $i=1,2,\dots,N_0$ ) biến đổi trên tất cả các  $N_0$  điểm ảnh trong hình.

- Định nghĩa 5: Hình dạng khác nhau giữa hai đường xương được gọi là một khoảng cách đường dẫn (path distance). Nếu  $R$  và  $R'$  biểu thị các vector bán kính của hai đường dẫn hình dạng  $p(u,v)$  và  $p(u',v')$  tương ứng, khoảng cách đường dẫn được định nghĩa là:

$$pd(p(u,v), p(u',v')) = \sum_{i=1}^M \frac{(r_i - r'_i)^2}{r_i + r'_i} + \alpha \frac{(l - l')^2}{l + l'} \quad (3.3)$$

$l$  và  $l'$  là chiều dài của  $p(u,v)$  và  $p(u',v')$  và  $\alpha$  là yếu tố khối lượng. Để kích thước là bất biến chiều dài đường đi phải được chuẩn hóa. Bằng cách này biểu diễn đường đi và khoảng cách các đường dẫn có kích thước là bất biến.

Để giải quyết vấn đề giống nhau của hình dạng khớp nối, khoảng cách đường dẫn trong (3.3) không xử phạt biến dạng đường đi (vd, biến dạng từ thẳng đến đường cong) không thay đổi các vector bán kính và độ dài đường đi. Điều này cho phép chúng ta nhận ra sự tương tự giữa các đối tượng biến dạng như rắn.

### 3.1.2.2 Đối sánh các nút cuối sử dụng đường dẫn (Matching End Nodes

#### Using Skeleton Paths)

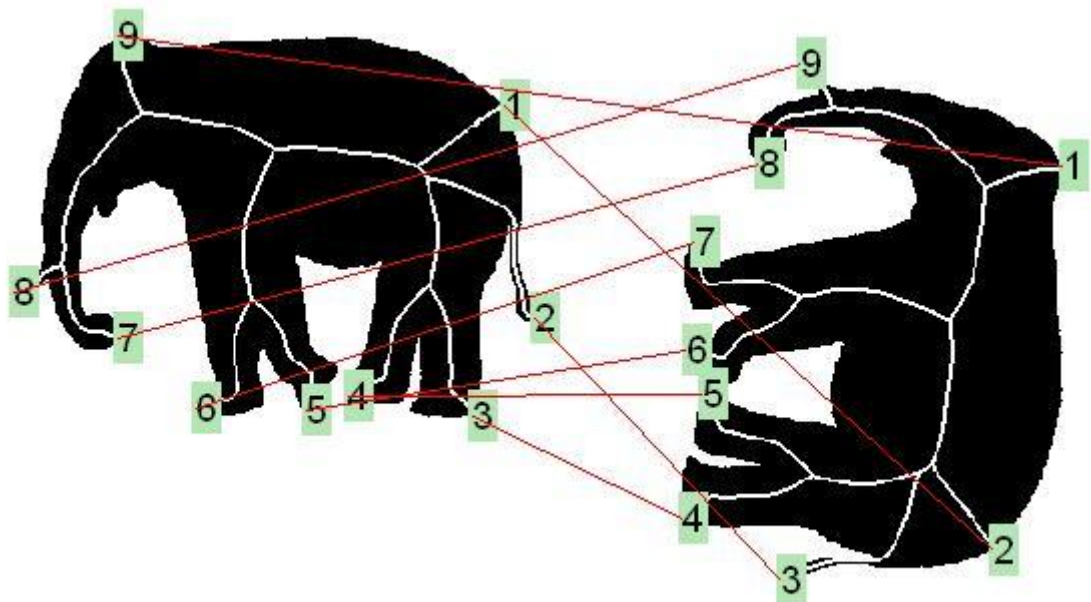
Trong đồ thị xương, mỗi nút cuối có những đường dẫn xương tới tất cả các nút khác trong đồ thị. Như chúng ta sẽ thấy, những đường xương là một mô tả hình dạng hữu ích.

Để  $G$  và  $G'$  biểu thị hai đồ thị được đối sánh,  $v_i$  và  $v'_j$  là một số nút kết thúc trong  $G$  và  $G'$ . Số lượng các nút kết thúc trong  $G$  và  $G'$  là  $K+1$  và  $N+1$ ,  $K \leq N$ . Chi phí đối sánh  $c(v_i, v'_j)$  giữa  $v_i$  và  $v'_j$  được ước tính dựa trên các đường dẫn đến tất cả các đỉnh khác phát tán trong  $G$  và  $G'$  từ  $v_i$  và  $v'_j$  tương ứng. Trước tiên, đặt tất cả các nút cuối cùng trong  $G$  theo đường viền theo chiều kim đồng hồ với các nút bắt đầu là  $v_i$  mà chúng biểu thị như  $v_{i0}$ . (Ở đây, có một thuận lợi là tất cả các điểm cuối của xương đều nằm trên đường biên.). Ta có được một chuỗi các nút kết thúc  $v_{i0}, v_{i1}, \dots, v_{ik}$  trong  $G$  và tương tự  $v_{j0}, v_{j1}, \dots, v_{jn}$  trong  $G'$ .

Sau đó sẽ tính toán khoảng cách đường đi giữa hai chuỗi (Biểu diễn cho những đường xuất phát từ  $v_i=v_{i0}$  ở  $G$  và  $v_j=v_{j0}$  trong  $G'$ ). Ta có một ma trận khoảng cách đường tính theo:

$$pd(v_i, v_j) = \begin{pmatrix} pd(p(v_{i0}, v_{i1}), p(v'_{j0}, v'_{j1})) & pd(p(v_{i0}, v_{i1}), p(v'_{j0}, v'_{j2})) & pd(p(v_{i0}, v_{iK}), p(v'_{j0}, v'_{jN})) \\ pd(p(v_{i0}, v_{i2}), p(v'_{j0}, v'_{j1})) & pd(p(v_{i0}, v_{i2}), p(v'_{j0}, v'_{j2})) & pd(p(v_{i0}, v_{i2}), p(v'_{j0}, v'_{jN})) \\ \vdots & \vdots & \vdots \\ pd(p(v_{i0}, v_{iK}), p(v'_{j0}, v'_{j1})) & pd(p(v_{i0}, v_{iK}), p(v'_{j0}, v'_{j2})) & pd(p(v_{i0}, v_{iK}), p(v'_{j0}, v'_{jN})) \end{pmatrix} \quad (3.4)$$

Để tính toán giá trị không giống nhau giữa hai nút cuối  $v_i$  và  $v_j$  các tác giả đã sử dụng phương pháp đối sánh hai chuỗi có độ dài khác nhau (OSB). Các thuộc tính của OSB có thể bỏ qua các yếu tố ngoại lai, tức là có thể bỏ qua một số điểm cuối của xương.



Hình 3.3 Kết quả đối sánh hai con voi

Bằng cách áp dụng OSB vào ma trận (3.4) ta có sự khác nhau của hai đầu nút  $v_i$  và  $v_j$ :

$$c(v_i, v'_j) = OSB(pd(v_i, v'_j)) \quad (3.5)$$

Với hai đồ thị  $G$  và  $G'$ , với các nút cuối cùng  $v_i$  ( $i=0,1,\dots,K$ ) và  $v'_j$  ( $j=0,1,\dots,N$ ), sau khi tính toán các giá trị không giống nhau giữa các nút cuối của hai đồ thị ta có ma trận mới như sau:

$$C(G, G') = \begin{pmatrix} c(v_0, v'_0) & c(v_0, v'_1) & \dots & c(v_0, v'_N) \\ c(v_1, v'_0) & c(v_1, v'_1) & \dots & c(v_1, v'_N) \\ \vdots & \vdots & \ddots & \vdots \\ c(v_K, v'_0) & c(v_K, v'_1) & \dots & c(v_K, v'_N) \end{pmatrix} \quad (3.6)$$

Cuối cùng, tính toán tổng số khác biệt  $c(G, G')$  giữa  $G$  và  $G'$  với thuật toán Hungary trên  $C(G, G')$ . Đối với mỗi nút cuối  $v_i$  trong  $G$  thuật toán Hungary có thể tìm thấy nút cuối tương ứng  $v'_j$  trong  $G'$ . Khi  $G$  và  $G'$  có thể có số lượng các nút cuối khác nhau, tổng giá trị khác biệt có thể bao gồm hình phạt cho các nút cuối mà không tìm được nút cuối tương ứng. Để đạt được điều này phải bổ sung thêm các hàng với giá trị hằng số tới [3.6] để  $C(G, G')$  trở thành một ma trận vuông. Giá trị hằng số là mức trung bình của tất cả các giá trị khác trong  $C(G, G')$ . Có nghĩa là chúng ta tìm kiếm một sự tương ứng 1-1 giữa các nút cuối trong đồ thị xương, với một số nút cuối có thể gán giá trị hằng số biểu diễn một nút giả.

Cách tiếp cận này không yêu cầu bất kỳ tương ứng nào với các nút giao nhau. Điều này là vô cùng quan trọng, vì trong nhiều trường hợp, sự tương ứng của các nút giao nhau là không thể thiết lập trực tiếp và do đó phương pháp chỉnh sửa đồ thị hoặc cây là cần thiết nếu cần sự tương ứng của các nút giao nhau. Một điều quan trọng cần tuân thủ là không thể thay đổi cấu trúc của các nút đường giao nhau với việc cắt tia xương mà không cần loại bỏ một số nút kết thúc quan trọng. Mặt khác, cắt tia xương có thể làm giảm các tập hợp các nút cuối để đến các nút cấu trúc liên quan bằng cách loại bỏ các nút giả.

Phương pháp này đo độ tương tự dựa trên hình dạng xương. Nó đối sánh đồ thị xương dựa trên sự tương tự của đường đi ngắn nhất, đường đi ngắn nhất giữa mỗi cặp điểm cuối xương được biểu diễn bằng chuỗi bán kính đĩa cực đại của điểm xương tương ứng.



Bảng 3.1

	1	2	3	4	5	6	7	8	9
1	35.45189	35.45158	35.50421	35.48862	35.4937	35.47303	35.47655	35.49843	35.48308
2	21.35382	21.33255	21.28259	21.29943	21.29645	21.30498	21.30378	21.34479	21.3382
3	19.47639	19.45347	19.44011	19.42756	19.42772	19.43372	19.43354	19.48357	19.47746
4	20.1647	20.15767	20.14791	20.13876	20.13836	20.16035	20.16054	20.15541	20.1464
5	35.07348	35.09428	35.08198	35.07433	35.07344	35.08386	35.08424	35.10599	35.09634
6	11.582	11.59656	11.58794	11.58936	11.58922	11.56833	11.56781	11.58588	11.57992
7	34.28898	34.31338	34.28353	34.29634	34.29302	34.27062	34.26897	34.26342	34.26782
8	11.59766	11.62881	11.6159	11.61471	11.61385	11.60145	11.6011	11.59846	11.59246
9	22.57204	22.606	22.65369	22.63052	22.63549	22.60589	22.60808	22.62612	22.61335

Hình 3.3 cho thấy bộ xương của hai con voi khác nhau tương ứng với các nút cuối cùng được liên kết bởi các đường. Ta lập chỉ mục các nút để các nút tương ứng có các chỉ số tương tự. Ma trận  $C(G, G')$  được thể hiện trong bảng trên. Các chi phí đối sánh giữa các nút kết thúc tương tự nhất được đánh dấu màu đỏ.

### 3.1.3 Tối ưu dãy song ánh (Optimal subsequence bijection)

Thuật toán OSB được sử dụng để đối sánh hai chuỗi có độ dài khác nhau  $m$  và  $n$ . Cụ thể hơn, hai chuỗi hữu hạn các nút cuối của xương  $a=(a_1, \dots, a_m)$  và  $b=(b_1, \dots, b_n)$ . Mục đích là để tìm chuỗi con  $a'$  của  $a$  và  $b'$  của  $b$  như là  $a'$  là phù hợp nhất với  $b'$ . Bỏ qua một số yếu tố của  $a$  và  $b$  là cần thiết bởi vì cả hai chuỗi có chứa một số yếu tố outlier. Tuy nhiên bỏ qua quá nhiều yếu tố của một chuỗi làm gia tăng kết quả không phù hợp. Để ngăn chặn điều này xảy ra phải có một hình phạt để bỏ qua các yếu tố. Hình phạt có thể được biểu diễn dưới dạng đối sánh thêm một số yếu tố  $b_\infty$ . Do đó, ta mở rộng chuỗi  $b$  thêm một yếu tố  $b_\infty$ .

Mục tiêu của phương pháp này là để tìm sự tương ứng tốt nhất có thể của một dãy  $a$  đến một dãy con  $b'$  của  $b$ . Ở đây xác định một sự tương ứng  $f: \{1, \dots, m\} \rightarrow \{1, \dots, n, \infty\}$  như một ánh xạ đơn điệu cho phạm vi giới hạn của hàm  $f: \{1, \dots, m\} \rightarrow \{1, \dots, n\}$  là một hàm  $f$  với  $f(i) < f(i+1) < \infty$  với  $a_i$  là ánh xạ đến  $b_{f(i)}$  của tất cả  $i \in \{1, \dots, m\}$  và cho phép một ánh xạ nhiều-một đến  $\infty$ .



Đặt  $f(i) = \infty$  có nghĩa chúng ta bỏ qua yếu tố  $i$  trong dãy  $a$ . Các bộ chỉ số  $i_k$  và  $(f(i_k))$  với  $f(i_k) < \infty$ ,  $i \in \{1, \dots, m\}$  xác định dãy con  $a'$  của  $a$  và  $b'$  của  $b$  sao cho  $f$  hạn chế  $i_k$  là một song ánh.

Giả định rằng các hàm khoảng cách  $d$  (distance function  $d$ ) được đưa ra có thể tính toán giá trị khác biệt giữa bất kỳ yếu tố của  $a$  và  $b$ , đó là  $d(a_i, b_j)$  với  $(i, j) \in (1, \dots, m) \times \{1, \dots, n, \infty\}$ . Ở đây không hạn chế chức năng khoảng cách  $d$  và do đó bất kỳ hàm khoảng cách nào cũng đều có thể thực hiện được. Trong phương pháp được đưa ra này sử dụng khoảng cách đường đi  $pd$  được định nghĩa trong (3.4) là khoảng cách  $d$ . Mặc dù hầu hết các ứng dụng  $d(a_i, b_j)$  được đưa ra cho  $(i, j) \in \{1, \dots, m\} \times \{1, \dots, n\}$  việc bổ sung yếu tố khoảng cách  $d(a_i, b_\infty)$  nên được lựa chọn cẩn thận. Thường  $d(a_i, b_\infty)$  là một hằng số cho tất cả  $i \in \{1, \dots, m\}$  để xác định chi phí bỏ qua bất kỳ yếu tố được đưa ra trong chuỗi  $a$ . Nó được gọi là hằng số *jumpcost*, được tính như sau:

$$\text{jumpcost} = \text{mean}_i(\min_j(d(a_i, b_j)) + \text{std}_i(d(a_i, b_j))) \quad (3.7)$$

Như vậy, mỗi yếu tố  $a_i$  tìm thấy yếu tố  $b_j$  gần nhất, và sau đó lấy trung bình cộng với một độ lệch chuẩn (std) của khoảng cách tới các yếu tố gần nhất. Ví dụ, nếu chuỗi  $a$  và  $b$  là tương tự với các trường hợp ngoại lệ của một trong những yếu tố outlier, gọi nó là  $a_k$ , mỗi  $a_i$  với  $i \neq k$  sẽ tìm thấy một phần tử  $b_j$  với một khoảng cách nhỏ  $d(a_i, b_j)$ . Do đó *jumpcost* sẽ là nhỏ, để khoảng cách đến các yếu tố gần nhất trong  $b$  cho  $a_k$  sẽ lớn hơn *jumpcost* và yếu tố  $a_k$  sẽ bị loại khỏi đối sánh với một hình phạt tương đối nhỏ. Đối với bất kỳ sự tương ứng nhất định, chúng ta có thể xác định khoảng cách giữa hai trình tự như sau:

$$d(a, b, f) = \frac{1}{m} \sum_{i=1}^m d(a_i, b_{f(i)})^2 \quad (3.8)$$

Mục tiêu là tìm một  $f$  tương ứng để  $d(a', b', f)$  là tối thiểu. Chính xác hơn là một  $f$  tương ứng tối ưu của các yếu tố trong dãy  $a$  tới các yếu tố trong dãy  $b$  được định nghĩa là giá trị nhỏ nhất của  $d(a, b, f)$  trên tất cả các  $f$  tương ứng:

$$f^{\wedge} = \arg \min \{d(a, b, f) : f \text{ là một tương ứng}\} \quad (3.9)$$

Cuối cùng, khoảng cách tối ưu được cho bởi công thức [3.8] cho  $f=f^*$ . Sự tương ứng tối ưu có thể được tìm thấy với các thuật toán đường đi ngắn nhất trên đồ thị có hướng(DAG). Biểu thị khoảng cách tối ưu  $d(a,b)$  trong [3.10] với  $OSB(a,b)$  cho DAG là tất cả các cặp chỉ số  $(i,j) \in \{1, \dots, m\} \times \{1, \dots, n\}$  và trọng số cạnh  $w$  được định nghĩa là:

$$w((i,j), (k,l)) = \begin{pmatrix} d(a_i, b_j) & \text{if } i+1=k \text{ and } j+1 \leq l \\ (k-i-1) \cdot \text{jumpcost} & \text{if } i+1 < k \text{ and } j+1 \leq l \\ \infty & \text{otherwise} \end{pmatrix} \quad (3.10)$$

## 3.2 Đối sánh đồ thị xương dựa trên các điểm quan trọng bằng cách sử dụng đường dẫn tương tự

### 3.2.1 Giới thiệu

Ý tưởng chính của phương pháp là đối sánh các điểm quan trọng (điểm giao và điểm cuối) trên đồ thị xương bằng cách so sánh các đường dẫn đo được giữa các điểm cuối và các điểm giao của xương. Phương pháp này được Yao Xu, Bo Wang, Wenyu Liu và Xiang Bai đề xuất dựa trên một thực tế là các điểm giao có thể mang thông tin về cấu trúc toàn cục của một đối tượng trong khi các đường dẫn giữa các điểm giao và các điểm cuối có thể đại diện cho thông tin cụ thể về hình dạng cho các bộ phận cục bộ. Phương pháp này mang lại độ chính xác đầy hứa hẹn trên hai tập dữ liệu hình có sự hiện diện của khớp nối, kéo dài, biến dạng ranh giới và xoay.

### 3.2.2 Trộn các nút giao nhau (Mergence of junction nodes)

Các nút giao trong xương mang thông tin về cấu trúc toàn cầu của một đối tượng nhưng có một thực tế là sẽ có rất nhiều điểm giao trong một đồ thị xương. Do đó việc xác định chính xác các điểm giao để đối sánh là một việc vô cùng quan trọng.

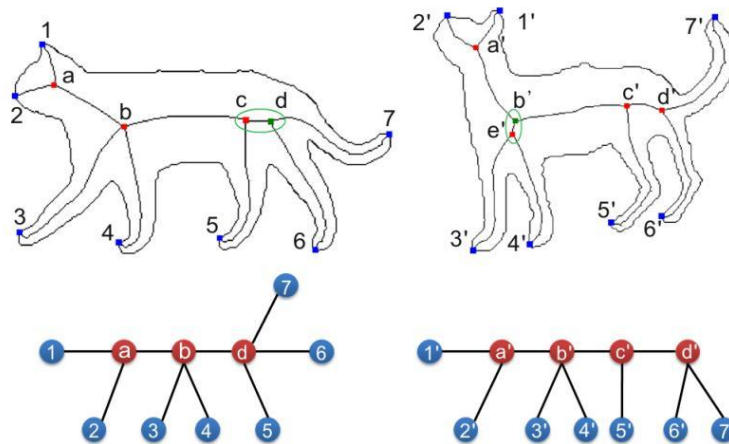
Giả sử có  $N$  đường giao nhau trong một xương. Chi phí để hợp nhất hai nút giao nhau  $V_i$  và  $V_j$  được định nghĩa như sau:

$$\text{cost}(V_i, V_j) = \sum_{k=1}^N \text{pd}(\text{sp}_{i,k}, \text{sp}_{j,k}) \quad (3.11)$$

Với  $sp_{i,k}$  và  $sp_{j,k}$  chỉ những đường junction-to-end giữa tất cả các nút kết thúc và các nút giao nhau,  $V_i$  và  $V_j$  và  $k$  là chỉ số của các nút kết thúc theo một hướng ngược chiều kim đồng hồ. Và điều kiện sát nhập là như sau:

$$\text{cost}(V_i, V_j) < N * \delta \quad (3.12)$$

Với  $\delta$  là một giá trị nhỏ như một ngưỡng. Bất kỳ cặp các nút giao nhau đáp ứng điều kiện (3.12) sẽ được sát nhập. Hình 3.4 minh họa một ví dụ về quá trình hợp nhất ở trên.



Hình 3.4 Minh họa hợp nhất các nút

Như hình trên cho thấy, các nút giao nhau  $c$  và  $d$  được sát nhập như một nút  $d$  duy nhất giao nhau vì chúng đáp ứng điều kiện (3.12), tương tự với  $b'$  và  $e'$ . Trong việc thực hiện này, các nút giao nhau  $c$  và  $d$  không được hợp nhất như một nút thực sự. Thay vào đó chỉ có một trong hai nút sẽ được lựa chọn làm nút đối sánh.

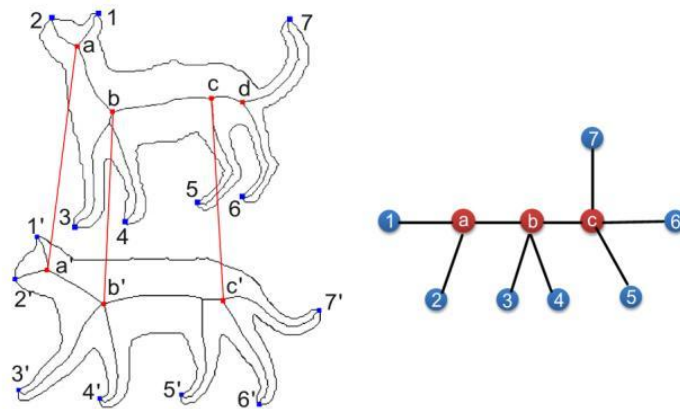
### 3.2.3 Đối sánh các nút quan trọng (Matching Critical Nodes)

Ta có  $G$  và  $G'$  biểu thị hai đồ thị để đối sánh, và để cho số các nút đường giao nhau trong  $G$  và  $G'$  tương ứng là  $K$  và  $K'$ . Giả sử  $K \leq N$ . Thật dễ để biết có  $C_N^K * K!$  các trường hợp phù hợp và mục tiêu đặt ra là có được đối sánh một-một tối ưu với chi phí tối thiểu. Trong trường hợp đó, hai đồ thị có số lượng các nút giao nhau khác nhau, hoạt động cắt giảm sẽ được thực hiện bằng cách bỏ qua

các nút giao nhau dư thừa. Ví dụ có một nút giao nhau  $V_1$  trong  $G$ , hai nút giao nhau  $V_1', V_2'$  trong  $G'$ , do đó có hai trường hợp:

$$V_1 \leftrightarrow V_1' \text{ or } V_1 \leftrightarrow V_2'$$

Trong hầu hết các trường hợp, sẽ xảy ra tình huống đối sánh phức tạp hơn. Trong hình 3.5 sau khi đối sánh các nút giao nhau và hoạt động cắt giảm sẽ thu được điểm quan trọng. Sau đó sẽ nhận được cấu trúc chung của xương.



Hình 3.5 Minh họa đối sánh các nút giao nhau

Đối với bất kỳ sự đối sánh nào của các cặp nút giao nhau  $V$  và  $V'$ , giả sử số lượng các nút cuối tiếp giáp với  $V$  và  $V'$  tương ứng là  $m$  và  $n$ . Giả định  $m \leq n$ . Như vậy có  $C_n^m * m!$  các sự lựa chọn phù hợp. Bằng cách này có thể nhận được  $U$  các loại lựa chọn phù hợp. Một sự lựa chọn phù hợp có chi phí phù hợp và mục tiêu là có được chi phí tối thiểu. Giả sử có  $P_k$  đường đối sánh và  $Q_k$  đường chưa đối sánh với  $k$ th lựa chọn đối sánh ( $k$  là chỉ số của sự lựa chọn phù hợp), do đó mô hình có thể được biểu diễn như sau:

$$\arg \min(cost_k), k = 1, 2, \dots, U.$$

$$cost_k = \left(1 + \frac{Q_k}{P_k}\right) \sum_{i=1}^{P_k} pd(sp_i, sp'_i) \quad (3.13)$$

Với  $sp_i$  và  $sp'_i$  biểu diễn cho đường xương trong các đồ thị để đối sánh,  $Q_k/P_k$  chức năng như là một yếu tố hình phạt.

---

## CHƯƠNG 4: THỰC NGHIỆM

### 4.1 Môi trường thực nghiệm

Chương trình được cài đặt trên môi trường Windows XP SP3, sử dụng ngôn ngữ Matlab Version 7.7.0.471 (R2008b).

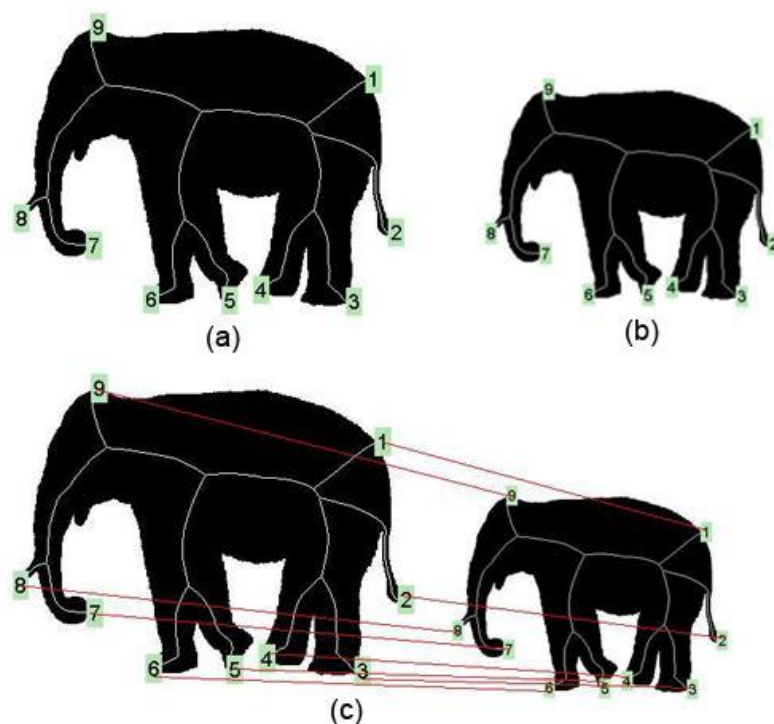
Dữ liệu được sử dụng trong thử nghiệm là tập dữ liệu thuộc MPEG-7.

Ảnh được dùng trong thử nghiệm thuộc định dạng .TIF với đối tượng đã được phân tách. Xương của đối tượng trong ảnh được xác định bởi kỹ thuật tìm xương DCE.

### 4.2 Một số kết quả

#### 4.2.1 Thay đổi kích thước ảnh mẫu

Ảnh mẫu thuộc kiểu .TIF, đối tượng là một con voi đã được cắt tia xương bằng kỹ thuật tìm xương DCE. Ảnh đối sánh là ảnh mẫu sau đã giảm kích thước 30%.

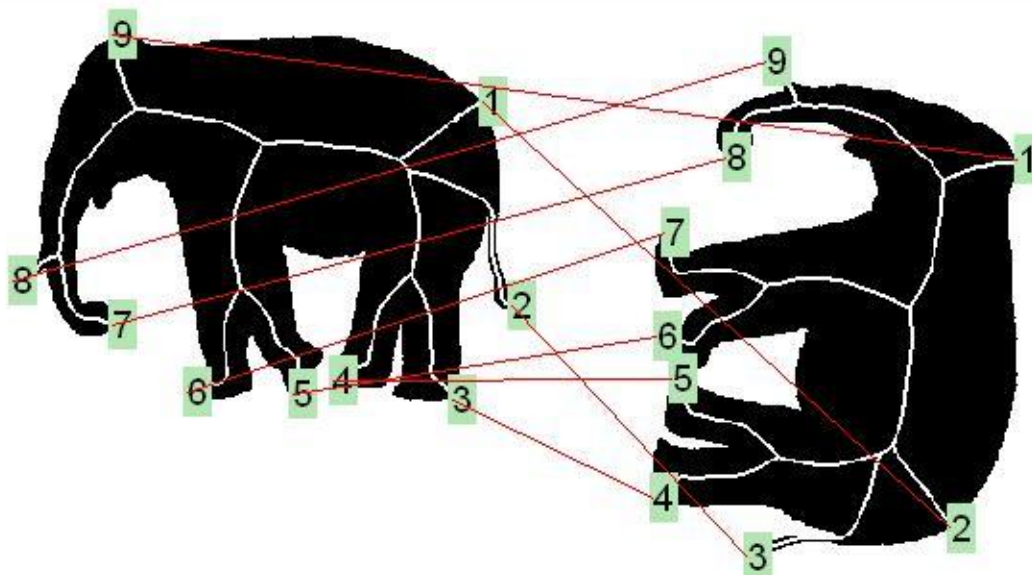


Hình 4.1 Kết quả thu được khi thay đổi kích thước ảnh mẫu

(a): Ảnh mẫu, (b): Ảnh mẫu sau khi giảm kích thước 30%, (c): Kết quả đối sánh.

### 4.2.2 Xoay ảnh mẫu một góc $\alpha$

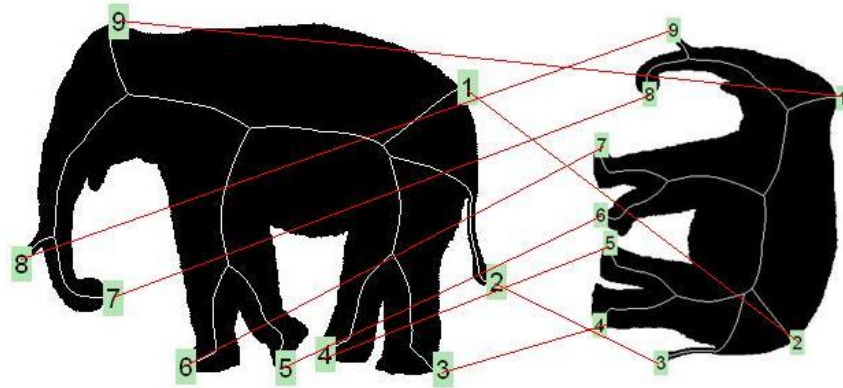
Ảnh mẫu thuộc kiểu .TIF, đối tượng là một con voi đã được cắt tia xương bằng kỹ thuật tìm xương DCE. Ảnh đối sánh là ảnh mẫu được quay 1 góc  $90^0$  theo chiều kim đồng hồ.



Hình 4.2 Kết quả thu được khi thay đổi góc quay của ảnh mẫu một góc  $\alpha=90^0$

### 4.2.3 Đồng thời thay đổi kích thước và góc xoay của ảnh

Ảnh mẫu thuộc kiểu .TIF, đối tượng là một con voi đã được cắt tia xương bằng kỹ thuật tìm xương DCE. Ảnh đối sánh là ảnh mẫu bị giảm kích thước 30% đồng thời bị xoay một góc  $\alpha=90^0$  theo chiều kim đồng hồ.



Hình 4.3 Kết quả thu được khi đồng thời thay đổi kích thước và góc quay ảnh

(a): Ảnh mẫu, (b): Ảnh mẫu giảm kích thước 30% và xoay  $90^{\circ}$ , (c): Kết quả

### 4.3 Một số nhận xét về chương trình

Chương trình mô tả phương pháp đối sánh dựa trên sự tương tự của đồ thị xương và các điểm quan trọng. Sau đây là danh sách một số tập tin và chức năng của nó trong chương trình:

- Matching.m: Hàm chính để chạy chương trình.
- GetSkeletonPath.m: Lấy xương của đối tượng, các điểm endpoints, giá trị bán kính đĩa tối đa tới biên của các điểm nằm trong đối tượng.
- GetBranchPath.m: Xác định các nhánh xương của đối tượng, và các nhánh xương sau khi được chuẩn hóa.

Các kết quả thử nghiệm cho kết quả khá tốt với ảnh thay đổi kích thước và thay đổi góc quay của ảnh vì độ dài của các đường dẫn xương đã được chuẩn hóa với số điểm cách đều  $M$  là như nhau với các đối tượng, thứ hai là phương pháp đối sánh được sử dụng căn cứ vào độ dài đường đi ngắn nhất giữa các điểm cuối mà không quan tâm đến sự tương ứng giữa các nút cuối. Tuy nhiên với các ảnh bị biến dạng thì kết quả nhận được có thể không như mong muốn vì phương pháp được sử dụng là không bất biến với sự biến dạng của đối tượng.



---

## KẾT LUẬN

Sau một thời gian tìm hiểu và nghiên cứu, đồ án đã trình bày được một số vấn đề sau:

- **Về lý thuyết:** Trình bày tổng quan về tra cứu ảnh dựa trên nội dung, một số phương pháp trích chọn đặc trưng dựa trên hình dạng và đặc biệt trình bày cụ thể phương pháp đối sánh dựa trên xương và các điểm quan trọng.

- **Về thực nghiệm:** Cài đặt chương trình thử nghiệm đối sánh ảnh dựa trên đồ thị xương phục vụ cho việc đối sánh và tra cứu ảnh.

Tuy nhiên trong quá trình thực hiện, thời gian không có nhiều, năng lực chuyên môn còn hạn chế, nên đề tài mới chỉ dừng ở mức dịch, hiểu và tóm lược về phương pháp.

Em rất mong nhận được sự đóng góp ý kiến của các Thầy Cô và các bạn để em có thêm kiến thức và kinh nghiệm để tiếp tục hoàn thiện nội dung nghiên cứu trong đề tài.

Em xin chân thành cảm ơn!



---

## TÀI LIỆU THAM KHẢO

Tài liệu tiếng Việt:

[1] Lương Thị Hoài Xuân, “*Tìm hiểu phương pháp DSE cho bài toán tìm xương của ảnh.*” Đồ án Tốt nghiệp, Đại học Dân lập Hải Phòng, 2011.

[2] Phùng Thị Lê, “*Tìm hiểu phương pháp trích chọn dấu hiệu của ảnh dựa vào đặc trưng hình dạng.*” Đồ án Tốt nghiệp, Đại học Dân lập Hải Phòng, 2011.

[3] Nguyễn Thị Lan, “*Tìm hiểu phương pháp BPR (Bending Potential Ratio) cho bài toán tìm xương của ảnh.*” Đồ án tốt nghiệp, Đại học Dân lập Hải Phòng, 2011.

Tài liệu tiếng Anh:

[3] Ritendra Datta, Jia Li, James Z. Wang, “*Content-Based Image Retrieval- Approaches and Trends of the New Age.*” The Pennsylvania State University, University Park, PA 16802, USA.

[4] Dr.Fuhui Long, Dr.Hongjiang Zhang and Prof. David Dagan Feng, “*Fundamentals of Content-Based Image Retrieval.*”.

[5] Dengsheng Zhang\*, Goujun Lu, “*Review of shape representation and description techniques.*” Gippsland School of Computing and Info. Tech., Monash University, Churchill, Vic 3842, Australia.

[6] Xiang Bai and Longin Jan Latecki, “*Path Similarity Skeleton Graph Matching.*” IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008.

[7] Yao Xu, Bo Wang, Wenyo Liu, and Xiang Bai, “*Skeleton Graph Matching Based on Critical Points Using Path Similarity.*” Department of Electronics and Information Engineering , Huazhong University of Science and Technology, Wuhan, 430074, China.