

MỤC LỤC

MỤC LỤC	1
DANH MỤC CÁC BẢNG	3
DANH MỤC CÁC HÌNH	4
DANH MỤC CHỮ VIẾT TẮT	6
LỜI CẢM ƠN	7
MỞ ĐẦU	8
CHƯƠNG 1:TỔNG QUAN VỀ TRA CỨU ẢNH DỰA TRÊN NỘI DUNG	9
1.1Giới thiệu về thuật ngữ “Tra cứu ảnh dựa trên nội dung”	9
1.2Thành phần chính của một thể thống tra cứu ảnh dựa trên nội dung.....	11
1.2.1 Công nghệ tự động trích chọn metadata.	11
1.2.2 Giao diện để lấy yêu cầu truy vấn của người sử dụng.....	11
1.2.3 Phương pháp để so sánh độ tương tự giữa các ảnh.....	12
1.2.4 Công nghệ tạo chỉ số và lưu trữ dữ liệu hiệu quả	13
1.3 Công nghệ sử dụng trong hệ thống tra cứu ảnh dựa trên nội dung.....	13
1.3.1Công nghệ trích chọn đặc trưng trực quan.....	13
1.3.2 Độ đo tương tự sử dụng đặc trưng trực quan.....	18
1.3.3 Phân cụm và phân lớp	20
1.3.4 Phản hồi liên quan.....	22
1.4 Ứng dụng của Tra cứu ảnh dựa trên nội dung	25
1.4.1 Văn hóa nghệ thuật	25
1.4.2 Truyện tranh	25
1.4.3 Bảo mật và hình ảnh.....	26
1.5 Những hướng phát triển hệ thống trong tương lai	27
CHƯƠNG 2:..TRA CỨU ẢNH DỰA TRÊN ĐẶC TRƯNG HÌNH DẠNG	28
2.1 Giới thiệu.....	28
2.2 Trích chọn đặc trưng IDSC	29

2.2.1 Giới thiệu.....	29
2.2.2 Ngữ cảnh hình dạng (Shapes Context)	31
2.2.3 Khoảng cách trong (THE INNER DISTANCE).....	32
2.2.4 Inner-Distance Shape Context	34
2.3 Đối sánh shape sử dụng quy hoạch động.....	35
2.4 Tra cứu ảnh với kỹ thuật học hàm khoảng cách	36
2.4.1 Học hàm khoảng cách (Learning New Distance Measures).....	39
CHƯƠNG 3: CÀI ĐẶT CHƯƠNG TRÌNH VÀ ỨNG DỤNG.....	42
3.1 Môi trường thực nghiệm	42
3.2 Bài toán.....	42
3.3 Các bước chính của chương trình	42
3.4 Một số hình ảnh của chương trình	44
KẾT LUẬN	53
TÀI LIỆU THAM KHẢO	54

DANH MỤC CÁC BẢNG

Bảng 1.2 : Bảng so sánh hai phương pháp phân cụm và phân lớp	20
Bảng 1.3: Tổng hợp các phương pháp phân cụm	21

DANH MỤC CÁC HÌNH

Hình 1.1. Kiến trúc tổng quan của hệ thống tra cứu	11
Hình 2.1: Ví dụ về khoảng cách trong của đối tượng	30
Hình 2.2: Ví dụ về khoảng cách trong của x và y trong hình O	32
Hình 2.3: Quá trình biểu diễn khoảng cách trong của đối tượng.....	33
Hình 2.4: Minh họa về góc trong (Inner - Angle).....	34
Hình 2.5: Ngữ cảnh hình dạng (SC) và khoảng cách trong ngữ cảnh hình dạng (IDSC)	35
Hình 2.6 : Ví dụ về khoảng cách trong ngữ cảnh hình dạng	37
Hình 2.7: Các hình đã biết trong cơ sở dữ liệu.	38
Hình 3.1: Giao diện chính của chương trình.....	44
Hình 3.2: Giao diện chọn thu mục chứa CSDL đặc trưng của ảnh	45
Hình 3.3: Giao diện chọn thư mục chứa ảnh truy vấn cần tra cứu	46
Hình 3.4: Giao diện hiển thị ảnh truy vấn cho ví dụ 1	47
Hình 3.5: Giao diện hiển thị kết quả tra cứu trước khi học hàm khoảng cách của ví dụ 1	47
Hình 3.6: Giao diện kết quả hiển thị sau khi áp dụng kĩ thuật học hàm khoảng cách của ví dụ 1	48
Hình 3.7: Giao diện hiển thị ảnh truy vấn của ví dụ 2	49
Hình 3.8: Giao diện hiển thị kết quả tra cứu trước khi học hàm khoảng cách của ví dụ 2	49
Hình 3.9: Giao diện hiển thị kết quả tra cứu sau khi học hàm khoảng cách của ví dụ 2	50
Hình 3.10: Giao diện hiển thị ảnh truy vấn của ví dụ 3	51
Hình 3.11: Giao diện hiển thị kết quả tra cứu của ví dụ 3	51
Hình 3.12: Giao diện hiển thị kết quả tra cứu sau khi học hàm khoảng cách của ví dụ 3	52

DANH MỤC CHỮ VIẾT TẮT

CBIR	CONTENT BASED IMAGE RETRIEVAL
IDSC	INNER DISTANCE SHAPE CONTEXT
TID	THE INNER DISTANCE
DP	DYNAMIC PROGRAM
SC	SHAPE CONTEXT
QBIR	QUERY BASED IMAGE RETRIEVAL

LỜI CẢM ƠN

Em xin gửi lời cảm ơn chân thành tới thầy giáo hướng dẫn Th.s Ngô Trường Giang, người đã định hướng nghiên cứu và tận tình chỉ bảo, giúp đỡ em trong quá trình làm đồ án, giúp em hoàn thành báo cáo thực tập đúng kế hoạch. Em xin chân thành cảm ơn các thầy cô trong khoa công nghệ thông tin trong Trường ĐHDL Hải Phòng đã tận tình giảng dạy, truyền đạt những kiến thức và kinh nghiệm vô cùng quý báu trong những năm học vừa qua. Em xin cảm ơn gia đình, bạn bè đã bên cạnh và động viên em trong quá trình làm đồ án để có được kết quả như ngày hôm nay.

Em xin chân thành cảm ơn!

Hải phòng, ngày.....tháng....năm 2014

Sinh viên

Bùi Đức Sơn

MỞ ĐẦU

Sự phát triển của công nghệ thông tin và mạng internet kéo theo sự gia tăng đáng kể của các loại dữ liệu đa phương tiện, dữ liệu số, nhu cầu tìm kiếm và chia sẻ những loại dữ liệu đó ngày càng tăng. Kiểu dữ liệu hình ảnh là một trong số đó. Tìm kiếm theo hình ảnh mang tính trực quan sẽ dễ dàng cho người tìm kiếm thông tin đi kèm ảnh, tìm kiếm ảnh mang nhiều mục đích như để giải trí, thỏa mãn nhu cầu cá nhân, phục vụ cho nhiều lĩnh vực trong cuộc sống như an ninh bảo mật, y tế, giáo dục... Do vậy việc nghiên cứu và phát triển các hệ thống tra cứu ảnh ngày càng trở nên cấp thiết.

Hai kiểu tìm kiếm phổ biến đó là tìm kiếm ảnh theo nội dung và tìm kiếm ảnh dựa vào từ khóa. Mỗi phương pháp đều có ưu nhược điểm riêng, ưu điểm của tìm kiếm ảnh theo từ khóa là nhanh và dễ dàng hơn cho các bộ máy tìm kiếm vì nó hoạt động dựa trên việc phân tích và so sánh các từ hoặc cụm từ tương ứng với nhau để đưa ra kết quả, kiểu dữ liệu là các dạng văn bản, từ ngữ cho nên sẽ nhanh chóng cho ra kết quả, không đòi hỏi phải có ảnh mẫu để so sánh. Nhược điểm của nó là độ chính xác của kết quả tìm kiếm không cao, bởi việc đánh từ khóa cho từng ảnh không phải lúc nào cũng đầy đủ theo ý của người tìm kiếm, nó phù hợp với việc đáp ứng như cầu người dùng thông qua việc mô tả bằng từ ngữ. Phương pháp thứ hai là phương pháp tra cứu ảnh dựa vào nội dung của hình ảnh. Phương pháp này cần một ảnh mẫu cho đầu vào rồi sau đó tìm ra bức ảnh tương ứng, độ chính xác của phương pháp này cao hơn, bởi nó mô tả tương đối đầy đủ nội dung của bức ảnh đầu vào sau đó đối chiếu với ảnh trong cơ sở dữ liệu, từ đó cho ra bức ảnh chính xác nhất.

CHƯƠNG 1: TỔNG QUAN VỀ TRA CỨU ẢNH DỰA TRÊN NỘI DUNG

1.1 Giới thiệu về thuật ngữ “Tra cứu ảnh dựa trên nội dung”

Tra cứu ảnh theo nội dung chính thức xuất hiện từ năm 1992, đánh dấu bằng Hội thảo về các hệ thống quản lý thông tin trực quan của Quỹ Khoa học Quốc gia của Hoa Kỳ.

Tra cứu ảnh dựa trên nội dung là một quá trình tìm kiếm ảnh trong một cơ sở dữ liệu ảnh những ảnh nào thỏa mãn một yêu cầu nào đó. Những tìm kiếm đặc thù vào tiêu biểu cho hệ thống này là: QBIC, VIR Image, Engine, VisualSEEK, NeTrA, MARS, Viper...

Thuật ngữ tra cứu ảnh dựa trên nội dung đã được Kato sử dụng đầu tiên để mô tả những thí nghiệm của ông về lĩnh vực tra cứu tự động những hình ảnh từ một cơ sở dữ liệu dựa trên đặc điểm về hình dạng và màu sắc. Từ đó, nó được sử dụng rộng rãi để mô tả quá trình tra cứu những hình ảnh mong muốn từ một tập hợp lớn hình ảnh dựa trên những đặc điểm về màu sắc, kết cấu và hình dạng, và những đặc điểm đó được trích rút một cách tự động từ chính những hình ảnh đó.

Tra cứu ảnh dựa trên nội dung (CBIR) là việc ứng dụng của thị giác máy tính (Computer Vision) vào việc tra cứu ảnh, tránh sử dụng miêu tả bằng từ khóa, thay vào đó là sử dụng các sự tương đồng trong nội dung của ảnh như : kết cấu, màu sắc, hình dạng.

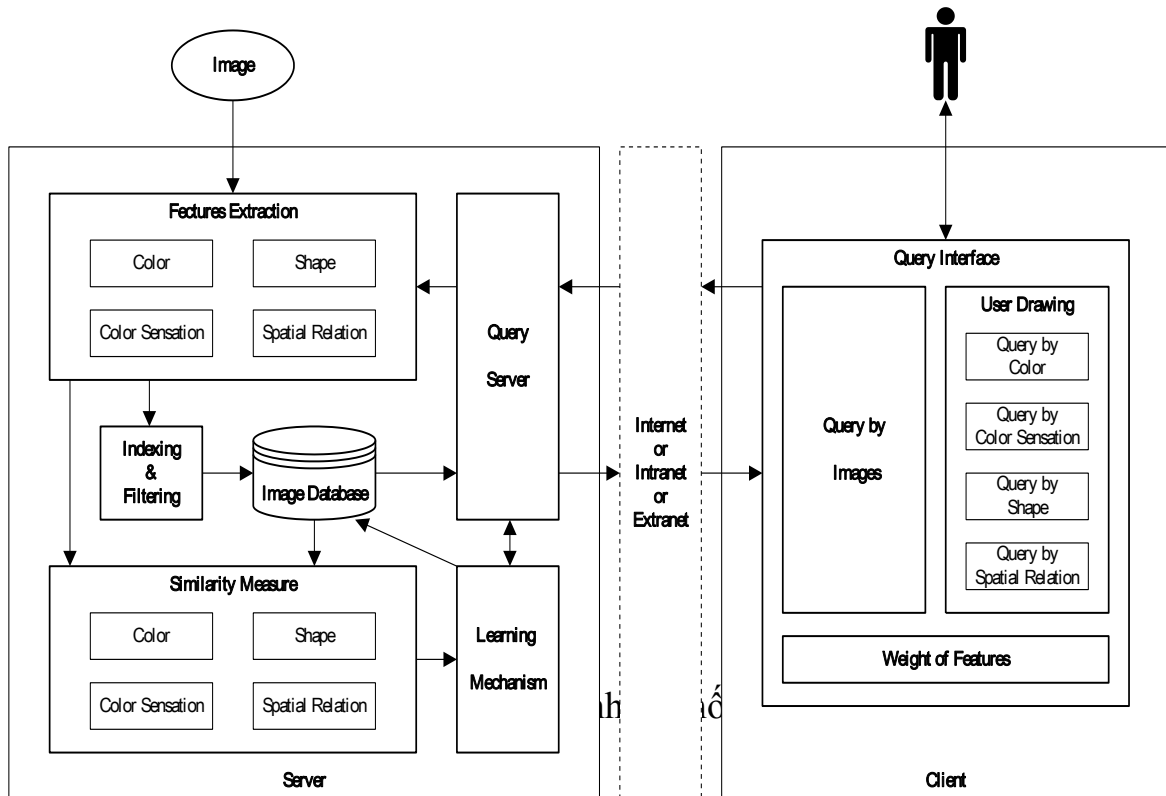
Tra cứu ảnh dựa trên nội dung trái ngược với tra cứu ảnh dựa trên từ khóa, nó dựa trên một số đặc trưng mức thấp (Low-Level features): Màu Sắc (Colors), hình dạng (Shape), kết cấu (Textures) và liên hệ không gian (Spatial relationship).

- *Màu sắc*: Là đặc trưng cơ bản và phổ biến, giúp con người dễ dàng nhận ra sự khác biệt giữa các hình ảnh, dùng lược đồ màu sắc (Color Histogram) để biểu diễn. Tìm kiếm ảnh theo màu sắc tiến hành tính toán biểu đồ cho ảnh để xác định vị trí các điểm ảnh chứa giá trị đặc biệt được xét trong không gian màu (RGB, CIE, HSV).
- *Hình dạng (Shape)*: Là đặc trưng khá quan trọng trong khâu trích chọn đặc trưng từ nội dung hình ảnh của đối tượng, chúng ta có thể tiếp cận dựa trên vùng (Region) của đối tượng, hoặc dựa trên biên của đối tượng hay còn gọi là các thuộc tính hình học của ảnh. Thuộc tính hình học được dùng trong phân lớp, so sánh và nhận dạng đối tượng.
- *Kết cấu*: Các thuộc tính của kết cấu (Tương phản, Hướng, Quy luật, Chu kỳ, Ngẫu nhiên), đặc trưng này được trích chọn nhằm tìm ra mô hình trực quan của ảnh và cách thức ảnh được xác định trong không gian, cung cấp thông tin về sự sắp xếp về mặt không gian của màu sắc và cường độ ảnh.
- *Liên hệ không gian*: Được dùng để phân biệt các đối tượng trong một ảnh theo hai cách đó là theo đối tượng và theo quan hệ.

Những phương pháp dựa trên đặc trưng mức thấp đang được phát triển tốt trong thời gian gần đây, tuy nhiên việc dựa trên những đặc trưng mức thấp không phải lúc nào cũng mang lại hiệu quả như mong muốn cho người sử dụng, do vậy phải kết hợp những đặc điểm đó lại với nhau, ngày nay, nhiều hệ thống CBIR đã được đưa vào sử dụng, tiêu biểu là hệ thống tra cứu hình ảnh Image Search của Google, Yahoo, Bing.

1.2 Thành phần chính của một thể thống tra cứu ảnh dựa trên nội dung

Một hệ thống tra cứu ảnh đòi hỏi các thành phần như hình 1.1



Hình 1.1. Kiến trúc tổng quan của hệ thống tra cứu

1.2.1 Công nghệ tự động trích chọn metadata.

Mỗi đặc điểm nguyên thủy của ảnh có định dạng đặc trưng của nó như biểu đồ màu được sử dụng rộng rãi để biểu thị đặc điểm màu sắc. Một ví dụ khác đặc điểm hình dạng có thể biểu thị bằng một tập các đoạn biên liên nhau. Với metadata thích hợp hệ thống tra cứu ảnh dựa trên nội dung có thể tra cứu ảnh bởi màu sắc, hình dạng, kết cấu và bởi sự kết hợp các đặc tính trên.

1.2.2 Giao diện để lấy yêu cầu truy vấn của người sử dụng

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

mình mong muốn với từ khóa nào đó trong thư viện. Với hệ thống tra cứu ảnh dựa trên nội dung thì quá trình tra cứu thường được thực hiện thông qua một hình ảnh mẫu được cung cấp bởi người sử dụng gọi là truy vấn bởi mẫu. Mặc dù vậy người sử dụng không thể luôn luôn đưa ra một ảnh mẫu cho hệ thống tra cứu. Hệ thống tra cứu ảnh dựa trên nội dung hiện nay giải quyết vấn đề này bằng cách đưa ra một giao diện để chỉ định hoặc chọn một số đặc điểm cơ bản cho việc cung cấp ảnh mẫu. Chẳng hạn như khi sử dụng hệ thống QBIC của IBM người sử dụng có thể chỉ định truy vấn đặc điểm màu sắc bằng cách chọn ra số lượng thành phần RED, BLUE, GREEN liên quan hoặc là có thể lựa chọn màu sắc ảnh mong muốn từ bảng màu, đồng thời người sử dụng có thể chọn kết cấu mong muốn cho đặc điểm kết cấu và vẽ ra một phác họa cho truy vấn đặc điểm hình dạng

1.2.3 Phương pháp để so sánh độ tương tự giữa các ảnh

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

quả của công nghệ đó phụ thuộc vào kiểu yêu cầu tra cứu mà người dùng sử dụng.

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

Đối với một tập dữ liệu ảnh lớn thì không gian lưu trữ cho metadata là rất cần thiết. Một hệ thống tra cứu ảnh dựa trên nội dung phải có những công nghệ hiệu quả để quản lý metadata đồng thời phải có chuẩn để mô tả nó. Chuẩn MP7 đang là chuẩn quan trọng nhất để mô tả metadata cho cả dữ liệu ảnh và dữ liệu video. Khi một truy vấn được xử lý trên một cơ sở dữ liệu lớn, việc so sánh độ tương tự giữa ảnh truy vấn và tất cả các hình ảnh từng cặp là không thể thực hiện được bởi người dùng chỉ cần những ảnh có độ tương tự cao so với ảnh mẫu. Những chỉ số cấu trúc có thể giúp tránh được việc tìm kiếm tuần tự và cải thiện truy vấn một cách hiệu quả nên được sử dụng trong hệ thống tra cứu ảnh dựa trên nội dung. Hơn nữa với những cơ sở dữ liệu ảnh thường xuyên thay đổi thì chỉ số cấu trúc động là rất cần thiết. Khi nội dung của ảnh được thể hiện bởi các vector low dimension và khoảng cách giữa các ảnh được định nghĩa(chẳng hạn như khoảng không gian được tính toán bằng khoảng cách Euclidean) cây R và các thành phần của nó có thể được sử dụng để đánh chỉ số cho ảnh. Khi khoảng cách không được định nghĩa như không gian vector hoặc khi không gian vector là High dimension hoặc khi mà những gì chúng ta có chỉ là một hàm khoảng cách tức là khoảng không metric thì những phương pháp để đánh chỉ số ảnh dựa trên hàm khoảng cách trong không gian metric là thích hợp.

1.3 Công nghệ sử dụng trong hệ thống tra cứu ảnh dựa trên nội dung

1.3.1 Công nghệ trích chọn đặc trưng trực quan

Hầu hết các hệ thống thực hiện việc trích chọn đặc trưng (Feature Extraction) như là một bước tiền xử lý, sau khi thu được, các đặc trưng trực

quan hoạt động như là đầu vào cho việc phân tích ảnh, cũng như so sánh độ tương tự, khái niệm và chú thích.

1.3.1.1 Phân đoạn ảnh

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

Có rất nhiều thuật toán để phân đoạn ảnh, K- mean là một trong số đó, ưu điểm chính của các thuật toán phân đoạn loại này là nó trích chọn các đường bao quanh từ một số lượng lớn các ảnh mà không chiếm thời gian và nỗ lực của con người, thuật toán phân cụm K-mean có tốc độ xử lý nhanh nhưng bên cạnh đó nó lại không được định nghĩa giống như các phương pháp nghiên cứu mới gần đây. Một trong những tiến bộ quan trọng nhất trong phân đoạn ảnh đó là phương pháp đồ thị cắt được Shi và Malik nghiên cứu rất kĩ.

Không thể phủ nhận rằng việc có được một phân đoạn tốt là một bước quan trọng trong việc hiểu rõ về ảnh, vấn đề khó khăn hiện nay đó là độ phức tạp của các thuật toán phân đoạn, độ tin cậy của phân đoạn tốt, và những phương pháp đánh giá một phân đoạn tốt.

1.3.1.2 Trích chọn toàn cục và trích chọn cục bộ

Trong hệ thống tra cứu ảnh, những đặc trưng được định nghĩa nhằm thu được những thuộc tính trực quan của ảnh, phương pháp trích chọn toàn cục khai thác trên toàn bộ ảnh, còn trích chọn cục bộ lại khai thác trên một nhóm các điểm ảnh gọi là nhóm pixel. Các đặc trưng này bao gồm: màu sắc, hình dạng, kết cấu, và các điểm nổi bật...

Trong trích chọn toàn cục, các đặc trưng sẽ được tính toán nhằm thu lại các đặc tính riêng của toàn bộ ảnh, ưu điểm của phương pháp này đó là tốc độ

tính toán cả trong quá trình trích chọn cũng như đo độ tương tự, tuy nhiên phương pháp này được cho là quá cứng nhắc cho việc mô tả hình ảnh

Trong trích chọn đặc trưng cục bộ, một tập hợp các tính năng được tính toán cho mỗi điểm ảnh bằng cách sử dụng những điểm lân cận của nó (hàng xóm). Để giảm lượng tính toán, ảnh đầu vào có thể được chia nhỏ thành các khối không chồng lấn, sau đó các tính năng sẽ được tính toán riêng cho mỗi khối. Các tính năng này vẫn là cục bộ bởi chúng nằm trong các khối đã được chia nhỏ, đa phần số lượng tính toán chỉ là một phần nhỏ các đặc trưng thu được xung quanh các điểm ảnh.

1.3.1.3 Đặc trưng màu sắc

Tầm quan trọng của đặc trưng màu sắc được sử dụng nhiều trong CBIR. Mới đây việc nghiên cứu không gian màu LUV cho kết quả tốt và sinh động hơn đối với người dùng so với không gian màu cơ bản RGB.

Việc nghiên cứu đặc trưng màu sắc tập trung hơn vào việc tổng hợp các màu sắc trong ảnh. Năm 2001, tác giả Manjunath et al [1] đã thử nghiệm kiểm tra một tập hợp màu sắc và kết cấu trong chuẩn MPEG-7, và kết quả là hoàn toàn phù hợp cho ảnh tự nhiên và video, bao gồm: mô tả dựa trên lược đồ, mô tả không gian màu, mô tả đặc trưng kết cấu phù hợp cho tra cứu ảnh.

1.3.1.4 Đặc trưng kết cấu

Kết cấu là một đặc trưng quan trọng khác của ảnh dùng để thu lại những đặc điểm gồ ghề và những dạng mẫu lặp lại nhiều lần của bề mặt ảnh. Các biểu diễn kết cấu đa dạng được nghiên cứu trong nhận dạng mẫu và thị giác máy tính. Về cơ bản các phương pháp kết cấu được phân làm hai loại: Kết cấu cấu trúc và kết cấu thống kê.

- Phương pháp cấu trúc bao gồm toán tử hình thái và đồ thị kề mô tả cấu trúc và nhận dạng cấu trúc, các luật sắp xếp của chúng, hiệu quả trong việc áp dụng với những kết cấu đều.
- Phương pháp thống kê gồm các kỹ thuật phổ năng lượng Fourier, các ma trận đồng khả năng, phân tích thành phần chính bất biến - trượt, đặc trưng Tamura, phân rã Wold, trường ngẫu nhiên Markov,...v.v. Một số biểu diễn kết cấu được sử dụng thường xuyên và đã được chứng minh là hiệu quả trong tra cứu nội dung ảnh.

Kết cấu cung cấp thông tin về sự sắp xếp về mặt không gian của màu sắc và cường độ của một ảnh, nó đặc trưng bởi sự phân bố không gian của những mức cường độ trong khu vực lân cận với nhau. Bao gồm các kết cấu gốc hay nhiều kết cấu gộp lại gọi là Texel.

Một số phương pháp dùng để trích chọn đặc trưng kết cấu

- Kim tự tháp “có thể lái được” (the steerable pyramid)
- Biến đổi đường viền (the cotourlet transform)
- Biến đổi sóng Gabor (The Gabor Wavelet transform)
- Biểu diễn ma trận đồng diện (co – occurrence matrix)

1.3.1.5 Trích chọn đặc trưng hình dạng

Hình dạng (Shape) là một đặc trưng quan trọng của việc phân đoạn vùng của ảnh, và tính hiệu quả và thiết thực của nó đóng vai trò quan trọng trong việc tra cứu ảnh. Phép biểu diễn hình dạng sử dụng đường cong rời rạc để làm đơn giản hóa đường viền giúp cho thuận lợi việc lọc nhiễu đã được hai tác giả Latecki và Lakamper nghiên cứu, ngoài ra việc sử dụng đường cong rời rạc còn loại bỏ được các đặc trưng hình dạng không thích hợp.

Một phương pháp mô tả hình dạng để đo độ tương tự đó chính là hình dạng ngữ nghĩa, phương pháp này đã được đề xuất bởi tác giả Belongie, ưu điểm của phương pháp này là nó khá tối ưu, đơn giản nhưng hiệu quả mang lại chưa cao cho việc liên quan đến biến đổi hình học và tra cứu dựa trên hình dạng.

Phương pháp thứ hai được nhắc tới đó là phương pháp quy hoạch động (Dynamic Program), phương pháp này dùng để đối sánh các Shape với nhau, các Shape sẽ được coi như dãy của các đoạn lồi lõm được tính đoán dựa trên mô tả Fourier và các momen, ưu điểm của phương pháp này là cho độ chính xác cao, các thuật toán tối ưu, nhược điểm của phương pháp này là tốc độ chậm.

Trong tìm kiếm ảnh theo nội dung, hình dạng là một đặc trưng 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. Các hệ thống tìm kiếm ảnh theo nội dung thường khai thác hai nhóm biểu diễn hình dạng sau :

- Biểu diễn hình dạng theo đường biên (cotour-based descriptor): Biểu diễn các đường biên bao bên ngoài
- Biểu diễn theo vùng (region-based descriptor): Biểu diễn một vùng toàn vẹn

1.3.1.6 Đặc trưng dựa trên cục bộ bất biến

Các nhà nghiên cứu thường chia đặc trưng cục bộ thành hai loại là: những điểm trích xuất được từ điểm "nhô ra" (salient points) của ảnh và đặc trưng SIFT được trích chọn từ các điểm hấp dẫn Harris (interest points) sử dụng truyền thống trong đối sánh ảnh cũng như tra cứu ảnh. Tầm quan trọng của mỗi điểm đặc biệt nằm trong những vùng quan trọng của ảnh, ảnh hưởng đến

việc hiệu quả trong lưu trữ và phân biệt ảnh, nó đặc biệt quan trọng trong phương pháp tra cứu dựa trên đối tượng.

Trong những nghiên cứu mới đây, đã có sự chuyển đổi mô hình từ biểu diễn đặc trưng toàn cục sang mô tả đặc trưng cục bộ, những loại đối tượng và các lớp trực quan sẽ được biểu diễn bởi sự kết của mô tả cục bộ và kết cấu không gian của chúng.

Chúng ta có thể kể đến các nghiên cứu gần đây như: việc sử dụng phương pháp phân đoạn để làm giảm số điểm nhô ra nhằm tăng số đối tượng đại diện đã được Zhang nghiên cứu vào năm 2006. Gouet và cộng sự của mình là Boujemaa thì đã cho ra kết quả nghiên cứu về ưu và nhược điểm của các loại điểm màu thu hút khác nhau. Hay như báo cáo của Mikolajczk và Schmid về hiệu suất của điểm hấp dẫn vào năm 2003.

1.3.2 Độ đo tương tự sử dụng đặc trưng trực quan

Sau khi các ảnh trong cơ sở dữ liệu được trích chọn đặc trưng tiêu biểu bởi các vec tơ đặc trưng nhiều chiều thì sẽ được lưu trong cơ sở dữ liệu đặc trưng. Khi người dùng tra cứu, sẽ cung cấp cho hệ thống ảnh truy vấn, ảnh này cũng được trích chọn đặc trưng, sau đó các đặc trưng của ảnh truy vấn sẽ được đối sánh với các đặc trưng của ảnh trong cơ sở dữ liệu để tìm ra sự tương đồng.

Thông thường có hai cách để đo độ tương đồng

- So sánh Một - Một: Mỗi vùng của ảnh đầu vào chỉ được đối sánh với một vùng của ảnh trong cơ sở dữ liệu. Sự đồng dạng sẽ được định nghĩa giống như là tổng của các đặc điểm giống nhau giữa 2 vùng này của 2 ảnh.
- So sánh Nhiều - Nhiều: mỗi vùng của ảnh đầu vào có thể so sánh với vài vùng của những ảnh khác trong cơ sở dữ liệu. Nhiều giải pháp đã

được đề xuất cho loại so sánh này như là Earth Mover' Distance, Integrated region matching.

1.3.2.1 Một số độ đo khoảng cách

Histogram intersection Distanc (Swain and Ball 1991) :

Đây là một trong những thước đo khoảng cách đầu tiên trong tra cứu ảnh dựa trên màu sắc. Khoảng cách được định nghĩa dựa trên cỡ phần chung của hai biểu đồ màu. Cho hai biểu đồ màu h_1 và h_2 , khoảng cách giữa chúng có thể được định nghĩa bằng công thức:

$$\text{Dist}_{H1} = 1 - \sum N_{i-1} \min(h_{1i}, h_{2i})$$

Việc đo khoảng cách này rất nhanh bởi nó dựa trên công thức đơn giản. Tuy nhiên thông tin màu không được sử dụng khi nhận được khoảng cách bởi vậy nó có thể dẫn tới những kết quả không tốt.

Khoảng cách Minkowski :

Khoảng cách L_1 , khoảng cách dạng Minkowski L_p : khoảng cách dạng Minkowski L_p giữa hai lược đồ được định nghĩa bằng công thức:

$$\text{Dist}_{H1} = 1 - \left(\sum_i |h_{1i} - h_{2i}|^p \right)^{1/p}$$

Earth Mover Distance :

Thước đo này dựa trên chi phí tối thiểu để chuyển một phân bố thành phân bố khác. Nếu chi phí của việc di chuyển một đơn vị đặc điểm đơn trong không gian đặc điểm là khoảng cách chung thì khoảng cách giữa hai phân bố sẽ là tổng cực tiểu của giá trị để di chuyển những đặc điểm riêng. Khoảng cách EMD có thể được định nghĩa bằng công thức sau:

$$\text{Dist}_{\text{EMD}} = \sum_{i,j} g_{ij} d_{ij} / \sum_{i,j} g_{ij}$$

Ở đây, g_{ij} biểu thị khoảng cách tương tự giữa bin i và bin j và $g_{ij} \geq 0$ là sự tối ưu hóa giữa hai phân số như là tổng giá trị được cực tiểu hóa

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

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

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

1.3.3 Phân cụm và phân lớp

Trong tra cứu ảnh, phân lớp được coi là bước tiền xử lý quan trọng nhằm cải thiện tốc độ tra cứu ảnh và độ chính xác trong cơ sở dữ liệu lớn, còn phân cụm giúp cho việc tra cứu ảnh hiệu quả và hiển thị ảnh, tuy nhiên, vấn đề thường gặp phải đó là các cụm không đủ để biểu diễn sự chính xác của phần hiển thị ảnh, trong khi đó, phân lớp có giám sát được xem là một phương pháp rất có hệ thống. Trong những năm đầu nghiên cứu, phân cụm và phân lớp chưa được chú nhiều mà thay vào đó là công nghệ trích chọn đặc trưng và so sánh độ tương tự, cùng với nhu cầu tạo ra một hệ thống có quy mô tốt cho việc xử lý hàng tỉ bức ảnh và hàng triệu người dùng dẫn đến việc nghiên cứu phân cụm và phân lớp ngày càng trở nên quan trọng.

Bảng sau đây mô tả phương pháp phân cụm và phân lớp ảnh:

Phương pháp	Ưu điểm	Công nghệ sử dụng	Hạn chế
Phân cụm	Cho kết quả tốt, trực quan, tốc độ tra cứu nhanh, khả năng tích lũy tốt.	Side-information, kernel mapping, k – means, hierarchical, metric learning	Giống đặc trưng mức thấp, thích nghi người dùng không cao
Phân lớp	Là bước tiền xử lý, kết quả tra cứu chính xác, cấu tạo tự động	SVM, MIL, statistical models, Bayesian classifiers, k -NN, trees	Nhiều lớp vô hình

Bảng 1.2 : Bảng so sánh hai phương pháp phân cụm và phân lớp

Phân cụm không giám sát là phương pháp tự nhiên phù hợp xử lý dữ liệu lớn như dữ liệu ảnh WEB. Bảng 1.3 sẽ tổng hợp các công nghệ phân cụm được cho là nguyên tắc của phân cụm và cũng sẽ chỉ ra những áp dụng của nhiều giải pháp khác nhau khi sử dụng chúng vào phương pháp học theo ví dụ Ta chia ra làm ba trường hợp: các vector, các tập hợp của các vector và các tiến trình ngẫu nhiên .

Các phương pháp phân cụm			
	Dựa trên khoảng cách pair-wise	Tối ưu hóa chất lượng cụm	Mô hình thống kê
Các vector	Cụm kết hợp	k-mean, k-center	Hỗn hợp tham số cơ bản (Gaussian, nonparametric mode based)
Tập hợp các vector		Phân cụm D2	Hỗn hợp thông qua lập bản đồ giả
Tiến trình ngẫu nhiên			Hỗn hợp HMMs

Bảng 1.3: Tổng hợp các phương pháp phân cụm

Trong tài liệu này lấy ví dụ thuật toán phân cụm K-Mean của phương pháp tối ưu hóa chất lượng cụm (Optimize of cluster). Trong thuật toán K - Mean, một vector trọng tâm sẽ được tính toán cho mọi cụm. Vector trọng tâm này được chọn để làm giảm thiểu tổng khoảng cách trong cụm, thuật toán này sử dụng khoảng cách Euclidean, khi khoảng cách Euclidean được sử dụng, nó có thể dễ dàng chỉ ra rằng vector trọng tâm phải là mức trung bình của các vector trong một cụm.

Phương pháp phân lớp phát huy tốt trong trường hợp dữ liệu ảnh được quy định tốt và được dán nhãn, ví dụ như: cơ sở dữ liệu ảnh y tế, ảnh viễn

thám, ảnh văn hóa và nghệ thuật... Phân lớp thường được áp dụng cho chú thích tự động, hoặc tổ chức các hình ảnh vô hình thành các loại rộng nhằm mục đích cho tra cứu ảnh. Phương pháp phân lớp có thể được chia làm hai loại chính là: phương pháp mô hình tách biệt và mô hình thể hệ.

Trong mô hình tách biệt, phân lớp đường biên của lớp được ước tính trực tiếp, ví dụ như kỹ thuật SVM hay kỹ thuật “cây quyết định”. Còn trong mô hình thể hệ, mật độ dữ liệu trong mỗi lớp được ước tính sau có áp dụng công thức Bayes để tính toán về sau. Mô hình tách biệt được sử dụng trực tiếp khi tối ưu hóa phân lớp đường biên, mặt khác mô hình thể hệ lại dễ dàng hơn khi kết hợp với kiến thức trước và có thể được sử dụng thuận tiện khi có nhiều lớp.

1.3.4 Phản hồi liên quan

Phản hồi liên quan là một kỹ thuật sửa đổi truy vấn bắt nguồn trong thông tin tra cứu qua đó sẽ tập hợp lại những đặc trưng tra cứu chính xác từ phía người dùng bằng việc lặp đi lặp lại việc phản hồi, sau đó hệ thống sẽ lọc ra thông tin chính xác. Nó có thể được coi là một mô hình tìm kiếm thay thế, bổ sung cho những mô hình khác như: tìm kiếm dựa trên từ khóa. Trong trường hợp không có một khuôn khổ đáng tin cậy để mô hình hóa ngữ nghĩa ảnh mức cao và nhận thức chủ quan, phản hồi người dùng sẽ cung cấp một cách để tìm hiểu các trường hợp cụ thể của ngữ nghĩa truy vấn. Có thể chia thành sáu loại kỹ thuật phản hồi như sau: *Cải tiến dựa trên học, phản hồi đặc điểm kỹ thuật, phản hồi dựa trên định hướng người dùng, xác suất, dựa trên vùng, các tiến bộ khác.*

1.3.4.1 Kỹ thuật dựa trên “học”

Kỹ thuật này dựa trên thông tin phản hồi có liên quan đến người dùng, phương pháp này thường được sử dụng một cách thích hợp để thay đổi các đặc trưng hoặc trong kỹ thuật so sánh độ tương tự. Tuy nhiên, trong thực tế, kết

quả của phản hồi liên quan người dùng chỉ là một số nhỏ của những ảnh được dán nhãn có liên quan đến khái niệm mức cao. Công nghệ học máy đã được nghiên cứu để giải quyết vấn đề này cũng như những vấn đề đáng quan tâm khác của phản hồi liên quan người dùng. Như là mô hình học một lớp (one - class learning), mô hình học tích cực (Active learning), mô hình học nhiều (manifold learning). Để giải quyết các vấn đề của việc học từ các tập hợp học như vậy, các nhà nghiên cứu đã đề xuất thuật toán phân biệt EM, thuật toán này sử dụng các hình ảnh không có nhãn trong cơ sở dữ liệu cho việc lựa chọn các tính năng phân biệt tốt hơn.

1.3.4.2 Phản hồi đặc điểm kỹ thuật tiến bộ

Theo truyền thống, phản hồi liên quan đã tiếp nhận thông tin từ phía người dùng qua nhiều vòng phản hồi, mỗi vòng gồm một tập hợp các ví dụ tích cực và tiêu cực liên quan đến truy vấn dự định. Tuy nhiên, các nghiên cứu mới đây đã giới thiệu đến các mô hình tiến bộ kỹ thuật khác trực quan hơn và hiệu quả hơn. Thông tin phản hồi trực tiếp dựa trên một ảnh đặc trưng ngữ nghĩa thích hợp được gọi là phản hồi ngữ nghĩa. Một kỹ thuật khác đó là phản hồi chào mời, vấn đề của kỹ thuật này là nó sẽ tạo ra nhiều vòng phản hồi để kiểm tra sự kiên nhẫn của người dùng, để giải quyết vấn đề trên, những log của người dùng đã phản hồi trước đó có thể được sử dụng trong truy vấn sàng lọc, do đó làm giảm lượng người tham gia sử dụng trong phản hồi liên quan, kỹ thuật này đã được Hoi và Lyu nghiên cứu vào năm 2004.

Năm 2003, Kim và Chung đã nghiên cứu kỹ thuật đa truy vấn, nơi mà nhiều ví dụ hình ảnh được sử dụng như là truy vấn và trong các bước trung gian của phản hồi liên quan. Tại mỗi vòng liên quan, các cụm của hình ảnh liên quan được tìm thấy dựa trên tính toán của bước trước đó trong phản hồi liên quan.

1.3.4.3 Phản hồi dựa trên định hướng người dùng

Trước đây, phân lớp, phản hồi liên quan tập trung vào việc học máy dựa vào phản hồi liên quan người dùng, ngày nay đã có một vài nghiên cứu quan tâm đến thiết kế mô hình phản hồi liên quan nhằm hỗ trợ, định hướng người dùng. Trong một vài nghiên cứu mới đây, đã có những nỗ lực trong việc cung cấp cho người dùng những dấu hiệu và gợi ý tìm kiếm để xây dựng truy vấn cụ thể. Một mô hình tìm kiếm tương tự đã được Fang và Geman đề xuất năm 2005, mô hình phản ứng liên tiếp người dùng sử dụng Bayesian, khuôn khổ lý thuyết thông tin. Với mục đích là để “học” một phân phối trên cơ sở dữ liệu ảnh đại diện và sử dụng sự phân phối này để tra cứu.

Một vấn đề khác được quan tâm, đó là việc lặp đi lặp lại các vòng phản hồi liên quan sẽ gây khó chịu cho người dùng, vấn đề này đã được giải quyết phần nào bởi nghiên cứu của Hoi và Lyu năm 2004 bằng cách sử dụng các bản ghi chứa thông tin phản hồi trước đó của người dùng.

1.3.4.4 Phương pháp xác suất

Phương pháp xác suất đã được Cox nghiên cứu năm 2000, các hệ thống PicHunter được đề xuất, nơi mà các mục tiêu không chắc chắn của người dùng được biểu diễn bởi một phân bố trên các mục tiêu tiềm năng, sau đó, hình ảnh đích sẽ được lựa chọn dựa trên luật của Bayesian. Trong nghiên cứu của Su năm 2003, phản hồi liên quan được kết hợp sử dụng một phân lớp Bayesian dựa trên xếp hạng của hình ảnh sau mỗi bước phản hồi. Giả thiết ở đây là, các đặc trưng của ví dụ dương bao gồm cả khả năng cư trú trong lớp ngữ nghĩa là như nhau, tất cả đều được tạo ra bởi một mật độ Gaussian cơ bản. Trong nghiên cứu của Vasconcelos và Lippman [1], phương pháp phản hồi liên quan dựa trên trực giác, độ tin cậy của hệ thống dựa trên ý định của người dùng là “tiền nghiệm”. Trong khi nhưng phản tiếp theo của người dùng là thông tin mới thu thập được, các khái niệm này giúp tính toán độ tin tưởng

mới về mục đích, bằng cách sử dụng luật của Bayesian cho việc trở thành “tiền nghiệm” của vòng phản hồi tiếp theo.

1.3.4.5 Phương pháp dựa trên vùng

Bên cạnh sự phát triển của phương pháp tra cứu ảnh dựa trên vùng, thì ngày nay, người ta đang nỗ lực nghiên cứu để thực hiện kết hợp vùng vào công nghệ phản hồi liên quan. Trong nghiên cứu của Jing [1], xét hai kịch bản phản hồi liên quan, và tra cứu ảnh được phù hợp cho việc hỗ trợ người dùng sửa đổi điểm truy vấn, và phân lớp dựa trên SVM. Trong tiến trình phản hồi liên quan này, vùng quan trọng cho mỗi phân đoạn vùng được “học” để việc tra cứu được tốt hơn. Ý tưởng cốt lõi ở đây, đó là tích hợp việc tra cứu dựa trên vùng cùng với thông tin phản hồi liên quan.

1.4 Ứng dụng của Tra cứu ảnh dựa trên nội dung

1.4.1 Văn hóa nghệ thuật

Văn hóa và nghệ thuật vẫn luôn đóng vai trò quan trọng trong đời sống con người. Trong những thế kỉ qua, hàng trăm những viện bảo tàng cũng như những triển lãm nghệ thuật được xây dựng và tổ chức nhằm gìn giữ những nền văn hóa của chúng ta nhằm góp phần làm nguồn hữu ích cho giáo dục. Tuy nhiên, thế hệ ngày này trải nghiệm những thứ thuộc về lịch sử, văn hóa đó hầu hết là trên các thiết bị số. Ứng dụng của kĩ thuật tra cứu ảnh dựa trên nội dung sẽ giúp chúng ta bảo tồn và phân tích lịch sử của chúng ta trong phương tiện kĩ thuật số, góp phần làm sinh động, trực quan hơn những ví dụ trong văn hóa nghệ thuật, giúp người xem dễ dàng nắm bắt được vấn đề một cách rõ ràng.

1.4.2 Truyện tranh

Trong khi vấn đề mối liên hệ giữa hình ảnh và từ ngữ được nghiên cứu khá tốt, thì việc liên hệ giữa ảnh và một câu truyện lại là vấn đề khá mới mẻ.

Ứng dụng này có thể minh họa khá cụ thể vấn đề được đưa ra dưới dạng hình ảnh, hoặc một bài báo, hoặc các câu chuyện trong sách giáo khoa...

Tuy nhiên vấn đề ở đây là mọi người có thể đính kèm những mức độ quan trọng khác nhau của những ý tưởng, khái niệm, và những nơi thảo luận trong câu truyện. Bất kì một hệ thống minh họa nào đều bị hạn chế bởi kho lưu trữ hình ảnh từ các hệ thống lựa chọn hình ảnh. Hệ thống thực tế đòi hỏi việc xác định các từ khóa có liên quan đến câu chuyện, và các hình ảnh phải được sắp xếp. Những vấn đề mới phát sinh gần đây như là việc bảo vệ quyền tác, việc chứng minh sự tương tác của con người với hệ thống. Bên cạnh việc phát triển những chương trình có trí tuệ nhân tạo cao thì việc sẽ sinh ra những chương trình có thể bắt chước được hành vi con người kèm theo đó là rủi ro an ninh cao đối với những chương trình này

1.4.3 Bảo mật và hình ảnh

Mối liên quan giữa CBIR (Tra cứu ảnh dựa trên nội dung) chưa được quan tâm nhiều cho đến thời gian gần đây. Những vấn đề mới phát sinh gần đây như là việc bảo vệ quyền tác, việc chứng minh sự tương tác của con người với hệ thống.

Bên cạnh việc phát triển những chương trình có trí tuệ nhân tạo cao thì việc sẽ sinh ra những chương trình có thể bắt chước được hành vi con người kèm theo đó là rủi ro an ninh cao đối với những chương trình này.

Các chương trình tấn công website nhằm chiếm tài nguyên của băng thông, đánh cắp thông tin tài khoản người dùng,... CAPTCHA là một giải pháp để giải quyết những vấn đề này, đây là giao diện để phân biệt giữa người hay là máy đang truy cập vào hệ thống website. Những văn bản sẽ được chỉnh sửa méo mó để người dùng có thể nhập chính xác vào trước khi truy cập vào

một website nào đó. Hiện nay, có nhiều website để áp dụng phương pháp này, Google, Yahoo, Bing,...

1.5 Những hướng phát triển hệ thống trong tương lai

Việc xây dựng hệ thống thế giới thực đòi hỏi phải bao hàm tất cả ý kiến phản hồi người dùng trong suốt quá trình thực thi giống như quá trình vòng đời một phần mềm.

- **Về độ hiệu quả:** Vấn đề đáng nói nhất được đưa ra là chất lượng của tra cứu và nó được khảo sát đánh giá như thế nào trong cộng đồng người sử dụng. Một trong những kết cách làm hiện nay được tập trung chứng minh hiệu quả thông qua sự liên kết của số phần trăm chính xác và phản hồi.
- **Học ngữ nghĩa:** Để xử lý vấn đề thiếu sót trong giao diện ngữ nghĩa của hệ thống CIBR, phương pháp học ngữ nghĩa ảnh từ những cơ sở dữ liệu đào tạo và phát triển tra cứu kỹ xảo.
- **Khối dữ liệu:** dữ liệu ảnh sẽ ngày một phát triển, hệ thống phần mềm phải có khả năng xử lý, lưu trữ và tra cứu một cách thông minh.
- **Giao diện người dùng:** Một kết quả đạt được tốt hơn là cần được thiết kế giao diện trực quan cho hệ thống tra cứu có như vậy, người dùng mới thực sự được sử dụng một công cụ cho lợi ích của họ
- **Tốc độ hoạt động:** thời gian xử lý online và thời gian hồi đáp trả lời người dùng cần được đáp ứng tốt để tăng tính hiệu quả. Các phần tử tính toán nên được sử dụng các thuật toán phù hợp và có hiệu năng cao nhất, đặc biệt là cho các hệ thống lớn.

CHƯƠNG 2: TRA CỨU ẢNH DỰA TRÊN ĐẶC TRƯNG HÌNH DẠNG**2.1 Giới thiệu**

Không như kết cấu, hình dạng là một khái niệm hoàn toàn rõ ràng, bằng chứng là những vật thể đầu tiên được nhận thấy bởi hình dạng của chúng. Số lượng những đặc điểm tiêu biểu của hình dạng đối tượng được tính toán cho mỗi đối tượng xác định trong mỗi ảnh được lưu trữ. Sau đó truy vấn được trả lời bởi việc tính toán tập những đặc điểm cho ảnh truy vấn, và việc tra cứu đặc điểm của những hình ảnh được lưu trữ này phải phù hợp với đặc điểm của truy vấn. Hai kiểu chính của đặc điểm hình dạng thường được sử dụng là đặc điểm tổng thể như tỷ lệ bên ngoài, hình tròn và những đặc điểm cục bộ như tập các đoạn biên liên tiếp. Các phương pháp khác đề cập tới sự đối sánh hình dạng bao gồm sự biến dạng co giãn của các khuôn dạng, sự so sánh của những biểu đồ định hướng của những biên được trích chọn từ ảnh, khung biểu diễn hình dạng của đối tượng có thể được so sánh bằng việc sử dụng những kỹ thuật đối sánh đồ thị. Những truy vấn đối với hệ thống tra cứu hình dạng thường được biểu diễn bằng cách xác định một hình ảnh mẫu để thực hiện như là hình thức truy vấn hoặc như là một bản phác thảo được vẽ ra bởi người sử dụng. Hình dạng có khuynh hướng chỉ đến một khu vực đặc biệt trong ảnh, hay hình dạng chỉ là biên của một đối tượng nào đó trong ảnh.

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.

Trong tìm kiếm ảnh theo nội dung, hình dạng là một đặc điểm cao cấp 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. Các hệ thống tìm kiếm ảnh theo nội dung thường khai thác hai nhóm biểu diễn:

- Biểu diễn hình dạng theo đường biên: là biểu diễn các đường biên bao quanh bên ngoài ảnh
- Biểu diễn hình dạng theo vùng: Biểu diễn một vùng toàn vẹn

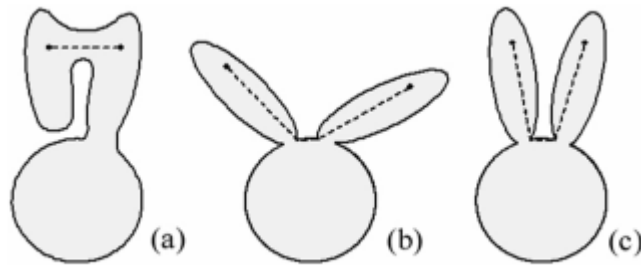
Trong nội dung của chương này, sẽ tập trung đi sâu vào khai thác phương pháp trích chọn đặc trưng IDSC (Inner Distance Shape Context) dựa theo biểu diễn hình dạng theo đường biên.

2.2 Trích chọn đặc trưng IDSC

2.2.1 Giới thiệu

Cấu trúc thành phần đóng vai trò quan trọng trong việc phân loại những hình dạng phức tạp. Tuy nhiên, việc thu lại được những cấu trúc thành phần chưa bao giờ là một công việc đơn giản, nhất là khi xét đến cấu trúc hình dạng có khớp nối. Những kiểu hình dạng này là sự biến đổi phi tuyến giữa các hình dạng, hơn nữa, một vài hình dạng có thể có cấu trúc “nhập nhằng”. Để giải quyết cho những vấn đề này, Haibin Ling [2] đã đề xuất ra một kỹ thuật biểu diễn hình dạng được gọi là khoảng cách trong.

Khoảng cách trong được định nghĩa là khoảng cách ngắn nhất của đường dẫn bên trong đường biên hình dạng nhằm xây dựng sự nhận diện hình dạng ảnh. Có thể dễ dàng thấy được, khoảng cách trong không nhạy cảm với các hình dạng khớp nối. Ví dụ trong hình 2.1



Hình 2.1: Ví dụ về khoảng cách trong của đối tượng

Ta có thể thấy, mặc dù trong hình (a) và hình (c) đều có sự phân bố không gian tương tự nhau, nhưng chúng lại hoàn toàn khác nhau về cấu trúc thành phần của chúng. Mặt khác, hình (c) và hình (b) lại xuất hiện từ cùng một loại hình dạng chỉ khác nhau ở các khớp nối. Khoảng cách trong giữa hai điểm được đánh dấu trong hình (a) và hình (b) là hoàn toàn khác nhau trong khi, phần lớn sự giống nhau lại nằm ở hình (b) và hình (c). Bằng trực giác, ví dụ này cho ta thấy rằng, khoảng cách trong là không nhạy cảm đối với cấu trúc khớp nối, và nhạy cảm đối với cấu trúc thành phần, một thuộc tính đáng đề hướng tới cho việc đối sánh các hình dạng phức tạp. Trong khi đó khoảng cách Euclidean không có những thuộc tính đó đối với ví dụ trên. Bằng chứng cho vấn đề này chính là khoảng cách trong được định nghĩa như là độ dài của những đoạn nét đứt giữa các điểm được đánh dấu, còn khoảng cách Euclidean thì không xem xét đến có những đoạn nét đứt chồng chéo lên nhau.

Việc sử dụng khoảng cách trong như là một giải pháp để thay thế cho những độ đo tương tự khác nhằm xây dựng một mô tả hình dạng mới mà có khả năng bất biến (không nhạy cảm) đối với hình dạng có cấu trúc khớp nối.

2.2.2 Ngữ cảnh hình dạng (Shapes Context)

Ngữ cảnh hình dạng được giới thiệu bởi Belongie [5]. Nó mô tả phân bố không gian liên quan của các điểm đã được đánh dấu xung quanh những điểm đặc trưng: cho n điểm mẫu $x_1, x_2, \dots, x_n$ trên một hình dạng. Ngữ cảnh hình dạng tại điểm x_i được định nghĩa như là biểu đồ tần suất h_i tọa độ liên quan của $n - 1$ điểm còn lại.

$$\text{Ta có công thức: } h_i(k) = \#\{x_j : j \neq i, x_j - x_i \in \text{bin}(k)\} \quad (1)$$

Trong đó: các bin được phân bố đều nhau trong không gian log-polar. Khoảng cách giữa hai biểu đồ ngữ cảnh hình dạng được định nghĩa bằng cách sử dụng thống kê χ^2 .

Để đối sánh hình dạng, Belongie đã sử dụng một framework kết hợp ngữ cảnh hình dạng và *thin-plate-splines*. Cho các điểm trên hai hình A và B, trước tiên các điểm phù hợp sẽ được tìm thấy thông qua đối sánh đồ thị vô hướng có trọng số, sau đó *thin-plate-splines* được sử dụng một cách lặp đi lặp lại để ước lượng sự biến đổi giữa chúng. Tiếp đó, độ tương tự D giữa A và B được đo bằng sự kết hợp của ba phần:

$$D = aD_{ac} + D_{sc} + bD_{be} \quad (2)$$

Trong đó:

D_{ac} : là độ đo sự khác biệt.

D_{be} : là độ đo khả năng uốn.

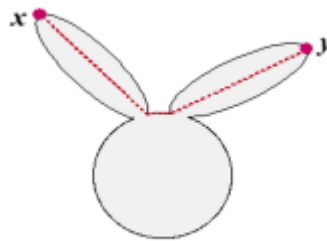
D_{sc} : là độ đo khoảng cách ngữ nghĩa, là độ đo khoảng cách trung bình giữa điểm trên A và những điểm tương tự nhất tương ứng trên B.

Ngữ cảnh hình dạng sử dụng khoảng cách Euclidean để đo không gian liên hệ giữa các điểm được đánh dấu. Khoảng cách trong là cách phù hợp để giải quyết vấn đề hình dạng có khớp nối do nó thu được những hình dạng cầu

trúc tốt hơn khoảng cách Euclidean. Khoảng cách trong ứng dụng mở rộng trong việc đối sánh hình dạng, ưu điểm của phương pháp này được thể hiện qua các thí nghiệm .

2.2.3 Khoảng cách trong (THE INNER DISTANCE)

Trước tiên, cho hình O là một tập đóng và có kết nối của $\mathbb{R}^2$, hai điểm x và y thuộc O , khoảng cách trong giữa x và y được ký hiệu là: $d(x, y; O)$ và được định nghĩa là độ dài của đường dẫn ngắn nhất kết nối hai điểm x và y ở trong hình O . Ví dụ hình 2.2



Hình 2.2: Ví dụ về khoảng cách trong của x và y trong hình O

Vấn đề đặt ra:

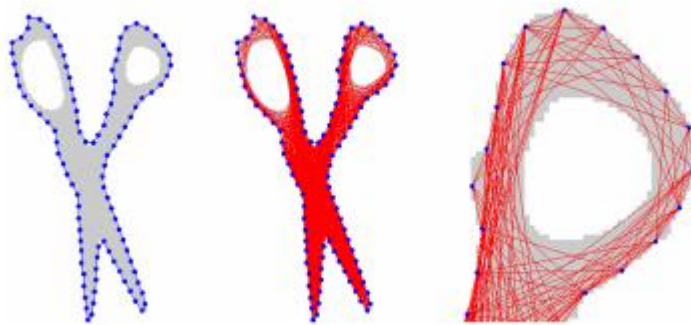
- Trong một vài trường hợp hiếm gặp, có thể tồn tại nhiều đường dẫn ngắn nhất giữa các điểm cho trước, khi đó, ta tùy ý chọn một đường dẫn ngắn nhất trong số đó.
- Chúng ta đã quen với việc định nghĩa Shapes bởi những đường biên của chúng, do đó, chỉ những điểm biên được sử dụng như là những điểm đánh dấu. Hơn nữa hình dạng được xấp xỉ bởi một hình đa giác, đa giác này được hình thành nên bởi những điểm được đánh dấu của chúng.

Cách đơn giản nhất để tính toán khoảng cách trong là sử dụng thuật toán tìm đường dẫn ngắn nhất, thuật toán này được chia là hai bước:

- Bước một: Xây dựng một đồ thị với các điểm mẫu. Trước tiên, mỗi điểm mẫu được coi như là một nút ở trong đồ thị, sau đó đối với mỗi cặp điểm mẫu p_1 và p_2 , nếu đoạn nối liền p_1 và p_2 nằm hoàn toàn trong đối tượng thì một cạnh giữa p_1 và p_2 được thêm vào đồ thị cùng với trọng số của nó là khoảng cách Euclidean $\|p_1 - p_2\|$. Ví dụ: hình 2.3

Một vài chú ý được đề cập tới đó là :

- Thứ nhất: các điểm biên láng giềng thì luôn luôn liên thông với nhau.
- Thứ hai: Khoảng cách trong không sử dụng những điểm mẫu của đường biên lỗ hổng.



Hình 2.3: Quá trình biểu diễn khoảng cách trong của đối tượng

- Bước thứ hai: Áp dụng thuật toán tìm đường đi ngắn nhất cho đồ thị. Nhiều thuật toán đã được áp dụng, trong đó có thuật toán Floyd-Warshall's có độ phức tạp là $O(n^3)$ với n là số điểm lấy mẫu. Thuật toán khoảng cách trong đã được tác giả chỉ ra có độ phức tạp thuật toán là $O(n^3)$. Trước tiên, mất một khoảng thời gian $O(n)$ để kiểm tra xem đoạn nối giữa hai điểm nằm trong hình dạng. Tiếp theo, việc xây dựng đồ thị có độ phức tạp là $O(n^3)$. Khi đồ thị đã được tính toán xong, thuật toán

dùng để tính toán tất cả các cặp có đường dẫn ngắn nhất có độ phức tạp thuật toán là $O(n^3)$. Do vậy, độ phức tạp tính toán toàn bộ là $O(n^3)$.

2.2.4 Inner-Distance Shape Context

Để mở rộng định nghĩa về ngữ cảnh hình dạng đã được trong công thức (1). Ling và Jacobs [4] đã định nghĩa lại các bin với khoảng cách trong, khoảng cách Euclidean được thay thế trực tiếp bằng khoảng cách trong. Sự định hướng liên quan giữa hai điểm có thể được định nghĩa như là phương tiếp tuyến tại điểm bắt đầu của đường dẫn ngắn nhất giữa chúng. Tuy nhiên, phương tiếp tuyến này bị nhạy cảm đối với các khớp nối.

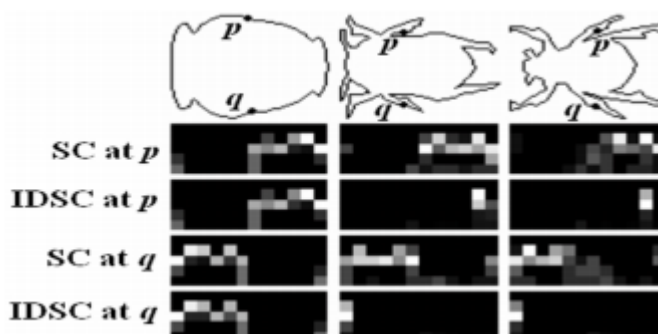
Thực tế, với điểm biên p và đường dẫn ngắn nhất $P(p, q, O)$ giữa điểm biên p và điểm q trong hình O thì góc được tạo bởi giữa tiếp tuyến tại q và hướng của $P(p, q, O)$ tại p là không nhạy cảm với khớp nối. Ta gọi góc này là góc trong (inner - angle) và kí hiệu nó là $\Theta(p, q, O)$ như hình 2.4. Góc trong này được sử dụng cho các bin hướng. Tuy nhiên, trong thực tế, đường biên hình dạng có thể bị bóp méo dẫn tới việc làm giảm sự ổn định của các góc trong. Để giải quyết vấn đề này, đường bao sẽ được làm mịn bằng việc sử dụng các “láng giềng” nhỏ trước khi tính toán góc trong.



Hình 2.4: Minh họa về góc trong (Inner - Angle)

Hình 2.5 là ví dụ về việc tính toán ngữ cảnh hình dạng bởi hai cách khác nhau. Có thể thấy rõ, ngữ cảnh hình dạng là giống nhau đối với cả ba hình, trong khi khoảng cách trong chỉ giống nhau đối với hai hình bên phải (hình

hai con bọ bên phải). Từ đó ta có thể thấy, khoảng cách trong tốt hơn so với ngữ cảnh hình dạng trong việc thu được các phần của hình dạng.



Hình 2.5: Ngữ cảnh hình dạng (SC) và khoảng cách trong ngữ cảnh hình dạng (IDSC)

2.3 Đối sánh shape sử dụng quy hoạch động

Bài toán đối sánh đường bao được phát biểu như sau: cho hai hình A và hình B, ta mô tả chúng bằng các dãy điểm trên đường bao của chúng. Ta có: $p_1, p_2, \dots, p_n$ là n điểm thuộc hình A và m điểm $q_1, q_2, \dots, q_m$ thuộc hình B. Giả sử $n \geq m$, sự đối sánh π từ A đến B là một ánh xạ từ $1, 2, \dots, n$ đến $0, 1, 2, \dots, m$ trong đó p_i được đối sánh với $q_{\pi(i)}$ nếu $\pi(i)$ khác 0 và ngược lại thì không đối sánh. π nên được cực tiểu hóa chi phí đối sánh và được định nghĩa là $H(\pi)$:

$$C(\pi) = \sum_{1 \leq i \leq n} c(i, \pi(i)) \quad (3)$$

Trong đó $c(i, 0) = \tau$ là hình phạt cho việc bỏ qua p_i không đối sánh, và cho $1 \leq j \leq m$, $c(i, j)$ là chi phí của đối sánh p_i với q_j . Điều này được đo bằng cách sử dụng hàm thống kê χ_2 như công thức sau:

$$c(i, j) = \frac{1}{2} \sum_{1 \leq k \leq K} \frac{[hA, i(k) - hB, j(k)]^2}{hA, i(k) + hB, j(k)} \quad (4)$$

Trong đó, $h_{A,i}$ và $h_{B,j}$ ở đây là những biểu đồ ngữ cảnh hình dạng của riêng p_i và q_j , và K là số biểu đồ các bin.

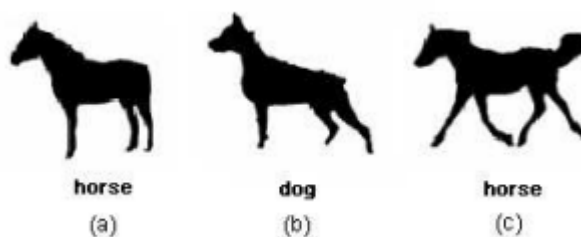
Do các đường viền cung cấp các thứ tự cho các chuỗi điểm $p_1p_2\dots p_n$ và $q_1q_2\dots q_m$, nên nó vô tình làm hạn chế sự đối sánh π cùng với sự sắp xếp này. Giải pháp sử dụng Quy hoạch động (DP) đã được đưa ra để giải quyết cho vấn đề đối sánh này. DP được sử dụng rộng rãi cho vấn đề đối sánh đường bao.

Mặc định, các phương pháp ở trên giả định hai đường bao đã được căn chỉnh tại điểm đầu và điểm cuối của chúng. Nếu không có giả định này, thì một giải pháp đơn giản là thử căn chỉnh tại tất cả các điểm trên đường bao đầu tiên và chọn ra một cái tốt nhất. Tuy nhiên, vấn đề của giải pháp này đó là nó làm tăng độ phức tạp đối sánh từ $O(n^2)$ đến $O(n^3)$. Với bài toán so sánh, nó thường đủ để có khả năng cho việc căn chỉnh các điểm đã được cố định, ở đây là k và thường thì k sẽ nhỏ hơn nhiều so với m và n , ví dụ: có $n, m = 100$, $k = 4$ hay 8 đều đủ thỏa mãn, với k lớn hơn thì cũng không chứng minh sự cải thiện đáng kể. Do đó, độ phức tạp vẫn là $O(kn^2) = O(n^2)$.

2.4 Tra cứu ảnh với kĩ thuật học hàm khoảng cách

Tra cứu hình dạng là một vấn đề rất quan trọng trong thị giác máy. Đã có nhiều phương pháp tra cứu hình dạng được đề xuất và nghiên cứu, đặc biệt là trong những năm trở lại đây. Tuy nhiên, tất cả các phương pháp đều tập trung vào bản chất tương tự của hình dạng. Nó có vẻ như là một lời tuyên bố rằng, hai hình dạng giống nhau hơn khi mà sự khác biệt của chúng là nhỏ hơn, sự khác biệt này được đo bằng hàm khoảng cách. Mặc dù vậy, tuyên bố này đã bỏ qua thực tế, đó là một vài sự khác biệt có liên quan, trong khi một vài sự khác biệt khác thì lại không phù hợp đối với độ tương tự của hình dạng. Tra cứu hình dạng rõ ràng có mối liên quan chặt chẽ đến việc hiểu được hình dạng

ở mức cao, đặc trưng hình dạng ở trong cùng một lớp có thể khác biệt đáng kể bởi sự biến dạng hoặc biến đổi mềm. Nói cách khác là kể cả hai hình có nằm trong cùng một lớp thì khoảng cách về sự khác nhau của chúng cũng là rất lớn nếu như độ đo khoảng cách không nắm bắt được tính chất bên trong của các hình này. Nó cho thấy rằng, tất cả những độ đo khoảng cách đã được công bố đều chưa giải quyết được vấn đề này. Ví dụ, dựa trên IDSC trong hình 2.6, ta thấy hình (a) giống với (b) hơn là đối với (c), nhưng rõ ràng là hình (a) và hình (c) là thuộc cùng một lớp (con ngựa). Kết quả không chính xác này là do trong thực tế, khoảng cách trong đã không nhận biết được sự mất tích của cái đuôi và một chân trước là không thích hợp cho sự quyết định độ tương tự giữa hai hình dạng này. Mặt khác, có nhiều những chi tiết nhỏ hơn giống tai của con chó (hình b) và hình dạng của cái đầu là những thứ có liên quan cao hơn ở đây. Không cần biết là thuật toán đối sánh hình dạng tốt đến mức nào, vấn đề về sự khác nhau liên quan và không liên quan của hình dạng phải được giải quyết nếu muốn các bộ máy có hiệu suất giống như con người. Điều này đòi hỏi phải có một mô hình nắm bắt bản chất của một số lớp hình dạng thay vì xem mỗi hình dạng như là một tập hợp các điểm hoặc tham số hàm.



Hình 2.6 : Ví dụ về khoảng cách trong ngữ cảnh hình dạng

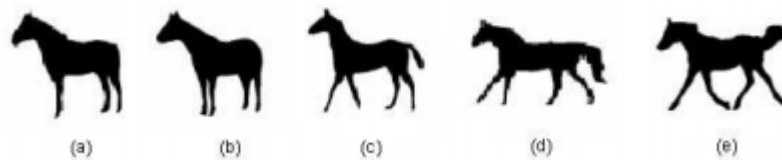
Để giải quyết vấn đề này, một phương pháp đã được đề xuất, đó là kỹ thuật học hàm khoảng cách. Thay vì tập trung vào việc tính toán độ tương tự

của các cặp hình dạng, thì ta tận dụng những lợi thế của sự đa dạng được hình thành bởi các hình dạng đã có sẵn.

Một cách đánh giá tốt hơn được học bằng cách truyền lại những độ đo tương tự chung tới truy vấn hình dạng và giữa các hình dạng có sẵn thông qua biểu đồ truyền dẫn.

Cho một cơ sở dữ liệu hình dạng, một truy vấn hình dạng, và một hàm khoảng cách hình dạng mà không cần tới một thước đo, ta sẽ học một hàm khoảng cách mới, nó được thể hiện bởi đường đi ngắn nhất trên hình đa dạng được cấu thành bởi những hình đã biết và truy vấn hình dạng.

Hàm khoảng cách mới được học này có thể dùng để kết hợp các kiến thức về các hình dạng liên quan và không liên quan khác nhau. Trong nghiên cứu của Xingwei Yang [3] chứng minh có thể làm được điều này mà không cần học chính xác hình đa dạng. Hàm khoảng cách này được học trong môi trường thiết lập không giám sát trong ngữ cảnh của hình dạng đã biết. Chúng ta có thể làm rõ hơn trong ví dụ sau :



Hình 2.7: Các hình đã biết trong cơ sở dữ liệu.

Trong một cơ sở dữ liệu những hình đã biết như hình 2.7 bao gồm các hình từ hình (a) cho tới hình (e), khi đó hàm khoảng cách mới đã được học như trình bày ở trên kia sẽ xếp hạng chính xác hình dạng ở trong hình 2.6, hình (a) sẽ giống hình (c) hơn là giống hình (b). Lý do là các chức năng của khoảng cách mới được học sẽ thay thế các chức năng của khoảng cách ban

đầu từ (a) tới (c) trong hình 2.6 cùng với khoảng cách bao gồm đường dẫn ngắn nhất giữa (a) và (e) trong hình 2.7

Về tổng quát, cho hai hình A và C, nếu như sự khác biệt (khoảng cách) giữa A và C là quá lớn nhưng tồn tại một hình B mà nó có sự khác biệt nhỏ đối với A và đối với C thì ta coi A và C là giống nhau. Tình trạng này xảy ra đối với hầu hết các khoảng cách hình dạng, bởi vì chúng không tuân theo bất đẳng thức tam giác: $d(A, C) \leq d(A, B) + d(B, C)$ cho tất cả các hình A, B, C. Do vậy, giải pháp được đưa ra đó là học hàm khoảng cách nhằm thay đổi khoảng cách ban đầu $d(A, C)$.

Nếu gặp trường hợp $d(A, C) > d(A, B) + d(B, C)$, thì ta sẽ tiến hành học một khoảng cách mới $d'(A, C)$ sao cho $d'(A, C) \leq d(A, B) + d(B, C)$, hơn nữa nếu tồn tại một đường dẫn ở trong không gian khoảng cách mà: $d(A, C) > d(A, B_k) + \dots + d(B_k, C)$ thì khi đó khoảng cách mới d' sẽ như sau: $d'(A, C) \leq d(A, B_k) + \dots + d(B_k, C)$. Lúc này đường dẫn đó sẽ tương ứng với biến dạng tối thiểu Morphing từ hình A đến hình B. Trong trường hợp này ta bỏ qua những hình dạng khác nhau không phù hợp, mà chỉ tập trung vào những hình dạng khác nhau phù hợp với khoảng cách mới này. Đó là những nét cơ bản về kỹ thuật học hàm khoảng cách, kỹ thuật này sẽ được trình bày kỹ hơn trong phần tiếp theo của đề án.

2.4.1 Học hàm khoảng cách (Learning New Distance Measures)

Trước tiên, Xingwei Yang [3] miêu tả những thiết lập của tra cứu độ tương tự, nó áp dụng cho nhiều kiểu tra cứu, như hình ảnh, văn bản, từ khóa, và hình dạng. Ví dụ, cho một tập hợp các đối tượng $X = \{x_1, \dots, x_n\}$ và một hàm độ tương tự sim: $X \times X \rightarrow R^+$, hàm này được gán một giá trị tương ứng (số nguyên dương) cho mỗi cặp các đối tượng. Coi x_1 là một đối tượng truy vấn (ví dụ: truy vấn hình dạng), $\{x_2, \dots, x_n\}$ là một tập hợp của cơ sở dữ liệu

các đối tượng đã biết (hoặc một tập hợp đào tạo) sau đó, sắp xếp các giá trị $\text{sim}(x_1, x_i)$ theo thứ tự giảm dần $i = 2, \dots, n$, vậy là đạt được sự xếp hạng của cơ sở dữ liệu các đối tượng theo độ tương tự của chúng cho truy vấn. Đối tượng giống với ảnh truy vấn nhất sẽ có giá trị cao nhất và được hiển thị lên đầu tiên. Đôi khi một độ đo khoảng cách (Sự khác nhau) được sử dụng để thay thế cho độ đo tương tự (sự giống nhau), trong trường hợp này, việc xếp hạng, sắp xếp các đối tượng trong cơ sở dữ liệu theo chiều tăng có nghĩa là, đối tượng nào trong cơ sở dữ liệu có độ khác nhau nhỏ nhất thì sẽ có giá trị nhỏ nhất và sẽ được hiển thị lên trước.

Một vấn đề gặp phải ở đây đó là hàm sim không hoàn toàn đúng cho nhiều cặp của những đối tượng, nó có thể trả về kết quả sai, mặc dù nó trả về kết quả đúng cho hầu hết các cặp. Một phương án để giải quyết cho vấn đề này là: học hàm tương tự mới gọi là sim_T , nó giúp cải thiện kết quả tra cứu của sim cho truy vấn x_1 .

Cho $w_{i,j} = \text{sim}(x_i, x_j)$, với $i, j = 1, \dots, n$ là một ma trận tương tự, mà có tên gọi khác là ma trận quan hệ. Ta khai báo một chuỗi những hàm gán nhãn $f_t : X \rightarrow [0, 1]$ với $f_0(x_1) = 1$ và $f_0(x_i) = 0$, $i = 2, \dots, n$. Ta có hàm đệ quy của f_t như sau:

$$f_t : f_{t+1}(x_1) = \frac{\sum_{j=1}^n w_{ij} f_t(x_j)}{\sum_{j=1}^n w_{ij}} \quad (1)$$

với $i = 2, \dots, n$ và ta đặt:

$$f_{t+1}(x_1) = 1 \quad (2)$$

Lúc này, ta chỉ còn một lớp chỉ chứa một phần tử có nhãn là truy vấn x_1 , khi đó hàm tương tự mới được học thay thế x_1 được định nghĩa như sau:

$$\text{sim}_t(x_1, x_i) = f_t(x_i) \quad (3)$$

Do đó, ta coi f_t như là một tập hợp của các giá trị tương tự đã được chuẩn hóa với truy vấn x_1 . Quan sát thấy, $sim_1(x_1, x_j) = w_{1,j} = sim(x_1, x_j)$.

Bước (1) và (2) được lặp đi lặp lại cho đến khi bước $t = T$. Sau đó, ta xếp hạng độ tương tự cho truy vấn x_1 với sim_T .

Bước (1) và bước (2) được sử dụng trong lan truyền nhãn. Đặc biệt là trong trường hợp chỉ có một lớp đã biết, đó là lớp của các đối tượng truy vấn, có nghĩa là trong thực tế, sự lan truyền nhãn có một cách giải quyết trong trường hợp $\lim_{t \rightarrow \infty} f_t(x_i) = 1$ với $i = 1, \dots, n$, có nghĩa là tất cả các đối tượng sẽ được gán lớp nhãn của truy vấn hình dạng. Do mục đích ở đây là xếp hạng các đối tượng trong cơ sở dữ liệu theo độ tương tự của chúng với truy vấn, nên việc tính toán sẽ dừng lại sau một số bước lặp phù hợp $t = T$. Thông thường, quá trình lặp đi lặp lại này phải đảm bảo được sự hội tụ. Tuy nhiên, việc tính toán được dừng lại khi sự khác biệt $\|f_{t+1} - f_t\|$ trở nên rất nhỏ.

Nếu cơ sở dữ liệu các đối tượng đã được biết trước rất lớn, thì việc tính toán với n các đối tượng có thể trở nên không thực tế. Vì vậy, xây dựng ma trận w chỉ sử dụng M sao cho $M < n$ đối tượng tương tự với truy vấn x_1 được sắp xếp theo hàm khoảng cách ban đầu sim .

CHƯƠNG 3: CÀI ĐẶT CHƯƠNG TRÌNH VÀ ỨNG DỤNG**3.1 Môi trường thực nghiệm**

Chương trình được xây dựng bằng bộ công cụ Matlab 7.7.0 (R2008) sử dụng ngôn ngữ lập trình Matlab. Sử dụng cơ sở dữ liệu ảnh MPEG7 gồm 1400 ảnh(Shape) được chia làm 70 lớp mỗi lớp gồm 20 ảnh. Nhóm các modul chính của chương trình Tra cứu ảnh theo nội dung ứng dụng phương pháp học hàm khoảng cách.

- Trích chọn đặc trưng toàn bộ ảnh trong cơ sở dữ liệu
- Chọn và hiển thị ảnh truy vấn
- Tra cứu ảnh truy vấn, hiển thị kết quả tra cứu.
- Học hàm khoảng cách giúp cải thiện kết quả truy vấn

3.2 Bài toán

Chương trình tra cứu hình dạng ảnh ứng dụng kỹ thuật học hàm khoảng cách nhằm phục vụ cho mục đích tra cứu và hiển thị hình ảnh có độ tương nhất đối với ảnh truy vấn.

- Đầu vào: Ảnh truy vấn (Shape)
- Đầu ra: Danh sách 20 ảnh giống nhất so với ảnh truy vấn đầu vào.

3.3 Các bước chính của chương trình

Bước 1 (Offline): Xây dựng CSDL đặc trưng

- Trích chọn đặc trưng 1400 ảnh trong CSDL bằng phương pháp khoảng cách trong (IDSC)
- Tính ma trận khoảng cách của 1 ảnh so với các ảnh còn lại trong CSDL. Độ tương tự của 2 ảnh được tính toán bằng cách sử dụng phương pháp quy hoạch động (Dynamic Program).

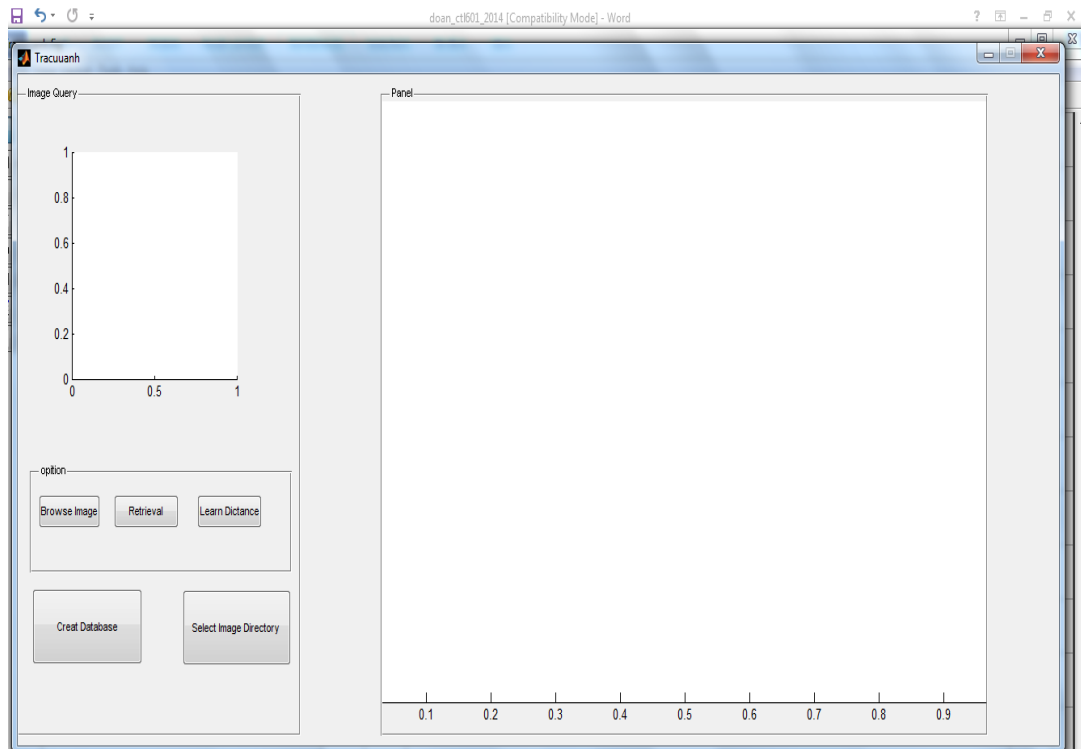
Bước 2 (Online): Tra cứu ảnh

- Chọn ảnh truy vấn và hiển thị ảnh truy vấn
- Trích chọn đặc trưng ảnh truy vấn sử dụng kỹ thuật IDSC
- Tính độ tương tự của ảnh truy vấn đối với các ảnh trong CSDL
- Sắp xếp ảnh kết quả theo độ tương tự tăng dần.
- Hiển thị 20 ảnh có độ tương tự giống nhất so với ảnh truy vấn

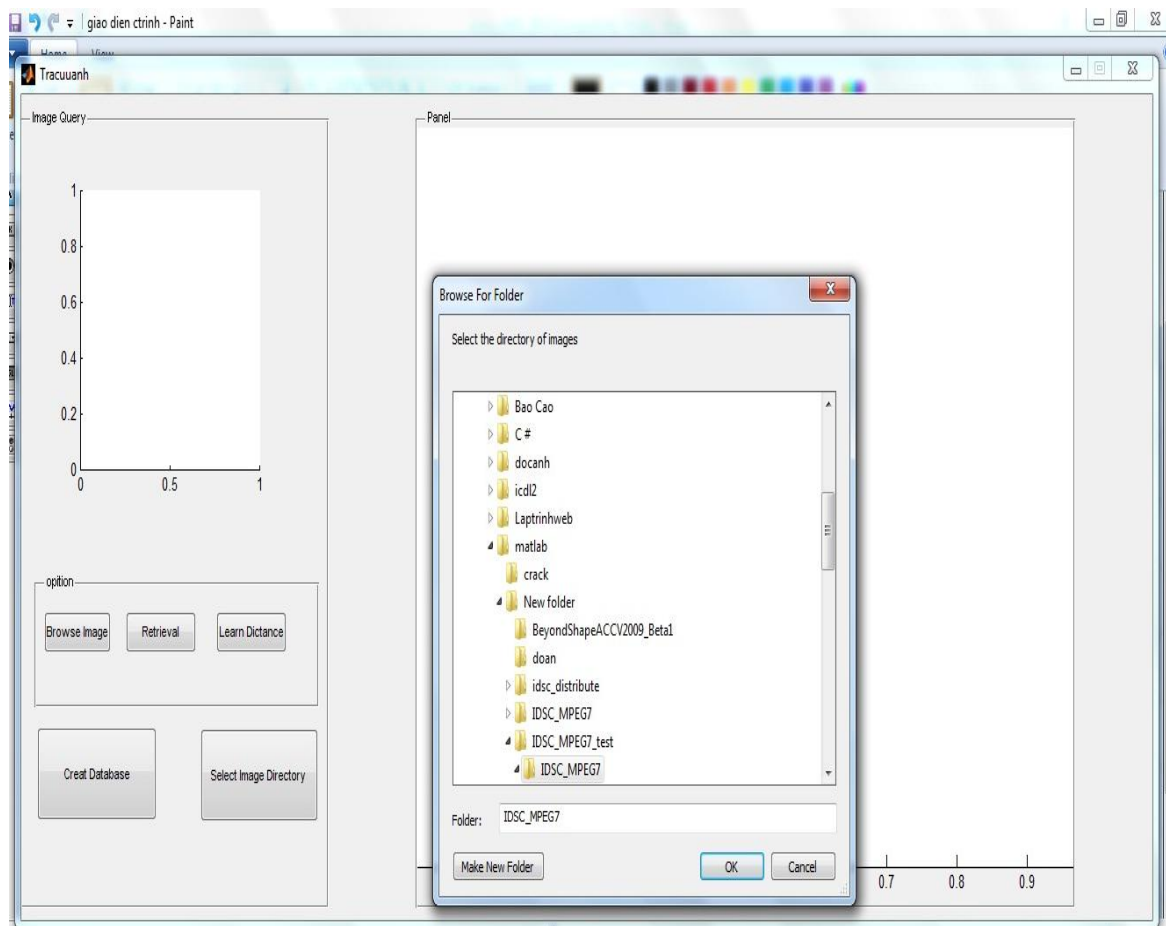
Bước 3: Học hàm khoảng cách

- Xây dựng ma trận khoảng cách của tất cả các ảnh
- Chuẩn hóa ma trận khoảng cách
- Học hàm khoảng cách
- Sắp xếp hàm khoảng cách mới giảm dần
- Lấy ra 20 ảnh đầu tiên để hiển thị

3.4 Một số hình ảnh của chương trình

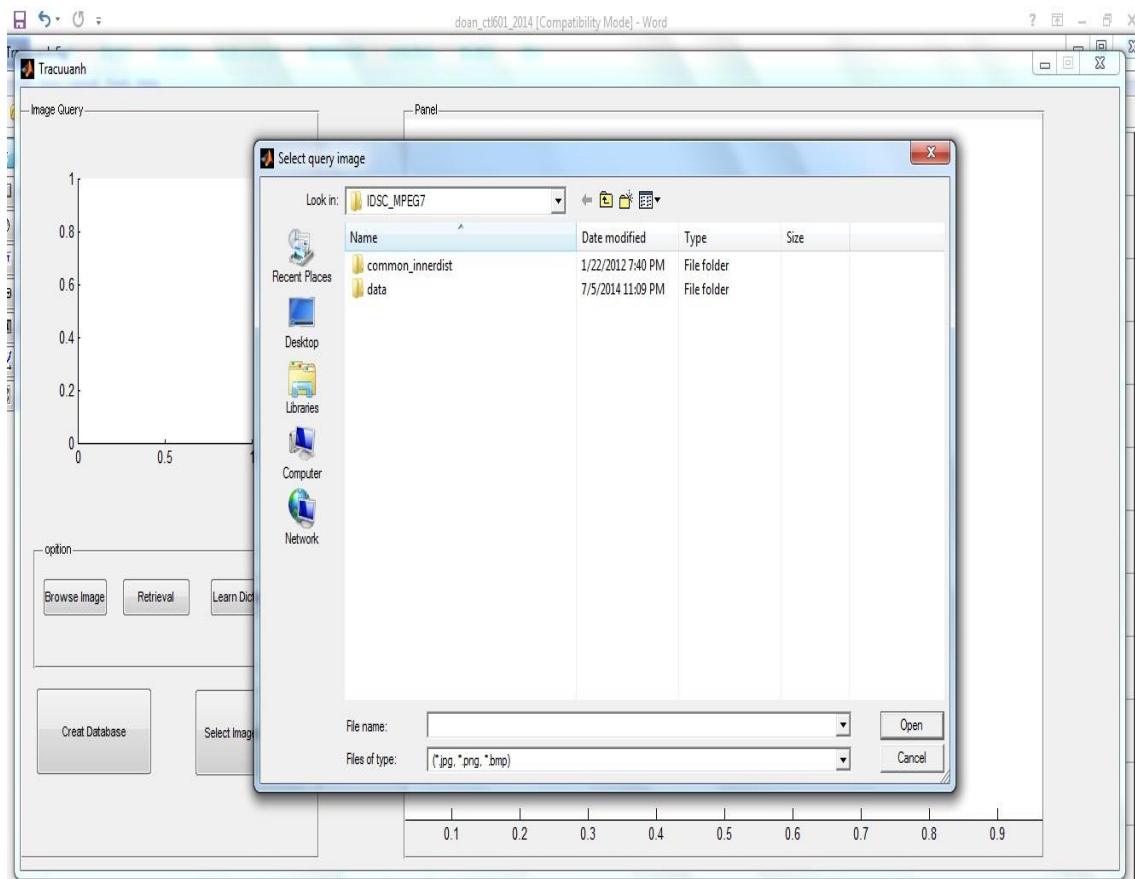


Hình 3.1: Giao diện chính của chương trình



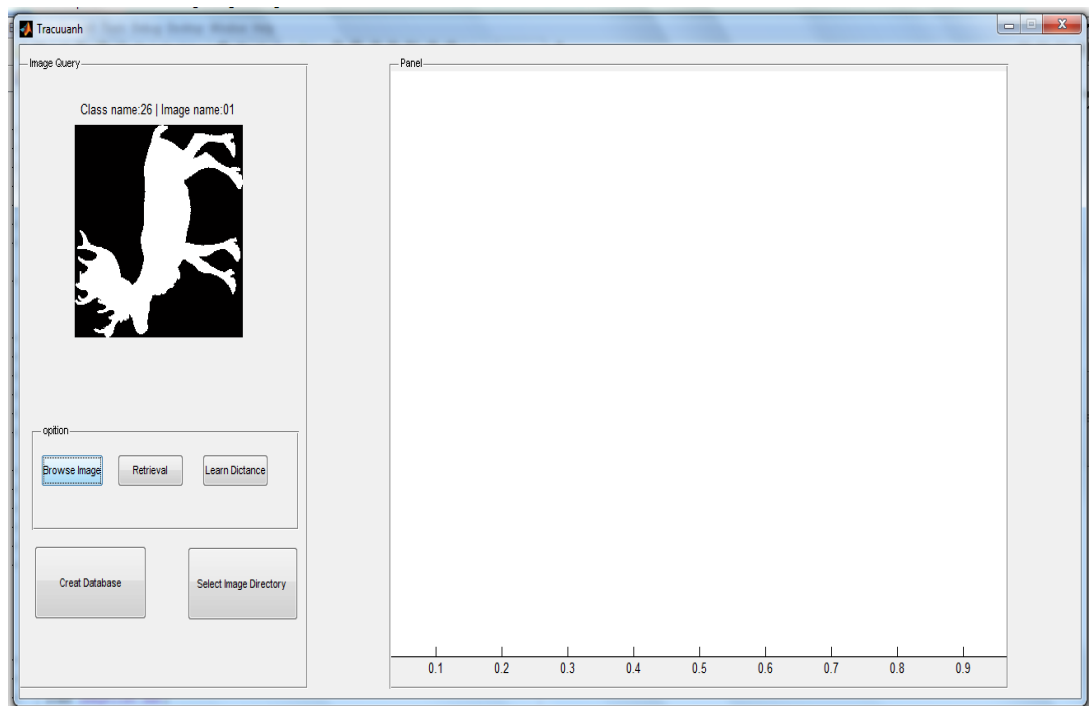
Hình 3.2: Giao diện chọn thư mục chứa CSDL đặc trưng của ảnh

Trong phần này CSDL đặc trưng nằm trong thư mục IDSC_MPEG7

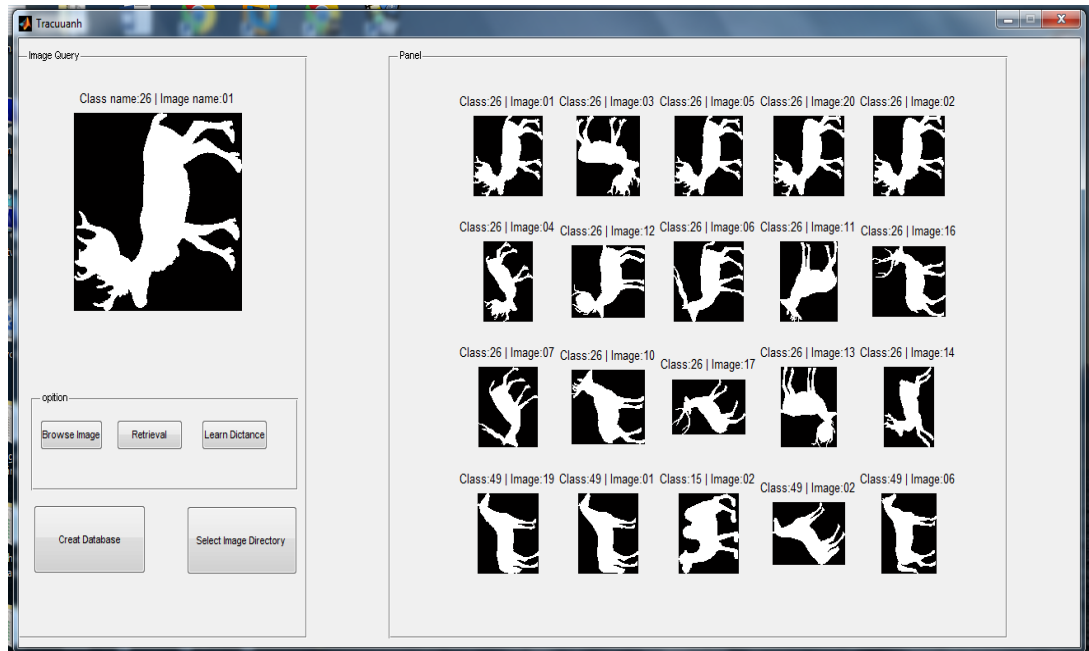


Hình 3.3: Giao diện chọn thư mục chứa ảnh truy vấn cần tra cứu
Trong phần này ảnh truy vấn nằm trong thư mục Data

Ví dụ 1: Ảnh truy vấn là ảnh 01 thuộc lớp 25

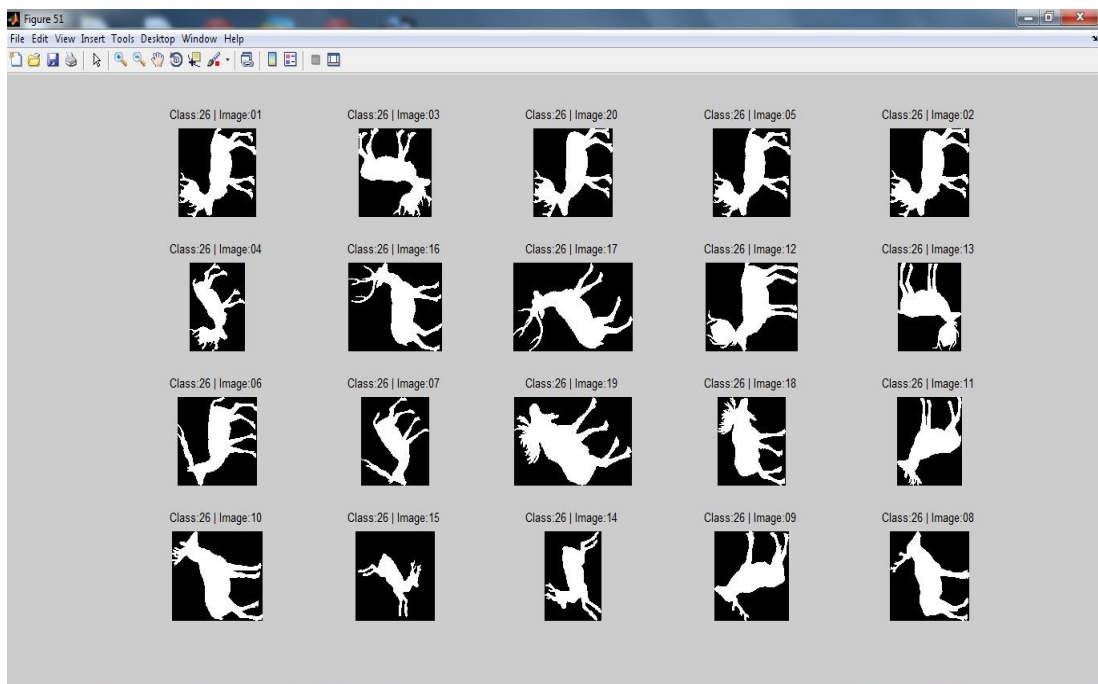


Hình 3.4: Giao diện hiện thị ảnh truy vấn cho ví dụ 1



Hình 3.5: Giao diện hiển thị kết quả tra cứu trước khi học hàm khoảng cách của ví dụ 1

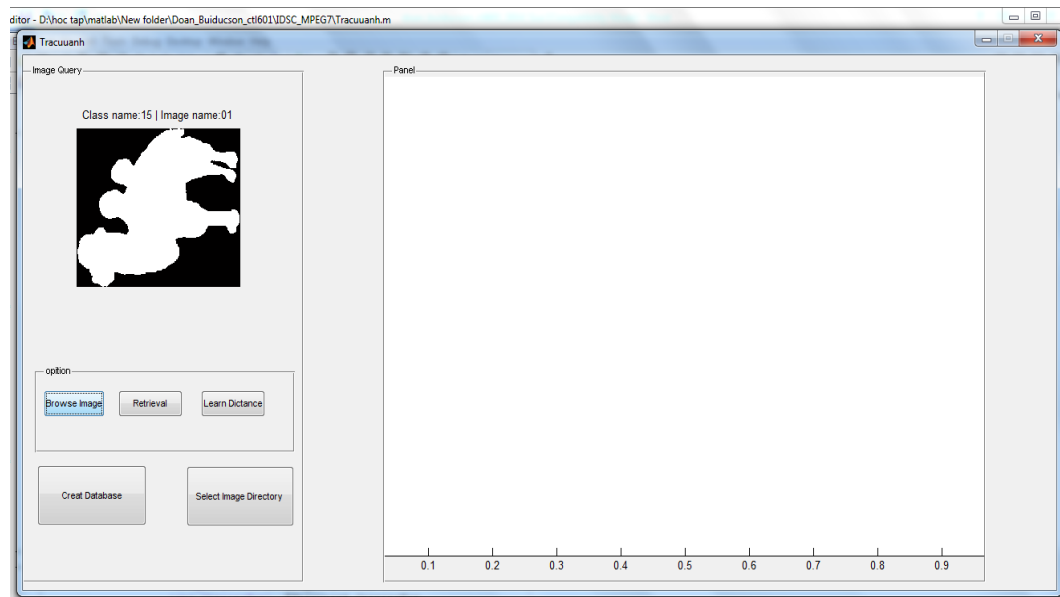
Ta có thể thấy kết quả tra cứu trả về một số ảnh có độ tương tự chưa giống so với ảnh truy vấn, như ảnh 19 của lớp 49 (hình con ngựa) hoặc ảnh 02 của lớp 15 (hình con lạc đà), vậy để giải quyết vấn đề trên chúng ta cần áp dụng kỹ thuật học hàm khoảng cách nhằm cải thiện lại độ chính xác của kết quả tra cứu



Hình 3.6: Giao diện kết quả hiển thị sau khi áp dụng kỹ thuật học hàm khoảng cách của ví dụ 1

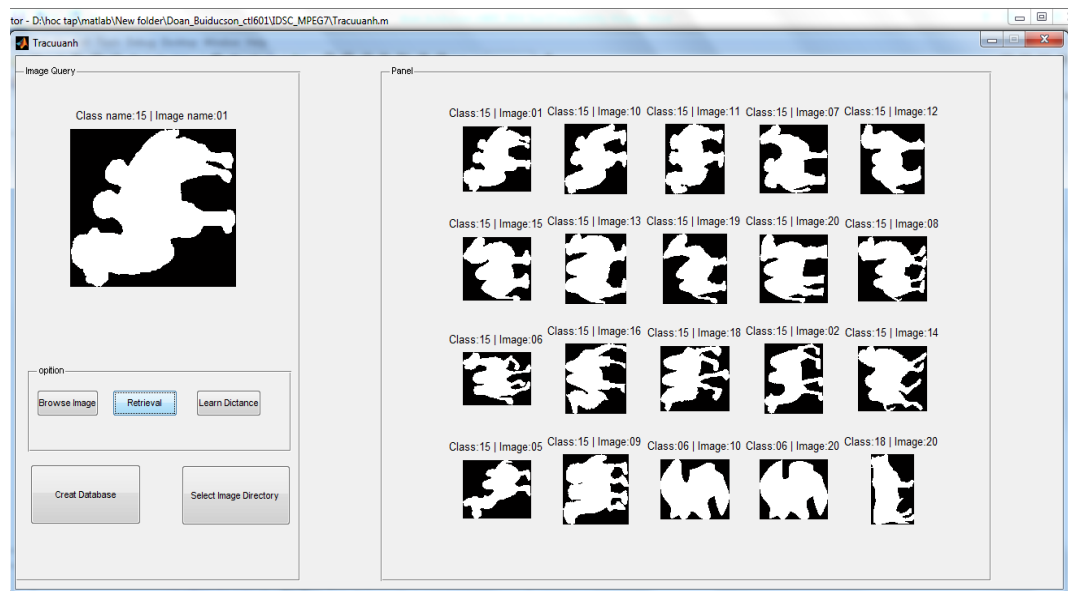
Chúng ta có thể thấy, lớp 49 (con ngựa) và lớp 15 (con lạc đà) đã không còn hiển thị trên kết quả tra cứu, thay vào đó là hiển thị các ảnh thuộc lớp 26 (lớp linh dương) cùng lớp với ảnh truy vấn, chứng tỏ việc học hàm khoảng cách đã giúp cải thiện lại độ chính xác của kết quả tra cứu.

Ví dụ 2: Thay đổi ảnh truy vấn đầu vào là ảnh 01 lớp 15



Hình 3.7: Giao diện hiển thị ảnh truy vấn của ví dụ 2

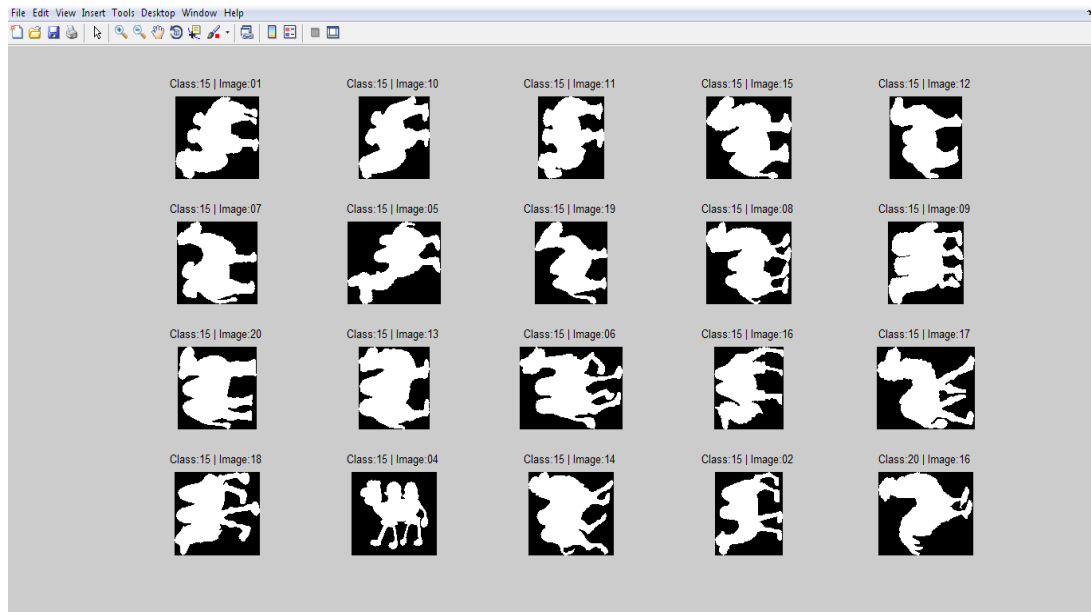
Kết quả tra cứu trước khi học hàm khoảng cách



Hình 3.8: Giao diện hiển thị kết quả tra cứu trước khi học hàm khoảng cách của ví dụ 2

Chúng ta có thể thấy rằng kết quả hiển thị chưa chính xác bởi vẫn còn nhiều ảnh không cùng lớp với ảnh truy vấn (lớp 6 và lớp 18)

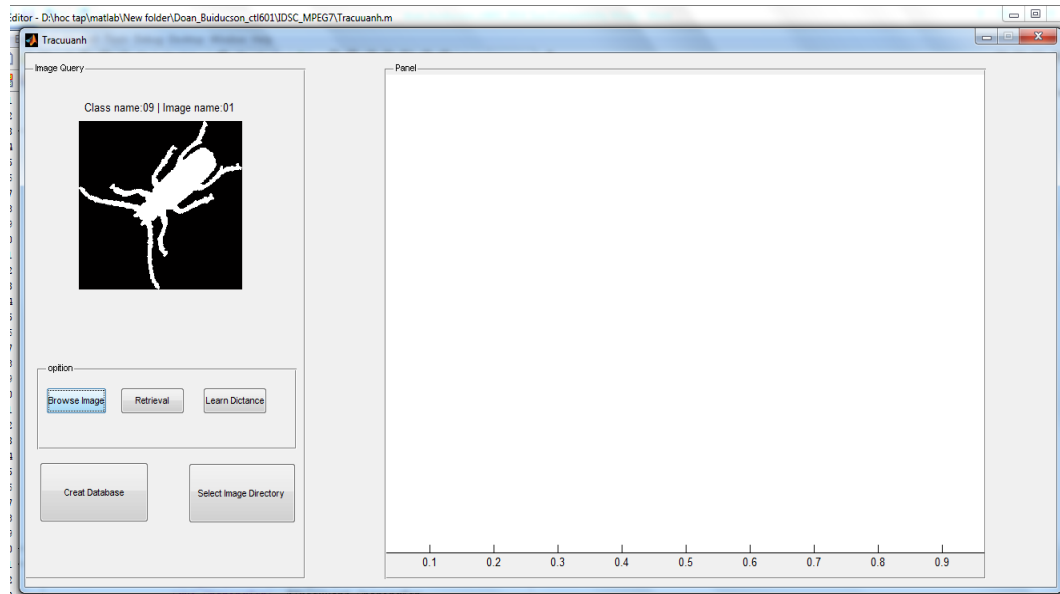
Ảnh kết quả sau khi được học



Hình 3.9: Giao diện hiển thị kết quả tra cứu sau khi học hàm khoảng cách của ví dụ 2

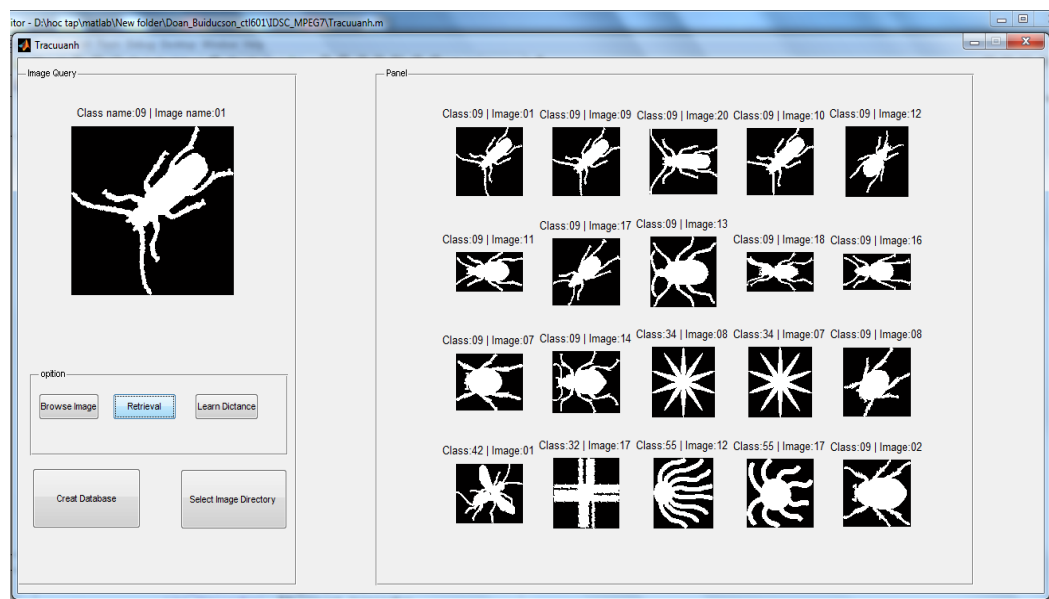
Nhìn vào kết quả hiển thị sau khi học hàm khoảng cách ta thấy được độ chính xác đã được cải thiện đáng kể, cụ thể là số ảnh thuộc cùng 1 lớp với ảnh truy vấn đã tăng lên, tuy nhiên vẫn còn tồn tại 1 ảnh không cùng lớp với ảnh truy vấn.

Ví dụ 3: Thay đổi ảnh truy vấn đầu vào là ảnh 01 thuộc lớp 09



Hình 3.10: Giao diện hiển thị ảnh truy vấn của ví dụ 3

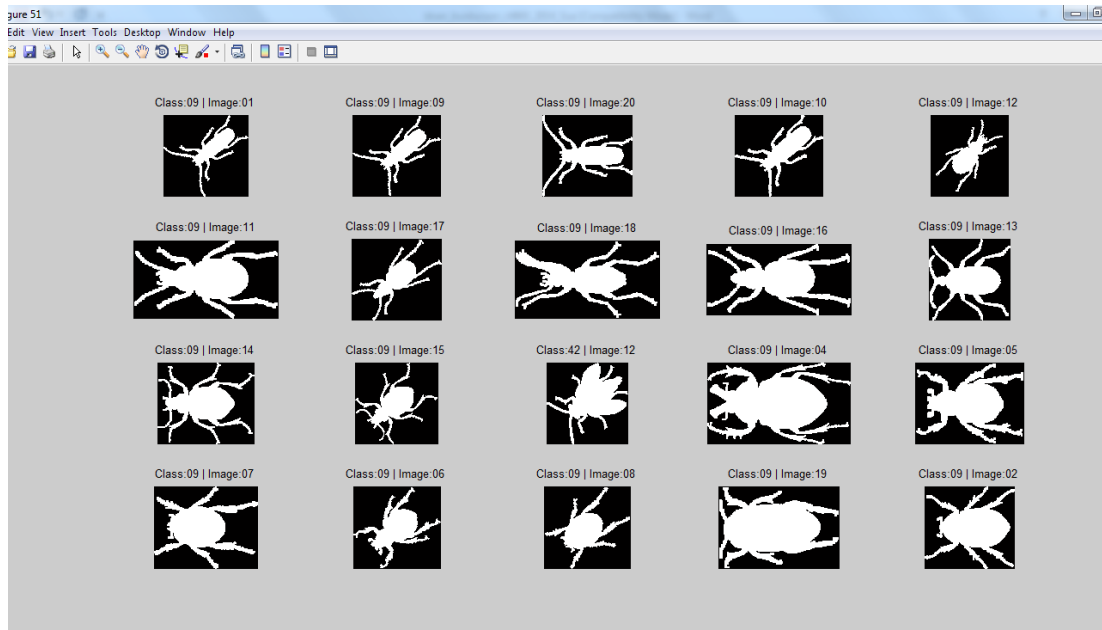
Kết quả trước khi được học hàm khoảng cách



Hình 3.11: Giao diện hiển thị kết quả tra cứu của ví dụ 3

Nhìn vào kết quả tra cứu ta thấy số ảnh khác lớp với ảnh truy vấn khá nhiều, bên cạnh đó ảnh số 09 thuộc lớp 09 giống với ảnh truy vấn hơn, nhưng lại được xếp sau ảnh 08 thuộc lớp 34 khác với ảnh truy vấn hơn.

Kết quả sau khi được học hàm khoảng cách



Hình 3.12: Giao diện hiển thị kết quả tra cứu sau khi học hàm khoảng cách của ví dụ 3

Nhìn vào kết quả tra cứu sau khi được học hàm khoảng cách ta thấy, các hình khác lớp với ảnh truy vấn đã không còn được hiển thị ở kết quả tra cứu, thay vào đó là các ảnh thuộc cùng lớp với ảnh truy vấn đã được hiển thị. Chứng tỏ, học hàm khoảng cách đã cải thiện độ chính xác của kết quả tra cứu.

KẾT LUẬN

Nội dung đồ án đã trình bày được một số các khái niệm về tra cứu ảnh theo nội dung, kĩ thuật trích chọn đặc trưng hình dạng sử dụng phương pháp khoảng cách trong và kĩ thuật học hàm khoảng cách để cải thiện hiệu quả tra cứu. Về mặt thực nghiệm, chương trình tra cứu hình dạng sử dụng học hàm khoảng đã được cài đặt bằng ngôn ngữ Matlab. Kết quả thực nghiệm cho thấy độ chính xác đã được cải thiện đáng kể so với kết quả trước khi được áp dụng hàm học, tuy nhiên vẫn còn một vài vấn đề hạn chế, như là tốc độ tra cứu chưa được nhanh, hay kết quả tra cứu chưa chính xác 100%.

Mặc dù đã hoàn thành được mục tiêu chính của đồ án nhưng do điều kiện về thời gian có hạn mà lĩnh vực cần tìm hiểu cũng tương đối rộng nên những gì tìm hiểu được trong đồ án sẽ khó tránh khỏi những thiếu sót. Chương trình thử nghiệm cũng chưa thực sự hoàn thiện nhưng đó cũng là một kết quả khả quan. Trong thời gian tới nếu có điều kiện em sẽ xây dựng lại chương trình tra cứu hình dạng ứng dụng học hàm khoảng cách một cách hoàn chỉnh hơn, tối ưu thuật toán nhằm tăng tốc độ tra cứu và độ chính xác của kết quả hiển thị.

TÀI LIỆU THAM KHẢO

- [1]. Datta, R., Joshi, D., Li, J., Wang, J.Z.: *Image retrieval: Ideas, influences, and trends of the new age*. *ACM Computing Surveys* 40 (2008)
- [2]. H. Ling and D. Jacobs, “*Shape classification using the inner-distance,*” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, . 29 (2007).
- [3]. Xingwei Yang “*Improving Shape Retrieval by Learning Graph Transduction*” In: *Proceedings of the 10th European Conference on Computer Vision* (2008).
- [4]. Datta, R., Joshi, D., Li, J., Wang, J.Z.: *Content-based image retrieval: approaches and trends of the new age*, In: *Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval* (2005).
- [5]. Xie, J., Heng, P.A., Shah, M.: *Shape matching and modeling using skeletal context*. *Pattern Recognition*. 41 (2008) .

Website: www.matworks.com