

**ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**

Nguyễn Thị Hoàn

**PHƯƠNG PHÁP TRÍCH CHỌN ĐẶC TRƯNG ẢNH
TRONG THUẬT TOÁN HỌC MÁY TÌM KIẾM ẢNH ÁP
DỤNG VÀO BÀI TOÁN TÌM KIẾM SẢN PHẨM**

KHOÁ LUẬN TỐT NGHIỆP ĐẠI HỌC HỆ CHÍNH QUY

Ngành: Công Nghệ Thông Tin

Hà Nội – 2010

**ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**

Nguyễn Thị Hoàn

**PHƯƠNG PHÁP TRÍCH CHỌN ĐẶC TRƯNG ẢNH
TRONG THUẬT TOÁN HỌC MÁY TÌM KIẾM ẢNH ÁP
DỤNG VÀO BÀI TOÁN TÌM KIẾM SẢN PHẨM**

KHOÁ LUẬN TỐT NGHIỆP ĐẠI HỌC HỆ CHÍNH QUY

Ngành: Công Nghệ Thông Tin

Cán bộ hướng dẫn: PGS.TS. Hà Quang Thụy

Cán bộ đồng hướng dẫn: Ths. Nguyễn Cẩm Tú

Hà Nội - 2010

Lời cảm ơn

Trước tiên, tôi xin gửi lời cảm ơn và lòng biết ơn sâu sắc nhất tới Phó Giáo sư Tiến sĩ Hà Quang Thụy và Thạc sĩ Nguyễn Cẩm Tú, người đã tận tình chỉ bảo và hướng dẫn tôi trong suốt quá trình thực hiện khoá luận tốt nghiệp.

Tôi chân thành cảm ơn các thầy, cô đã tạo những điều kiện thuận lợi cho tôi học tập và nghiên cứu tại trường Đại học Công nghệ.

Tôi cũng xin gửi lời cảm ơn tới các anh chị, các bạn và các em sinh viên trong phòng nghiên cứu SIS-KTLab đã giúp tôi rất nhiều trong việc hỗ trợ kiến thức chuyên môn để hoàn thành tốt khoá luận.

Cuối cùng, tôi muốn gửi lời cảm vô hạn tới gia đình và bạn bè, những người thân yêu luôn bên cạnh và động viên tôi trong suốt quá trình thực hiện khoá luận tốt nghiệp.

Tôi xin chân thành cảm ơn !

Sinh viên

Nguyễn Thị Hoàn

Tóm tắt

Sự phát triển mạnh mẽ của công nghệ ảnh số làm lượng ảnh lưu trữ trên web tăng lên một cách nhanh chóng đòi hỏi phải có các công cụ hỗ trợ tìm kiếm ảnh hiệu quả và tiện lợi. Mặc dù các công cụ tìm kiếm ảnh theo văn bản đi kèm ảnh ra đời cho phép người dùng tìm kiếm ảnh với thời gian đáp ứng khá nhanh, tuy nhiên, các công cụ này vẫn còn hạn chế trong việc giải quyết nhập nhằng giữa nội dung câu truy vấn và nội dung hiển thị của ảnh trả về. Sự ra đời của các công cụ tìm kiếm ảnh theo nội dung ảnh đã giải quyết được những nhập nhằng trên.

Mục tiêu của khóa luận là nghiên cứu các phương pháp biểu diễn đặc trưng ảnh để nâng cao chất lượng tìm kiếm ảnh. Đầu tiên, khóa luận khảo sát phương pháp trích chọn đặc trưng ảnh trong tìm kiếm và xếp hạng ảnh. Tiếp đó, dựa theo phương pháp lượng tử hóa tích của Hervé Jégou và cộng sự [12], khóa luận đưa ra một mô hình tìm kiếm k lảng giềng gần nhất kết hợp độ đo tương đồng về khoảng cách giữa các vector đặc trưng và tiến hành thực nghiệm mô hình. Thực nghiệm ban đầu cho thấy, từ một ảnh truy vấn đầu vào hệ thống trả về 10 ảnh tương đồng nhất đối với mỗi truy vấn với độ chính xác 80.4% và đây là một kết quả khả quan.

Mục lục

Mở đầu	1
Chương 1. Khái quát về trích chọn đặc trưng ảnh và tìm kiếm theo đặc trưng ảnh	3
1.1. Đặt vấn đề.....	3
1.2. Đặc trưng văn bản đi kèm ảnh và tìm kiếm ảnh theo văn bản đi kèm ảnh.....	3
1.3. Đặc trưng nội dung ảnh và tìm kiếm theo đặc trưng nội dung.....	5
Tổng kết chương 1.....	8
Chương 2. Các phương pháp lựa chọn đặc trưng và độ đo tương đồng giữa các ảnh	10
2.1. Đặt vấn đề.....	10
2.2. Đặc trưng màu sắc.....	11
2.2.1. Đặc trưng màu sắc	11
2.2.2. Độ đo tương đồng cho màu sắc	11
2.3. Đặc trưng kết cấu	12
2.3.1. Đặc trưng kết cấu.....	12
2.3.2. Độ đo tương đồng cho kết cấu	12
2.4. Đặc trưng hình dạng.....	13
2.4.1. Đặc trưng hình dạng.....	13
2.4.2. Độ đo tương đồng cho hình dạng	13
2.5. Đặc trưng cục bộ bất biến.....	13
2.5.1. Đặc trưng cục bộ bất biến	14
2.5.2. Độ đo tương đồng cho đặc trưng cục bộ bất biến	18
2.6. Lựa chọn đặc trưng	18
Tổng kết chương 2.....	20
Chương 3. Một số phương pháp tìm kiếm ảnh theo nội dung.....	21
3.1. Phương pháp PageRank cho tìm kiếm ảnh sản phẩm.....	21
3.2. CueFlik: Một phương pháp xếp hạng lại ảnh dựa trên luật của người dùng ...	22

3.3. Phương pháp tìm kiếm ảnh dựa trên màu sắc, hình dạng, kết cấu của ảnh	24
3.3.1. Lưới.....	25
3.3.2. Tích hợp các đối sánh ảnh.....	25
3.3.3. Hình dạng:	26
3.4. Phương pháp tìm kiếm ảnh dựa vào nội dung sử dụng các phân vùng ảnh như mẫu truy vấn.....	26
Tổng kết chương 3	27
Chương 4. Mô hình k láng giềng gần nhất sử dụng bộ lượng tử hóa	28
4.1. Đặt vấn đề.....	28
4.2. Cơ sở lý thuyết.....	28
4.2.1. Các ký hiệu và khái niệm.....	28
4.2.2. Tìm kiếm sử dụng lượng tử hóa	30
4.2.3. Tìm kiếm không toàn bộ	31
4.3. Mô hình bài toán	33
4.3.1. Trích chọn đặc trưng ảnh	33
4.3.2. Tìm kiếm K láng giềng gần nhất	34
Tổng kết chương 4	35
Chương 5. Thực nghiệm và đánh giá	36
5.1. Môi trường và các công cụ sử dụng cho thực nghiệm.....	36
5.2. Xây dựng tập dữ liệu ảnh	37
5.3. Quy trình, phương pháp thực nghiệm	38
5.4. Kết quả thực nghiệm	38
Tổng kết chương 5	41
Kết luận	42
Tài liệu tham khảo	43

Danh sách các bảng

Bảng 1. Cấu hình phần cứng sử dụng trong thực nghiệm	36
Bảng 2. Công cụ phần mềm sử dụng trong thực nghiệm	36
Bảng 3. Một số thư viện sử dụng trong thực nghiệm	37
Bảng 4. Kết quả độ chính xác trung bình của 10 truy vấn	40
Bảng 5. Độ chính xác mức k của một số truy vấn	40

Danh sách các hình vẽ

Hình 1. Ví dụ hiển thị một ảnh.....	4
Hình 2. Ví dụ truy vấn của Google.....	5
Hình 3. Ví dụ truy vấn của Google.....	5
Hình 4. Ví dụ về một số loại kết cấu	6
Hình 5. Một kết quả trả về của Google Image Swirl.....	7
Hình 6. Một kết quả trả về của Tiltomo.....	7
Hình 7. Một kết quả trả về của Byo Image Search	8
Hình 8. Biểu đồ mô phỏng việc tính toán các DoG ảnh từ các ảnh kè mờ	15
Hình 9. Mỗi điểm ảnh được so sánh với 26 láng giềng của nó.....	16
Hình 10. Quá trình lựa chọn các điểm hấp dẫn.....	17
Hình 11. Biểu diễn các vector đặc trưng	18
Hình 12. Ví dụ các ảnh sản phẩm trả về từ hệ thống của Jing	22
Hình 13. Tổng quan về mô hình của hệ thống tìm kiếm theo màu sắc, kết cấu và hình dạng.....	25
Hình 14. Mô hình hệ thống IVFADC.....	33
Hình 15. Mô hình giải quyết bài toán.....	34
Hình 16. 10 kết quả trả về đầu tiên của hệ thống với truy vấn Apple.....	41

Danh sách các từ viết tắt

STT	Từ viết tắt	Từ viết đầy đủ
1	ADC	Asymmetric distance computation
2	AP	Average Precision
3	BDA	Biased Discriminant analysis
4	CBIR	Content Based Images Retrieval
5	DoG	Difference of Gaussian
6	IVFADC	Inverted file asymmetric distance Computation
7	JSD	Jensen-Shannon divergence
8	MAP	Mean Average Precision
9	MDA	Multiple Discriminant analysis
10	QBIC	Query Based Image Content
11	SDC	Symmetric distance computation
12	SIFT	Scale Invariant feature transform
13	SMMS	Symmetric maximized minimal distance in subspace

Danh sách tham chiếu thuật ngữ Anh – Việt

STT	Thuật ngữ tiếng Anh	Thuật ngữ tiếng Việt
1	Asymmetric distance	Khoảng cách bất đối xứng
2	Biased Discriminant analysis	Phân tích biệt thức không đối xứng
3	Boosting manner	Tăng khuyếch đại
4	Content Based Images Retrieval	Tìm kiếm ảnh theo nội dung
5	Co-occurrence matrix	Ma trận đồng xuất hiện
6	Cotourlet transform	Biến đổi đường viền
7	Discriminant analysis	Phân tích biệt thức
8	Distribution based method	Phương pháp dựa vào phân phối
9	Feature contrast Model	Mô hình tương phản đặc trưng
10	Feature selection	Lựa chọn đặc trưng
11	Gabor Wavelet transform	Biến đổi sóng Gabor
12	Global texture descriptor	Đặt tả kết cấu toàn cục
13	Image Segment	Phân vùng ảnh
14	Interest point	Điểm hấp dẫn
15	Inverted file asymmetric distance computation	Tính toán khoảng cách bất đối xứng file chỉ mục ngược
16	Inverted list	Danh sách chỉ mục ngược
17	Local features	Đặc trưng cục bộ
18	Local interest Point	Điểm hấp dẫn cục bộ
19	Local scale – invariant feature	Đặc trưng cục bộ bất biến
20	Mean Average Precision	Độ chính xác trung bình
21	Metadata	Siêu dữ liệu
22	Non exhaustive search	Tìm kiếm không toàn bộ
23	Product quantization	Lượng tử hóa tích
24	Quantization code	Mã lượng tử hóa
25	Query Based Image Content	Truy vấn theo nội dung ảnh
26	Similarity measurement	Độ đo tương đồng
27	Symmetric distance	Khoảng cách đối xứng
28	Texture	Kết cấu
29	The complex directional filter	Bộ lọc định hướng phức tạp
30	The steerable pyramid	Kim tự tháp có thể lái được
31	Visual hyperlinks	Siêu liên kết trực quan

Mở đầu

Cùng với sự bùng nổ thông tin trên web và sự phát triển của công nghệ kỹ thuật số, lượng ảnh lưu trữ trên Web cũng tăng một cách nhanh chóng. Vì vậy, việc xây dựng các hệ thống tìm kiếm và xếp hạng ảnh là rất cần thiết và thực tế đã có nhiều công cụ tìm kiếm ảnh thương mại xuất hiện. Các công cụ tìm kiếm ảnh thường dựa vào hai đặc trưng chính là văn bản đi kèm ảnh hoặc nội dung ảnh. Một số công cụ tìm kiếm ảnh theo văn bản đi kèm như Google Image Search, Yahoo!, MSN,... Một số công cụ tìm kiếm ảnh dựa vào nội dung ảnh như Google Image Swirl, Bing, Tiltomo, Tineye,... Tuy nhiên, việc tìm kiếm chỉ dựa vào văn bản đi kèm còn có nhiều nhầm lẫn giữa nội dung hiển thị ảnh và nội dung văn bản đi kèm ảnh trong quá trình tìm kiếm. Ví dụ, với truy vấn “Apple”, máy tìm kiếm khó phân biệt được người dùng muốn tìm hình ảnh quả táo hay logo của hãng Apple. Những công cụ tìm kiếm ảnh theo nội dung của các bức ảnh ra đời tỏ ra ưu thế vì hạn chế được những nhầm lẫn trên.

Tìm kiếm ảnh theo nội dung đã nhận được nhiều sự quan tâm của các nhà khoa học. Nhiều công trình nghiên cứu về tìm kiếm ảnh theo nội dung được đăng trên các tạp chí như International Journal of Computer Vision, IEEE conference... Nhóm nghiên cứu chúng tôi đã tiến hành một số nghiên cứu bước đầu liên quan đến xếp hạng ảnh dựa vào độ tương đồng theo nội dung ảnh trong công tác sinh viên nghiên cứu khoa học.

Khóa luận “Phương pháp trích chọn đặc trưng ảnh trong học máy tìm kiếm ảnh và ứng dụng trong tìm kiếm sản phẩm” nhằm khảo sát, phân tích một số phương pháp trích chọn đặc trưng ảnh phổ biến và tìm kiếm ảnh theo ảnh mẫu, thử nghiệm hệ thống trong ứng dụng tìm kiếm sản phẩm.

Ngoài phần **MỞ ĐẦU** này, khóa luận bao gồm các nội dung sau:

- Chương 1. Khái quát về lựa chọn đặc trưng cho tìm kiếm ảnh. Các đặc trưng về văn bản đi kèm ảnh và đặc trưng về nội dung ảnh.
- Chương 2. Các phương pháp lựa chọn đặc trưng và độ đo tương tự giữa ảnh. Trình bày một số đặc trưng về nội dung ảnh và một số độ đo tương đồng tương ứng với các đặc trưng.

- Chương 3. Một số phương pháp tìm kiếm và xếp hạng ảnh dựa trên nội dung của ảnh. Giới thiệu một số công trình nghiên cứu liên quan đến tìm kiếm ảnh theo nội dung ảnh.
- Chương 4. Mô hình tìm kiếm K láng giềng gần nhất. Giới thiệu mô hình tìm kiếm K láng giềng gần nhất, phương pháp lưu trữ và đánh chỉ mục trong tìm kiếm.
- Chương 5. Thực nghiệm. Trình bày quá trình thực nghiệm, kết quả, nhận xét, đánh giá khi áp dụng mô hình K láng giềng gần nhất với các đặc trưng trích chọn trong tìm kiếm ảnh sản phẩm.
- Cuối cùng là phần **KẾT LUẬN**. Tổng kết các kết quả chính của khóa luận và phương hướng nghiên cứu tiếp theo.

Chương 1. Khái quát về trích chọn đặc trưng ảnh và tìm kiếm theo đặc trưng ảnh

1.1. Đặt vấn đề

Sự phát triển mạnh mẽ của công nghệ ảnh số làm lượng ảnh lưu trữ trên web tăng lên một cách nhanh chóng. Mỗi ngày, có hàng triệu bức ảnh được đăng tải trên các trang ảnh trực tuyến như: Flickr¹, Photobucket², Facebook³,.... Theo thống kê, có 10 tỉ ảnh trên Facebook (tính đến tháng 10/2008), 3 tỉ ảnh trên Flickr (tính đến tháng 11/2008), 6.2 tỉ ảnh trên Photobucket(tính đến tháng 10/2008) [36].

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

Chương 1 trình bày về các đặc trưng của ảnh gồm đặc trưng văn bản đi kèm ảnh và đặc trưng về nội dung ảnh(màu sắc, kết cấu, hình dạng, đặc trưng cục bộ) và một số vấn đề về tìm kiếm ảnh.

1.2. Đặc trưng văn bản đi kèm ảnh và tìm kiếm ảnh theo văn bản đi kèm ảnh.

Mỗi ảnh trên web thường có các văn bản đi kèm như là tên ảnh (title), các thẻ (tags), bình luận (comment),...để mô tả các thông tin về ảnh, đây là các siêu dữ liệu

¹ Flickr: <http://www.flickr.com>

² Photobucket: <http://www.photobucket.com>

³ Facebook: <http://www.facebook.com>

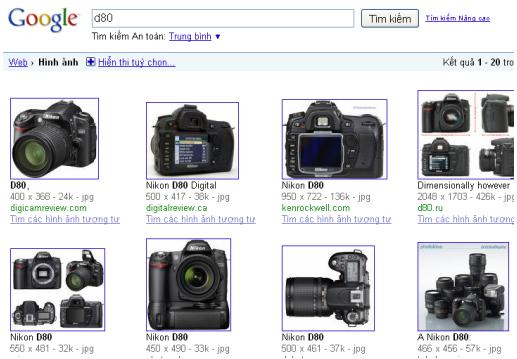
(metadata) về ảnh. Các dữ liệu này thường do người dùng tạo ảnh gắn cho mỗi ảnh, vì vậy chúng đều mang một ý nghĩa nhất định. Độ quan trọng của các loại siêu dữ liệu khác nhau cũng khác nhau. Ví dụ, các thẻ thường quan trọng hơn tên ảnh, tên ảnh quan trọng hơn bình luận. Dưới đây là một ví dụ về văn bản đi kèm một ảnh:

- Title: “Red_Rose Flower”
- Tags: “redRoseflower, hongkongflowershow, 2009, bokeh, causewaybay, hongkong, jonnoj, jonbinalay, nikond80, interestingness50”
- Description: “HEAVEN SCENT”...FOR THE LOVE OF THE RED RED ROSE...
- Content:



Hình 1. Ví dụ hiển thị một ảnh

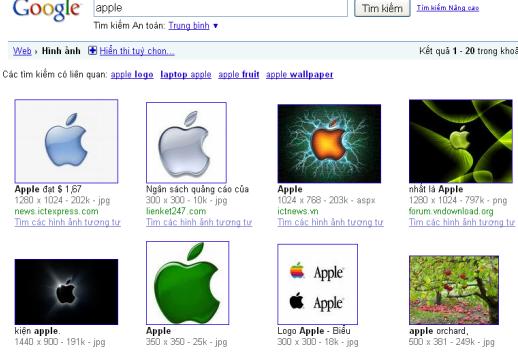
Vì văn bản đi kèm ảnh mang ngữ nghĩa về nội ảnh cho nên hai bức ảnh có nội dung giống nhau thường có tên giống nhau và các thẻ tương tự nhau. Vì vậy, các công cụ tìm kiếm ảnh theo văn bản đi kèm thường tập trung khai thác nội dung của các văn bản này để tìm kiếm và xếp hạng ảnh. Phương pháp này cho kết quả khả quan cũng như đáp ứng nhanh nhu cầu của người sử dụng. Tuy nhiên, với các câu truy vấn mang ý nghĩa nhập nhằng có thể các kết quả trả về sẽ không đúng với yêu cầu đặt ra. Ví dụ khi truy vấn là “d-80”, một máy ảnh phổ biến của Nikon, thì các hệ thống trả về kết quả khá tốt (hình 2). Tuy nhiên, với truy vấn “apple”, nếu người dùng muốn tìm quả táo thì kết quả trả về đầu tiên không thỏa mãn (logo của hãng Apple) (hình 3):



Hình 2. Ví dụ truy vấn của Google

Kết quả với truy vấn “d-80”

Mặt khác, các albumn cá nhân thường không có các thẻ hoặc văn bản đi kèm ảnh. Cùng với số lượng ảnh số được chụp thêm mỗi ngày, việc gán thủ công các thẻ cho ảnh rất tốn kém. Một hướng nghiên cứu nhằm khắc phục vấn đề trên là tìm kiếm theo chính các đặc trưng xuất từ nội dung của ảnh.



Hình 3. Ví dụ truy vấn của Google

Kết quả với truy vấn “Apple”

Tìm kiếm ảnh theo nội dung (Content Based Images Retrieval CBIR) hay truy vấn theo nội dung ảnh (Query Based Image Content QBIC) là một ứng dụng của thị giác máy tính đối với bài toán tìm kiếm ảnh [30][35]. “Dựa vào nội dung ảnh (Content-Based)” nghĩa là việc tìm kiếm sẽ phân tích nội dung thực sự của các bức ảnh. Nội dung ảnh ở đây được thể hiện bằng màu sắc, hình dạng, kết cấu (texture), các đặc trưng cục bộ (local features), ... hay bất cứ thông tin nào có từ chính nội dung ảnh. Cụm từ CBIR được T.Kato đưa ra vào năm 1992 trong quá trình thu thập ảnh một cách tự động từ cơ sở dữ liệu dựa trên biểu diễn màu sắc và hình dạng của ảnh. Tee Cheng Siew đã giới thiệu một số đặc trưng nội dung ảnh[23]:

- Đặc trưng màu sắc: Màu sắc là một đặc trưng nổi bật và được sử dụng phổ biến nhất trong tìm kiếm ảnh theo nội dung. Mỗi một điểm ảnh (thông tin màu sắc) có thể được biểu diễn như một điểm trong không gian màu sắc ba chiều. Các không gian màu sắc thường dùng là: RGB, Munsell, CIE, HSV. Tìm kiếm ảnh theo màu sắc tiến hành tính toán biểu đồ màu cho mỗi ảnh để xác định tỉ trọng các điểm ảnh của ảnh mà chứa các giá trị đặc biệt (màu sắc). Các nghiên cứu gần đây đang cố gắng phân vùng ảnh theo các màu sắc khác nhau và tìm mối quan hệ giữa các vùng này.

- Đặc trưng kết cấu: Trích xuất nội dung ảnh theo kết cấu nhằm tìm ra mô hình trực quan của ảnh và cách thức chúng được xác định trong không gian. Kết cấu được biểu diễn bởi các texel mà sau đó được đặt vào một số các tập phụ thuộc vào số kết cấu được phát hiện trong ảnh. Các tập này không chỉ xác định các kết cấu mà còn chỉ rõ vị trí các kết cấu trong ảnh. Việc xác định các kết cấu đặc biệt trong ảnh đạt được chủ yếu bằng cách mô hình các kết cấu như những biển thể cấp độ xám 2 chiều. Ví dụ về một số loại kết cấu[41]

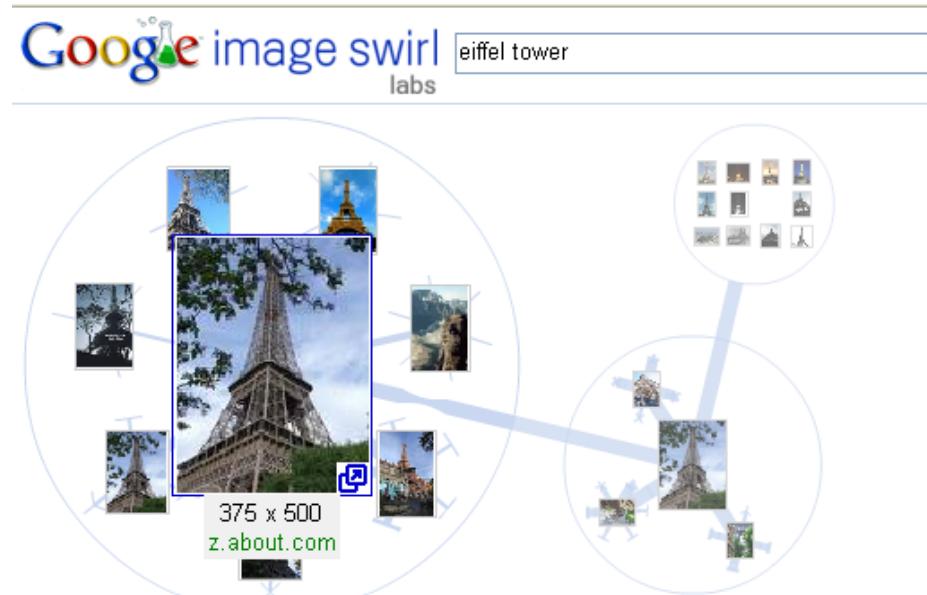


Hình 4. Ví dụ về một số loại kết cấu

- Đặc trưng hình dạng: Hình dạng của một ảnh hay một vùng là một đặc trưng quan trọng trong việc xác định và phân biệt ảnh trong nhận dạng mẫu. Mục tiêu chính của biểu diễn hình dạng trong nhận dạng mẫu là đo thuộc tính hình học của một đối tượng được dùng trong phân lớp, so sánh và nhận dạng đối tượng.

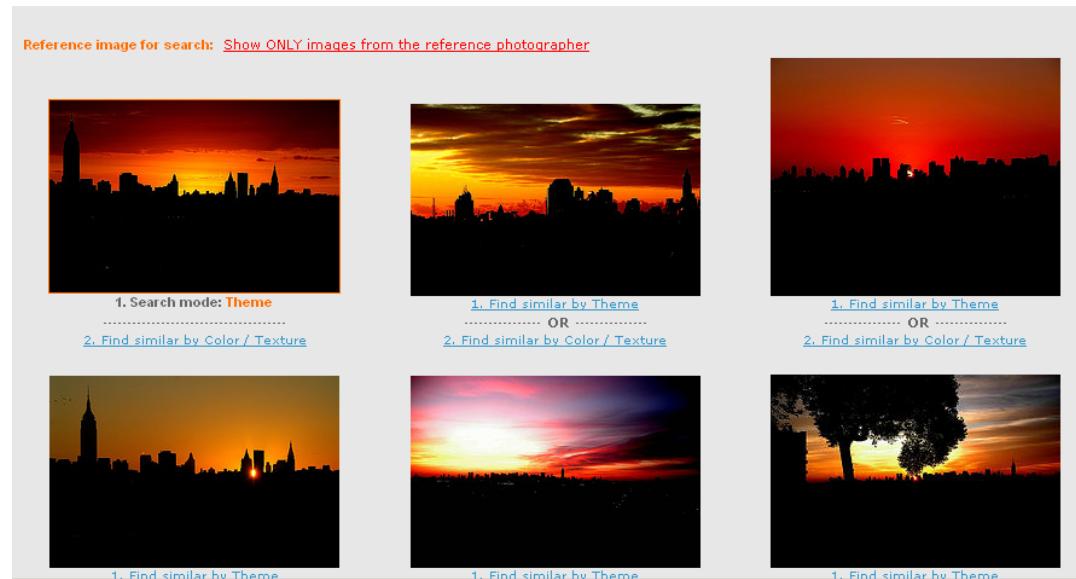
Thực tế, đã có nhiều máy tìm kiếm cho phép tìm kiếm ảnh theo nội dung ảnh, tuy nhiên, các máy tìm kiếm này thường chỉ khai thác vào một phần nội dung của ảnh.

- [Google Image Swirl](#): Là một thử nghiệm tìm kiếm hình ảnh theo nội dung của Google, trong đó, kết quả tìm kiếm được sẽ được tổ chức lại dựa vào hiển thị trực quan và độ tương đồng ngữ nghĩa giữa các ảnh. Google Image Swirl phân cụm tốp đầu các kết quả trả về cho trên 200.000 câu truy vấn và cho phép hiển thị hình ảnh dưới dạng các cụm và mối quan hệ giữa các ảnh.



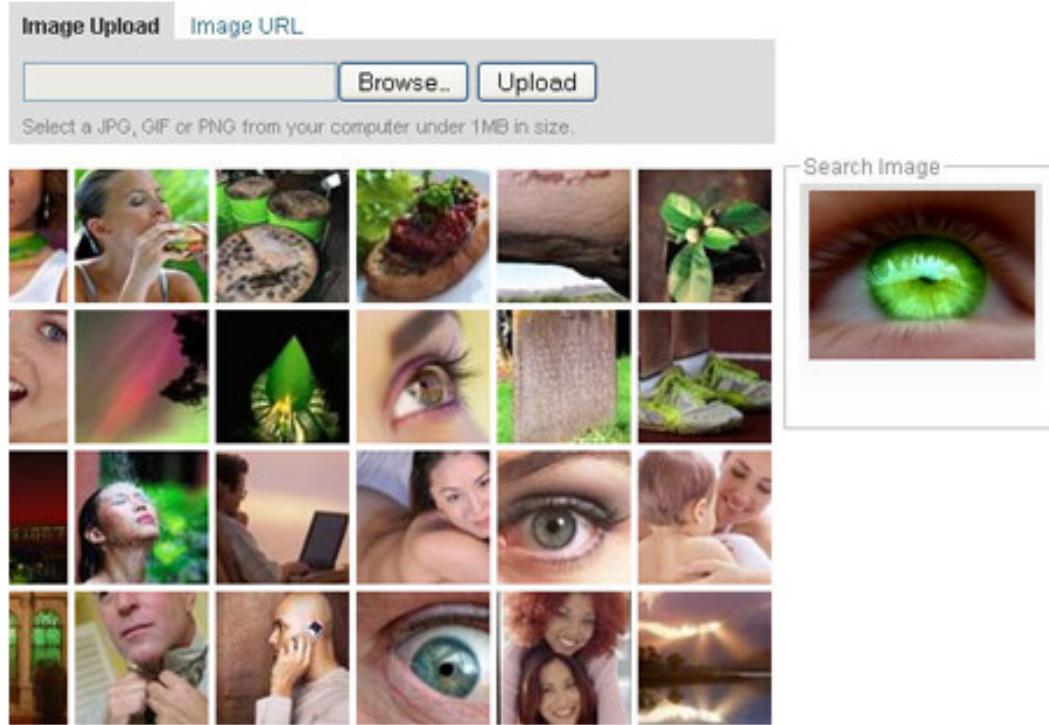
Hình 5. Một kết quả trả về của Google Image Swirl

- Tiltomo: Là một công cụ dựa trên Flickr và duy trì chính cơ sở dữ liệu ảnh của Flickr. Nó cho phép tìm kiếm ảnh dựa vào độ tương đồng về chủ đề, màu sắc hay kết cấu.



Hình 6. Một kết quả trả về của Tiltomo

- [Byo Image Search](#): Tìm kiếm ảnh theo độ tương đồng về màu sắc với mẫu ảnh mà người dùng tải lên từ máy tính hoặc từ một địa chỉ URL. Công cụ tìm kiếm này không hỗ trợ tính năng tìm kiếm ảnh dựa vào độ tương đồng về chủ đề.



Hình 7. Một kết quả trả về của Byo Image Search

Tìm kiếm ảnh theo mẫu (example-based image search): Tìm kiếm ảnh theo mẫu là một dạng của tìm kiếm ảnh dựa vào nội dung. Trong hệ thống đó, đầu vào là một ảnh, hệ thống tìm kiếm và trả lại cho người dùng những ảnh tương đồng với ảnh mẫu.

Trong nội khóa luận này, chúng tôi tập trung vào bài toán tìm kiếm ảnh dựa theo mẫu, tìm hiểu các phương pháp trích chọn đặc trưng nội dung cũng như các độ đo tương đồng để tìm kiếm tập ảnh sản phẩm gần với ảnh mẫu nhất trong tập cơ sở dữ liệu các ảnh sản phẩm.

Tổng kết chương 1

Trong chương này, chúng tôi trình bày khái quát đặc trưng văn bản đi kèm ảnh và đặc trưng nội dung của ảnh, và giới thiệu một số công cụ tìm kiếm dựa vào nội dung ảnh. Phương pháp tìm kiếm ảnh theo nội dung đã khắc phục được một phần

nhược điểm của phương pháp tìm kiếm ảnh theo văn bản đi kèm ảnh và cho ra những kết quả khả quan. Chương 2, khóa luận sẽ trình bày một số công trình nghiên cứu khoa học liên quan đến bài toán tìm kiếm ảnh theo nội dung.

Chương 2. Các phương pháp trích chọn đặc trưng và độ đo tương đồng giữa các ảnh

2.1. Đặt vấn đề

Trong tìm kiếm ảnh theo nội dung, việc lựa chọn các đặc trưng thích hợp với từng loại truy vấn và miền ứng dụng cùng với các độ đo tương đồng tương ứng là thành phần quan trọng và then chốt nhất[31]. Việc lựa chọn các đặc trưng và độ đo thích hợp sẽ giúp tăng cả tốc độ và mức độ chính xác của các hệ thống. J.V.Jawahe và cộng sự [32] đã nêu ra các yêu cầu cơ bản đối với thành phần lựa chọn đặc trưng cho ảnh:

- Thành phần lựa chọn đặc trưng phải lựa chọn được một tập các đặc trưng cung cấp đầu vào tốt nhất cho hệ thống tìm kiếm ảnh. Nếu số lượng các đặc trưng quá nhiều sẽ làm “che khuất” các “tín hiệu” (giảm các “tín hiệu” đối với tỉ lệ nhiễu), mặt khác, nếu số lượng các đặc trưng quá ít sẽ khó phân biệt được ảnh trong tìm kiếm.
- Nó phải giảm bớt được độ phức tạp trong lúc tính toán tổng thể bằng cách đa chiều của bài toán phân lớp.
- Khi người dùng muốn sử dụng các đặc trưng đó cho mọi truy vấn, thì việc sử dụng các đặc trưng này phải hiệu quả. Vì số lượng các đặc trưng có thể là hàng ngàn, do đó thời gian xử lý của module phải tuyến tính với số lượng đặc trưng.
- Vì thời gian xử lý của thành phần lựa chọn đặc trưng tuyến tính với số lượng đặc trưng, do đó việc lựa chọn các đặc trưng cũng nên tuyến tính dựa trên phân lớp.
- Thành phần lựa chọn đặc trưng có thể xử lý được với kích thước tập mẫu nhỏ (khoảng 5 mẫu).

Trong chương này, chúng tôi sẽ trình bày sơ bộ về các vấn đề về đặc trưng của ảnh(màu sắc, kết cấu, hình dạng, đặc trưng cục bộ SIFT), một số độ đo tương đồng tương ứng với các đặc trưng và phương pháp lựa chọn đặc trưng ảnh để tăng chất lượng tập đặc trưng.

2.2. Đặc trưng màu sắc

2.2.1. Đặc trưng màu sắc

Tìm kiếm ảnh theo lược đồ màu 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. Đây có thể xem là bước lọc đầu tiên cho những bước tìm kiếm sau. Một số lược đồ màu được sử dụng như: lược đồ màu RGB, lược đồ màu HSI, lược đồ HSI cải tiến. Trong đó, lược đồ màu RGB được sử dụng phổ biến nhất[18][20].

- Lược đồ màu RGB:

Đối với ảnh 256 màu, lược đồ màu của ảnh tương đương với lược đồ màu của ảnh xám. Đối với ảnh 24 bit màu, lược đồ miêu tả khả năng kết nối về cường độ của ba kênh màu R, G, B. Lược đồ màu này được định nghĩa như sau:

$$h_{R,G,B}[r,g,b] = N * \text{Prob}\{R = r, G = g, B = b\} \quad (1)$$

Trong đó N là số lượng điểm có trong ảnh.

Lược đồ màu này được tính bằng cách rời rạc hóa từng màu trong ảnh, sau đó đếm số điểm ảnh của mỗi màu. Khi mà số lượng màu là có hạn, để thuận tiện hơn, người ta thường chuyển đổi ba kênh màu thành một biến giá trị duy nhất. Một cách khác để tính lược đồ màu của ảnh RGB là ta phân ra làm 3 lược đồ riêng biệt $h_r[], h_g[], h_b[]$. Khi đó, mỗi lược đồ được tính bằng cách đếm kênh màu tương ứng trong mỗi điểm ảnh.

2.2.2. Độ đo tương đồng về màu sắc

Một số độ đo tương đồng được sử dụng như: Độ đo khoảng cách Oclit, độ đo Jensen-Shannon divergence (JSD).

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

- Khoảng cách Oclit:

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

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

Hoặc:

$$\text{Intersection}(h(I), h(M)) = \sum_{j=1}^K |h(I)_j - h(M)_j| \quad (3)$$

- Độ đo Jensen-Shannon divergence (JSD) :

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

$$d_{JSD}(H, H') = \sum_{m=1}^M H_m \log \frac{2H_m}{H_m + H'_m} + H'_m \log \frac{2H'_m}{H'_m + H_m} \quad (4)$$

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

2.3. Đặc trưng kết cấu

2.3.1. Đặc trưng kết cấu

Hiện tại, vẫn chưa có một định nghĩa chính thức cụ thể về kết cấu. Kết cấu là một đối tượng dùng để phân hoạch ảnh ra thành những vùng quan tâm để phân lớp những vùng đó[27][24][18][23].

Kết cấu cung cấp thông tin về sự sắp xếp về mặt không gian của màu sắc và cường độ một ảnh. Kết cấu được đặc trưng bởi sự phân bổ không gian của những mức cường độ trong một khu vực lảng giềng với nhau. Kết cấu gồm các kết cấu gốc hay nhiều kết cấu gộp lại đôi khi gọi là texel.

Một số phương pháp dùng để trích xuất các đặc trưng kết cấu như[18]:

- Kim tự tháp "có thể lái được" (the steerable pyramid)
- Biến đổi đường viền (the contourlet transform)
- Biến đổi sóng Gabor (The Gabor Wavelet transform)
- Biểu diễn ma trận đồng hiện (co-occurrence matrix)
- Hệ thống bộ lọc định hướng phức tạp (The complex directional filter bank)

2.3.2. Độ đo tương đồng cho kết cấu ảnh

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

2.4. Đặc trưng hình dạng

2.4.1. Đặc trưng hình dạng

Màu sắc và kết cấu là những thuộc tính có khái niệm toàn cục trong một ảnh. Trong khi đó, hình dạng không phải là một thuộc tính của ảnh. Nói tới hình dạng không phải là nhắc đến hình dạng của một ảnh. Thay vì vậy, 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.

Trong tìm kiếm ảnh theo nội dung, hình dạng là một cấp cao hơn so với màu sắc và kết cấu. Nó đòi hỏi sự phân biệt giữa các vùng để tiến hành xử lý về độ đo của hình dạng. 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 (contour-based descriptor) : Biểu diễn các đường biên bao bên ngoài
- Biểu diễn theo vùng (region-based descriptor): Biểu diễn một vùng toàn vẹn

2.4.2. Độ đo tương đồng cho hình dạng

Độ đo về hình dạng rất nhiều trong phạm vi lý thuyết của bộ môn xử lý ảnh. Chúng trải rộng từ những độ đo toàn cục dạng thô với sự trợ giúp của việc nhận dạng đối tượng, cho tới những độ đo chi tiết tự động tìm kiếm những hình dạng đặc biệt. Lược đồ hình dạng là một ví dụ của độ đo đơn giản. Kỹ thuật dùng đường biên hiệu quả hơn phương pháp trước, chúng tìm kiếm những hình dạng đối tượng gần giống với đường biên nhất. Phương pháp vẽ phác họa là phương pháp có nhiều đặc trưng rõ ràng hơn, không chỉ tìm kiếm những đường biên đối tượng đơn, mà còn đối với tập những đối tượng đã được phân đoạn trong một ảnh mà người dùng vẽ hay cung cấp.

2.5. Đặc trưng cục bộ bất biến

Người ta thường chia đặc trưng cục bộ thành 2 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 Haris (interest points). Trong phần này, chúng tôi sẽ trình bày chi tiết về việc trích chọn các đặc trưng cục bộ bất biến (Scale Invariant Feature Transform SIFT) của ảnh.

2.5.1. Đặc trưng cục bộ bất biến

Phần này trình bày phương pháp trích rút các đặc trưng cục bộ bất biến SIFT của ảnh. Các đặc trưng này bất biến với việc thay đổi tỉ lệ ảnh, quay ảnh, đôi khi là thay đổi điểm nhìn và thêm nhiều ảnh hay thay đổi cường độ chiếu sáng của ảnh. Phương pháp được lựa chọn có tên là Scale-Invariant Feature Transform (SIFT) và đặc trưng trích rút được gọi là đặc trưng SIFT (SIFT Feature). Các đặc trưng SIFT này được trích rút ra từ các điểm hấp dẫn cục bộ (Local Interest Point) [17][30][16].

Điểm hấp dẫn (Interest Point (Keypoint)): Là vị trí (điểm ảnh) "hấp dẫn" trên ảnh. "Hấp dẫn" ở đây có nghĩa là điểm đó có thể có các đặc trưng bất biến với việc quay ảnh, co giãn ảnh hay thay đổi cường độ chiếu sáng của ảnh.

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

- **Phát hiện các điểm cực trị Scale-Space** (Scale-Space extrema detection): Bước đầu tiên này tiến hành tìm kiếm các điểm hấp dẫn trên tất cả các tỉ lệ và vị trí của ảnh. Nó sử dụng hàm different-of-Gaussian để xác định tất cả các điểm hấp dẫn tiềm năng mà bất biến với quy mô và hướng của ảnh.
- **Định vị các điểm hấp dẫn** (keypoint localization): Một hàm kiểm tra sẽ được đưa ra để quyết định xem các điểm hấp dẫn tiềm năng có được lựa chọn hay không?
- **Xác định hướng cho các điểm hấp dẫn** (Orientation assignment): Xác định hướng cho các điểm hấp dẫn được chọn
- **Mô tả các điểm hấp dẫn** (Keypoint descriptor): Các điểm hấp dẫn sau khi được xác định hướng sẽ được mô tả dưới dạng các vector đặc trưng nhiều chiều.

2.5.1.1. Phát hiện điểm cực trị Scale-space

Các điểm hấp dẫn với đặc trưng SIFT tương thích với các cực trị địa phương của bộ lọc difference -of-Gaussian (DoG) ở các tỉ lệ khác nhau. Định nghĩa không gian tỉ lệ của một hình ảnh là hàm $L(x,y,k\sigma)$ được mô tả như sau:

$$L(x,y,\sigma) = G(x,y,k\sigma) * I(x,y) \quad (5)$$

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

$I(x, y)$: Ảnh đầu vào

* là phép nhân chập giữa x và y

Và

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (6)$$

Để phát hiện được các điểm hấp dẫn, ta đi tìm các cực trị của hàm DoG được định nghĩa:

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ D(x, y, \sigma) &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \quad (7)$$

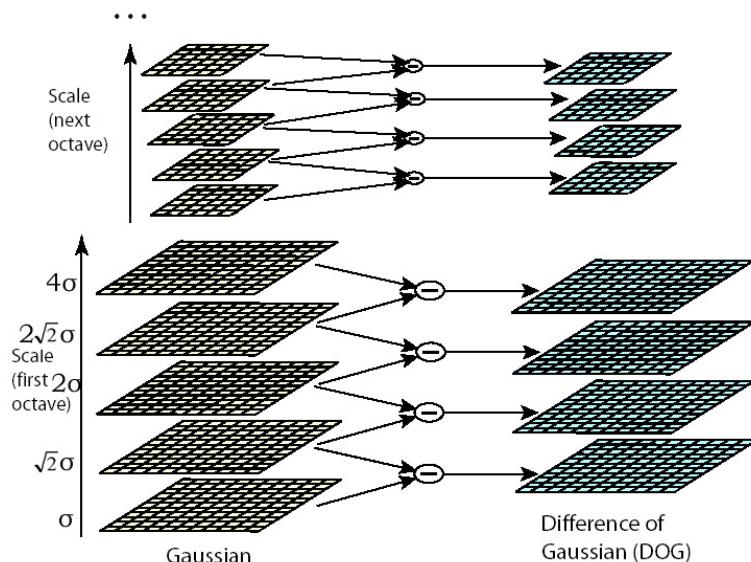
Giá trị hàm DoG được tính xấp xỉ dựa vào giá trị scale-normalized Laplacian of Gaussian ($\sigma^2 \nabla^2 G$) thông qua các phương trình (5)(6)(7)

$$\frac{\partial G}{\partial \sigma} = \sigma \nabla^2 G \quad (8)$$

$$\sigma \nabla^2 G = \frac{\partial G}{\partial \sigma} \approx \frac{G(x, y, k\sigma) - G(x, y, \sigma)}{k\sigma - \sigma} \quad (9)$$

$$G(x, y, k\sigma) - G(x, y, \sigma) \approx (k-1)\sigma^2 \nabla^2 G \quad (10)$$

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

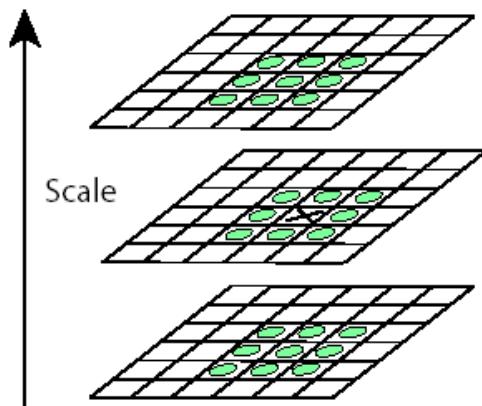


Hình 8. Biểu đồ mô phỏng việc tính toán các DoG ảnh từ các ảnh kề mờ

Các ảnh cuộn được nhóm thành các octave (mỗi octave tương ứng với giá trị gấp đôi của σ). Giá trị của k được chọn sao cho số lượng ảnh mờ (blured images) cho

mỗi octave là cố định. Điều này đảm bảo cho số lượng các ảnh DoG cho mỗi octave không thay đổi.

Các điểm hấp dẫn được xác định là các cực đại hoặc cực tiểu của các ảnh DoG qua các tỉ lệ. Mỗi điểm ảnh trong DoG được so sánh với 8 điểm ảnh láng giềng của nó ở cùng tỉ lệ đó và 9 láng giềng kè ở các tỉ lệ ngay trước và sau nó. Nếu điểm ảnh đó đạt giá trị cực tiểu hoặc cực đại thì sẽ được chọn làm các điểm hấp dẫn ứng viên.



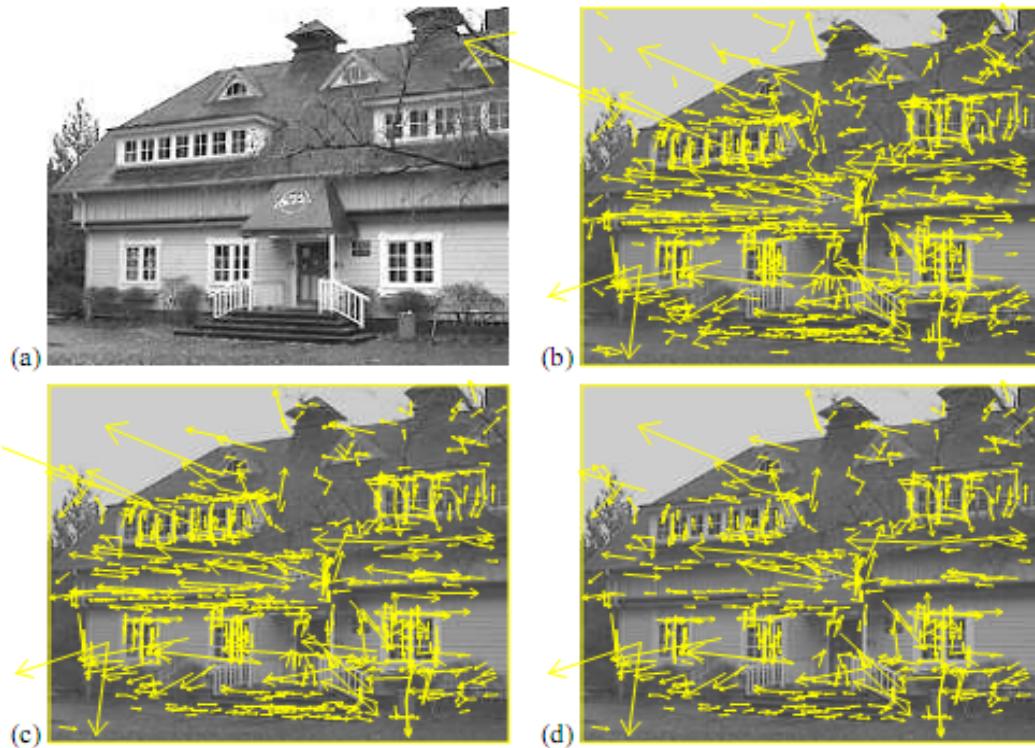
Hình 9. Mỗi điểm ảnh được so sánh với 26 láng giềng của nó

2.5.1.2. Định vị điểm hấp dẫn:

Mỗi điểm hấp dẫn ứng viên sau khi được chọn sẽ được đánh giá xem có được giữ lại hay không:

- Loại bỏ các điểm hấp dẫn có độ tương phản thấp
- Một số điểm hấp dẫn dọc theo các cạnh không giữ được tính ổn định khi ảnh bị nhiễu cũng bị loại bỏ.

Các điểm hấp dẫn còn lại sẽ được xác định hướng.



Hình 10. Quá trình lựa chọn các điểm hấp dẫn

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

2.5.1.3. Xác định hướng cho điểm hấp dẫn:

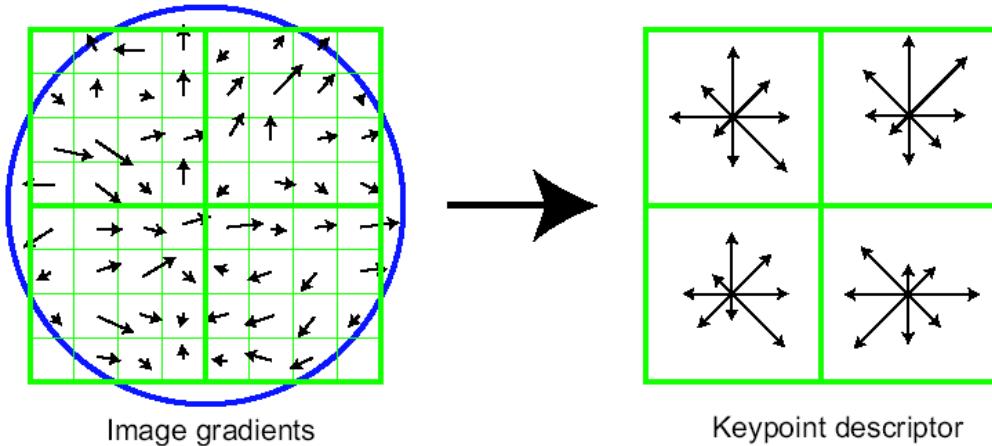
Để xác định hướng cho các điểm hấp dẫn, người ta tính toán biểu đồ hướng Gradient trong vùng lảng giềng của điểm hấp dẫn. Độ lớn và hướng của các điểm hấp dẫn được xác định theo công thức:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (11)$$

$$\theta(x, y) = \tan^{-1}((L(x, y+1) - L(x, y-1)) / (L(x+1, y) - L(x-1, y))) \quad (12)$$

2.5.1.4. Biểu diễn vector cho điểm hấp dẫn

Điểm hấp dẫn sau khi được xác định hướng sẽ được biểu diễn dưới dạng các vector $4 \times 4 \times 8 = 128$ chiều.



Hình 11. Biểu diễn các vector đặc trưng

2.5.2. Độ đo tương đồng cho đặc trưng cục bộ bất biến

Một số độ đo tương đồng cho ảnh sử dụng đặc trưng SIFT như[33] :

- Độ đo Cosin :

$$d(x, y) = \frac{x \cdot y}{\|x\| \cdot \|y\|} \quad (13)$$

- Khoảng cách góc :

$$d(x, y) = \cos^{-1}(x \cdot y) \quad (14)$$

- Độ đo Euclide :

$$d(x, y) = \sqrt{\sum_{i=1}^n |x_i - y_i|^2} \quad (15)$$

- Độ đo Jensen-Shannon divergence :

$$d_{JSD}(H, H') = \sum_{m=1}^M H_m \log \frac{2H_m}{H_m + H'_m} + H'_m \log \frac{2H'_m}{H_m + H'_m} \quad (16)$$

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

2.6. Lựa chọn đặc trưng

Sau khi trích chọn được các đặc trưng nội dung của ảnh, tập các đặc trưng có thể được tối ưu hóa bằng các phương pháp lựa chọn đặc trưng để tăng chất lượng và hiệu quả khi sử dụng các tập đặc trưng.

Một cách tổng quát, lựa chọn đặc trưng là phương pháp giảm thiểu các đặc trưng nhằm chọn ra một tập con các đặc trưng phù hợp trong học máy để xây dựng mô hình

học tốt nhất. Mục đích của lựa chọn đặc trưng là tìm ra không gian con các đặc trưng tối ưu sao cho các tập ảnh “thích hợp” và “không thích hợp” được tách biệt nhất.

Có nhiều phương pháp lựa chọn đặc trưng được đề xuất như: phương pháp tăng khuyếch đại (boosting manner) kết hợp với nền tảng Real Adaboost của Wei Jian và Guihua Er [25]. Mingjing Li[26] đưa ra tiêu chí lựa chọn các đặc trưng là: Mô hình tương phản đặc trưng được tổng quát hóa (Generalized Feature Contrast Model) dựa trên mô hình tương phản đặc trưng (Feature Contrast Model). Một số phương pháp có điển khác như phương pháp dựa vào phân phối (distribution based). Phương pháp dựa vào phân tích biệt thức (Discriminant analysis DA) ví dụ như Phân tích đa biệt thức (Mutiple Discriminant analysis MDA)), phân tích biệt thức không đối xứng (biased Discriminant analysis BDA). Phương pháp tối đa khoảng cách tối thiểu đối xứng trong không gian con (symmetric maximized minimal distance in subspace SMMS)...

Một số phương pháp lựa chọn đặc trưng[23]:

STT	Phương pháp	Mô tả, nhận xét
1	Phương pháp dựa vào phân phối (Distribution based approaches)	<p>Không xét đến yêu cầu về tính bất đối xứng trong hệ thống CBIR.</p> <p>Khó đánh giá phân phối mẫu vì một số mẫu huấn luyện không đặc tả được hết toàn bộ tập dữ liệu. Vì vậy, phương pháp này không thích hợp cho hệ thống tìm kiếm ảnh học online.</p>
2	Phương pháp khuyếch đại thông thường (conventional Boosting method)	<p>Không xét đến yêu cầu về tính bất đối xứng trong hệ thống CBIR.</p> <p>Không được đánh giá tốt vì khả năng tổng quát hóa thấp do tiêu chí lựa chọn đặc trưng dựa trên lỗi huấn luyện.</p>
3	Phương pháp phân tích biệt thức	<p>Phương pháp DA tổng hợp các phân tích biệt thức tuyến tính và giả thiết rằng các ảnh “thích hợp” được nhóm vào với nhau như một cụm.</p> <p>Với những ảnh “không thích hợp”, phương pháp DA giả thiết rằng chúng không nằm trong một phân phối một cụm.</p> <p>Phương pháp MDA giả thiết rằng mỗi ảnh “không</p>

		<p>thích hợp” đến từ một lớp khác nhau.</p> <p>Phương pháp BDA giả thiết rằng mỗi ảnh “không thích hợp” đến từ một số không xác định các lớp.</p> <p>SMMS lựa chọn không gian đặc trưng con trực giao với không gian con kéo dài bằng các mẫu “thích hợp”.</p>
4	Phương BiasMap (BDA hạt nhạt)	Ánh xạ mẫu huấn luyện đến một không gian nhiều chiều hơn để giải quyết vấn đề giả thuyết một cụm.
5	Phương pháp khuyếch đại (Boosting manner)	Tăng các đặc trưng được học thành phân lớp toàn bộ giảm lỗi huấn luyện.

Có nhiều phương pháp để đánh giá kết quả của tập con đặc trưng. Vì vậy, kết quả đối với những mô hình lựa chọn đặc trưng khác nhau là khác nhau. Hai mô hình phổ biến cho lựa chọn đặc trưng là: Mô hình Filter và mô hình Wrapper.

- Mô hình Filter: đánh giá mỗi phần tử bằng một vài tiêu chuẩn hay độ đo nào đó, rồi chọn ra tập con các thuộc tính được đánh giá cao nhất.
- Mô hình Wrapper: Sử dụng một thuật toán tìm kiếm để đánh giá tập con các thuộc tính coi như là một nhóm hơn là một phần tử riêng lẻ. Cốt lõi của mô hình Wrapper là một thuật toán học máy cụ thể. Nó đánh giá độ tốt của những tập con đặc trưng tùy theo độ chính xác học của tập con, điều này xác định thông qua một tiêu chí nào đó.

Tổng kết chương 2

Trong chương 2, khóa luận đã trình bày tóm tắt phương pháp trích chọn các đặc trưng nội dung ảnh(màu sắc, kết cấu, hình dạng và đặc trưng cục bộ SIFT) và một số độ đo tương đồng tương ứng với các đặc trưng. Một số phương pháp lựa chọn đặc trưng để tối ưu hóa tập đặc trưng. Trong chương 3, chúng tôi sẽ trình bày một số công trình nghiên cứu khoa học liên quan đến tìm kiếm ảnh theo nội dung ảnh trích chọn được.

Chương 3. Một số phương pháp tìm kiếm ảnh theo nội dung

3.1. Phương pháp PageRank cho tìm kiếm ảnh sản phẩm

Yushi Jing và cộng sự giới thiệu hệ thống xếp hạng lại các kết quả tìm kiếm hình ảnh của Google dựa trên nội dung của các bức ảnh. Hệ thống xây dựng một đồ thị tương đồng với mỗi đỉnh là một ảnh, các ảnh được liên kết với nhau theo độ tương đồng giữa chúng và áp dụng phương pháp PageRank để xếp hạng lại các ảnh. Hệ thống cho kết quả tốt với 2000 truy vấn về những sản phẩm phổ biến nhất[30].

Hệ thống xây dựng một đồ thị từ tập dữ liệu ảnh và sau đó xếp hạng các ảnh dựa trên các siêu liên kết trực quan (visual hyperlinks) giữa các ảnh. Nhận định trực quan của việc sử dụng các siêu liên kết trực quan này là nếu một người dùng xem một ảnh, thì người đó có thể cũng sẽ quan tâm đến một ảnh khác gần giống với ảnh vừa xem. Đặc biệt, nếu ảnh u có siêu liên kết trực quan đến ảnh v , thì sẽ có một xác suất để người dùng chuyển từ u sang v . Bằng trực giác, ta có thể thấy các ảnh có liên quan tới truy vấn sẽ có nhiều ảnh khác trỏ tới chúng và do đó sẽ được thăm thường xuyên. Các ảnh mà được thăm thường xuyên thường được coi là quan trọng. Hơn nữa, nếu một ảnh v là quan trọng và nó có liên kết tới ảnh w , thì nó sẽ gộp độ quan trọng của nó cho độ quan trọng của w vì bản thân v là quan trọng;

Hạng của một bức ảnh được định nghĩa lại như sau:

$$IR = S^* \times IR \quad (17)$$

Trong đó, S^* là ma trận kè cắt giảm theo cột của S , với $S_{u,v}$ là độ tương đồng giữa 2 ảnh u và v . Bằng cách lặp đi lặp lại phép nhân IR với S^* ta sẽ thu được véc tơ đặc trưng nổi bật (dominant eigenvector) của ma trận S^* .

ImageRank (IR) hội tụ chỉ khi ma trận S^* không tuần hoàn hoặc tối giản. Điều kiện không tuần hoàn thường đúng đối với Web còn điều kiện tối giản thường yêu cầu một đồ thị liên thông mạnh. Do đó, định nghĩa một hệ số hăm d để tạo một đồ thị liên thông mạnh, thỏa mãn điều kiện hội tụ và để làm giảm hạng của các đỉnh, tránh trường hợp một số trang có thứ hạng quá cao.

Với một tập n ảnh, IR được định nghĩa:

$$IR = dS^* \times IR + (1-d)p \quad \text{với} \quad p = \left[\frac{1}{n} \right]_{n \times 1} \quad (18)$$

Một cách trực quan, điều này tạo một xác suất nhỏ cho việc duyệt ngẫu nhiên đến các ảnh trong đồ thị, mặc dù nó có thể không có liên kết tới ảnh hiện tại. Trong thực nghiệm, hệ số hâm d thường được chọn giá trị $d > 0.8$.

Trong hệ thống của mình, Jing và cộng sự đã sử dụng đặc trưng SIFT (2.6) và biểu diễn đặc trưng ảnh dưới dạng biểu đồ hướng đặc trưng. Sau khi biểu diễn ảnh thành các vector đặc trưng tương ứng, độ tương đồng hai ảnh được tính một cách đơn giản bằng số điểm hấp dẫn chung chia cho số điểm hấp dẫn trung bình của hai ảnh.

Hệ thống thử nghiệm với các ảnh trả về từ Google cho 2000 câu truy vấn của những sản phẩm phổ biến nhất. Kết quả cho thấy ở top10 kết quả đầu tiên, tỉ lệ ảnh không phù hợp của hệ thống chỉ là 0.47 trong khi của Google là 2.82 và top 3 của hệ thống là 0.2 so với 0.81 của Google. Xét về hiệu xuất tổng thể trên các truy vấn, có 762 truy vấn của hệ thống chứa ít ảnh không hợp lý hơn so với Google và chỉ 70 truy vấn cho kết quả kém hơn Google.



Hình 12. Ví dụ các ảnh sản phẩm trả về từ hệ thống của Jing

3.2. CueFlik: Một phương pháp xếp hạng lại ảnh dựa trên luật của người dùng

Tìm kiếm ảnh trên web là một nhiệm vụ gặp nhiều khó khăn vì từ khóa thường không đặc tả được hết các đặc trưng trực quan của ảnh. Một số công cụ tìm kiếm phổ biến đã bắt đầu cung cấp các thẻ dựa trên một số đặc điểm cơ bản của ảnh ví dụ như

ảnh đen, trắng, ảnh có chứa khuôn mặt,... Tuy nhiên, phương pháp này còn hạn chế trong việc xác định rõ ràng thẻ mà người dùng mong muốn được sử dụng trong kết quả tập ảnh tìm kiếm từ web. Để giải quyết vấn đề này, James Fogarty và cộng sự đã công bố phương pháp CueFlik[14], một ứng dụng tìm kiếm ảnh trên web, cho phép người dùng tạo nhanh các luật riêng của họ để xếp hạng lại các ảnh dựa trên các đặc trưng trực quan của chúng. Sau đó, người dùng có thể xếp hạng lại bất kỳ kết quả tìm kiếm ảnh nào dựa trên các luật mà họ đã đưa ra. Phương pháp này đã được thử nghiệm, cho phép người dùng tạo nhanh các luật của các khái niệm như: “product photos”, “portraits of people”, “clipart”.

CueFlik kế thừa việc tìm kiếm ảnh dựa vào từ khóa. Tuy nhiên, CueFlik cho phép người dùng sắp xếp lại các ảnh theo các luật được xây dựng từ các đặc trưng trực quan của ảnh. Mỗi luật được định nghĩa như là lớp láng giềng gần nhất, việc tính toán xác định mức độ tương đồng của một ảnh so với các ảnh mẫu dùng để huấn luyện các luật đó. Việc huấn luyện các luật như vậy yêu cầu học một hàm khoảng cách từ các ảnh mẫu cung cấp bởi người dùng.

CueFlik xếp hạng các ảnh được lấy từ truy vấn đến Microsoft’s Live (1000 bức ảnh), Các luật sẽ tính điểm cho các ảnh dựa vào công thức:

$$imageScore(i) = \sum_{r \in ActiveRules} weight_r * score_r(i) \quad (19)$$

Với các weight có giá trị từ -1 đến 1

Active Rules là các luật áp dụng với ảnh đó

Mỗi luật được định nghĩa là lớp láng giềng gần nhất gồm tập các mẫu “tích cực” (positive examples), các mẫu “tiêu cực” (negative examples) và một độ đo khoảng cách. Theo đó, một luật tính điểm cho mỗi bức ảnh dựa theo công thức:

$$score_r(i) = 1 - \frac{\min dist_p}{\min dist_p + \min dist_N} \quad (20)$$

Trong đó:

$score(i)$ có giá trị từ 0 đến 1. $score(i)$ có giá trị 1 khi gần với ảnh mẫu tích cực nhất và bằng 0 khi gần ảnh mẫu tiêu cực nhất.

$\min dist_p$ là khoảng cách đến ảnh mẫu “tích cực” gần nhất, $\min dist_N$ là khoảng cách đến ảnh mẫu “tiêu cực” gần nhất.

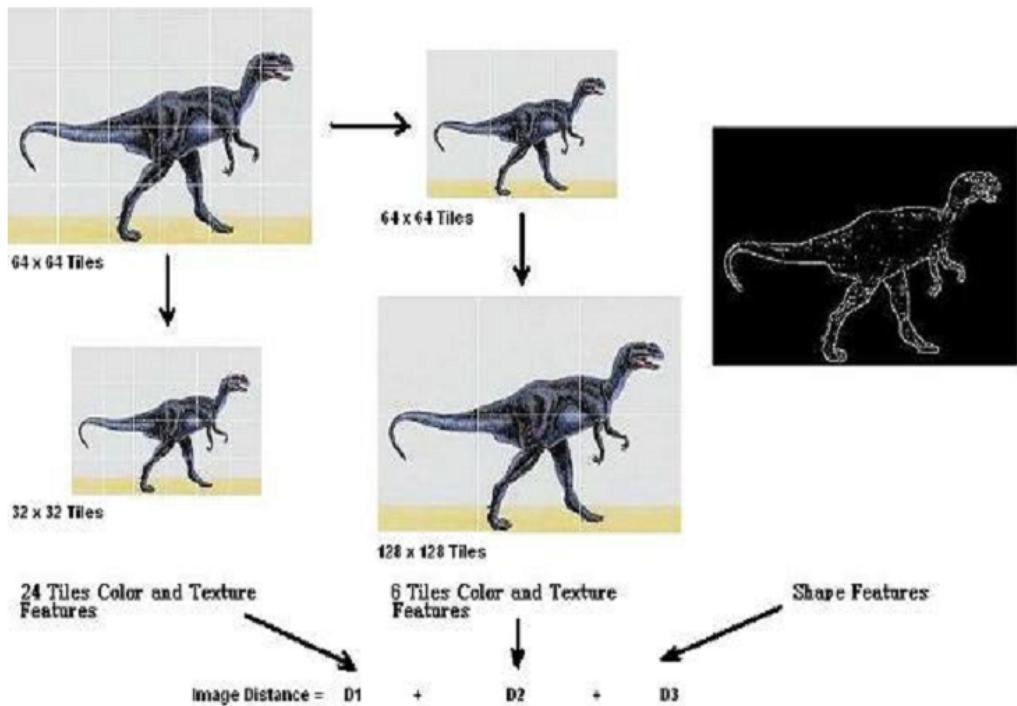
Khoảng cách giữa 2 ảnh i, j là tổng hợp các độ đo khoảng cách được sử dụng.

$$Dis tan ce(i, j) = \sum_{m \in Metrics} weight_m * dis tan ce_m(i, j) \quad (21)$$

CueFlik có thể học được các khoảng cách đo thành phần, sử dụng các độ đo khoảng cách dựa vào biểu đồ màu sắc, độ bão hòa màu, cường độ chiếu sáng của các điểm ảnh, biểu đồ cạnh, biểu đồ hình toàn cục, biểu đồ kết cấu. CueFlik tính toán chúng cho mỗi ảnh và sử dụng để đo khoảng cách giữa các ảnh với nhau. CueFlik học các luật từ các mẫu tích cực và tiêu cực để đưa ra được các luật là tương đồng với bức ảnh hay không?. Việc học các luật này được đưa về việc học các trọng số dựa trên độ đo khoảng cách tương thích nhất với các bức ảnh mẫu cung cấp. Việc học này dựa trên các lý thuyết của Globerson và Roweis [34].

3.3. Phương pháp tìm kiếm ảnh dựa trên màu sắc, hình dạng, kết cấu của ảnh

Màu sắc, kết cấu, hình dạng là những đặc trưng được sử dụng đầu tiên trong các hệ thống tìm kiếm ảnh dựa vào nội dung. P.S. Hirematch và Jagadeesh Pujari [20] đã trình bày phương pháp kết nối cả ba đặc trưng màu sắc, kết cấu và hình dạng để đạt hiệu quả cao trong tìm kiếm hình ảnh. Trong phương pháp này, ảnh và phần hỗ trợ của nó được chia thành các ô vuông (tiles) cùng kích thước và không chồng lấp lên nhau. Những đặc trưng được rút ra từ những biểu đồ xảy ra đồng thời có điều kiện giữa các ô vuông của ảnh và ô vuông của các thành phần hỗ trợ tương ứng được coi như là những đặc trưng cục bộ của màu sắc và kết cấu. Một đề xuất tích hợp nguyên tắc độ ưu tiên cao nhất cho cái tương đồng nhất (most similar highest priority principle) và dạng đồ thị 2 phần (bipartite graph) sử dụng các ô vuông của truy vấn và của ảnh đích, được sử dụng để đối sánh giữa 2 ảnh. Đặc trưng theo hình dạng được trích rút nhờ việc tính toán cạnh của ảnh dựa vào Gradient Vector Flow. Việc kết nối đặc trưng màu sắc, kết cấu giữa ảnh và thành phần hỗ trợ của nó cộng thêm các đặc trưng về hình dạng đã đưa ra được một tập các đặc trưng mạnh mẽ trong tìm kiếm ảnh theo nội dung .



Hình 13. Tổng quan về mô hình của hệ thống tìm kiếm theo màu sắc, kết cấu và hình dạng

3.3.1. Lưới

Mỗi ảnh được phân thành 24 ô vuông (4×6 hoặc 6×4 như hình 12) không trùng lặp nhau. Các ô vuông này sẽ được xử lý như đặc trưng màu sắc và kết cấu cục bộ của ảnh. Những đặc trưng rút ra từ biểu đồ xảy ra đồng thời có điều kiện giữa các ô vuông của ảnh và ô vuông của các thành phần hỗ trợ tương ứng được sử dụng cho độ tương đồng về màu sắc và kết cấu. Với mỗi ảnh (kích thước 256×384 hoặc 384×256) được phân thành vùng 6×4 hoặc 4×6 , mỗi ô vuông sẽ có kích thước là 64×64 , sau đó ảnh lại được phân rã thêm một bậc thành có kích thước $M/2 \times N/2$ với M và N là số hàng và cột của ảnh gốc. Việc phân chia này giúp chúng ta nắm bắt được các thông tin ảnh khác nhau trong quá trình giải quyết.

3.3.2. Tích hợp các đối sánh ảnh

Trong phương pháp này, một ô vuông từ ảnh truy vấn được phép đối sánh với bất kỳ ô vuông nào của ảnh đích. Tuy nhiên, một ô vuông có thể chỉ tham gia chỉ một lần trong quá trình đối sánh. Thuật toán sử dụng ma trận kè để giảm thiểu quá trình tính toán cho độ ưu tiên cao nhất cho độ tương đồng lớn nhất. Ở đây, ma trận khoảng cách được tính như một ma trận kè, khoảng cách tối thiểu d_{ij} được tính trong ma trận này, khoảng cách này được ghi lại và hàng tương ứng với ô vuông i và cột tương ứng với ô vuông j được đánh dấu lại (thay thế bằng một giá trị cao như: 999).

Điều này tránh việc ô vuông i của ảnh truy vấn và ô vuông j của ảnh đích tiếp tục tham gia trong việc xử lý đối sánh. Khoảng cách giữa ô vuông i và những ô vuông khác của ảnh đích và khoảng cách của ô vuông j với những ô vuông khác của ảnh truy vấn được bỏ qua. Quá trình này tiếp tục cho đến khi tất cả các ô vuông được đổi sánh. Khoảng cách đổi sánh tối thiểu giữa các ảnh được định nghĩa bởi công thức:

$$D_{qt} = \sum_{i=1,n} \sum_{j=1,n} d_{ij} \quad (22)$$

Trong đó: d_{ij} là khoảng cách đổi sánh tốt nhất giữa ô vuông i của ảnh truy vấn và ô vuông j của ảnh đích. D_{qt} là khoảng cách giữa ảnh q và ảnh t .

3.3.3. Hình dạng:

Thông tin về hình dạng thu được từ khuôn khổ các cạnh của ảnh cấp độ xám tương đương. Nhóm tác giả sử dụng Gradient Vector Flow để thu thập đặc trưng cạnh của ảnh.

Giải thuật tính toán cạnh của ảnh:

- Đọc ảnh và chuyển đổi ảnh sang ảnh cấp xám
- Làm mờ ảnh sử dụng bộ lọc Gaussian
- Tính toán các biểu đồ Gradient của ảnh bị làm mờ
- Tính toán Gradient Vector Flow (GVF)
- Lọc ra các phản hồi cạnh mạnh sử dụng $k\delta$ với δ là độ lệch tiêu chuẩn của GVF
- Hội tụ vào các điểm ảnh cạnh thỏa mãn điều kiện cân bằng sinh ra các ảnh cạnh.

3.4. Phương pháp tìm kiếm ảnh dựa vào nội dung sử dụng các phân vùng ảnh như mẫu truy vấn

Một phương pháp phổ biến để tìm kiếm ảnh dựa vào nội dung là sử dụng ảnh mẫu làm truy vấn. Awang Iskandar James và cộng sự trình bày phương pháp tìm kiếm ảnh sử dụng các mẫu truy vấn là các phân vùng ảnh[4]. Nhóm tác giả so sánh hiệu quả khi sử dụng các đặc trưng trích chọn từ toàn bộ bức ảnh làm truy vấn với sử dụng đặc trưng trích chọn từ phân vùng đơn và nhiều phân vùng. Hiệu quả của bài toán khi sử dụng thêm đặc trưng hình dạng so với việc phân lớp sử dụng giải thuật học máy cũng được nhắc đến trong bài.

Hai phương pháp được sử dụng rộng rãi để việc miêu tả và biểu diễn hình dạng là dựa vào phân vùng và đường biên trên. Trong phương pháp dựa vào phân vùng, các đặc trưng được trích xuất từ toàn vùng. Phương pháp dựa vào đường biên trên biểu diễn các hình dạng bằng cách lấy mẫu thô rời rạc chu vi của nó. Biểu diễn hình dạng dựa vào đường biên bao gồm các vành đai, khoảng cách Haus-dorff, biểu diễn Fourier,...

Trong bài báo, tác giả kết hợp cả 2 phương pháp dựa vào phân vùng và dựa vào đường biên trên của trích xuất các đặc trưng hình dạng của các vùng quan tâm: Area, mean, circularity và boundary. Area là tổng số điểm ảnh có trong một vùng, mean là giá trị cấp xám trung bình trong một vùng được tính bằng giá trị sám của tất cả các điểm ảnh chia cho tổng số điểm ảnh.

Tập ảnh dữ liệu được thu thập từ các tập truyện tranh Groat. Với mỗi bức ảnh, sẽ xác định và trích xuất ra 2 phân vùng. Bài báo dùng 30 phân vùng được trích xuất để truy vấn ảnh dựa vào các mẫu phân vùng sử dụng đơn và đa vùng và huấn luyện dữ liệu cho giải thuật học máy

Kết luận tác giả đã chỉ ra rằng, việc sử dụng phân vùng đơn làm mẫu truy vấn hiệu quả hơn so với việc sử dụng toàn bộ ảnh làm truy vấn và sử dụng đa phân vùng lại vượt trội hơn so với sử dụng phân vùng đơn. Việc sử dụng kết hợp truyền tính trọng số bằng nhau đơn giản hơn nhưng mang lại hiệu quả tương đương so với sử dụng giải thuật học máy.

Tổng kết chương 3

Chương 3 khóa luận đã tóm tắt một số công trình nghiên cứu khoa học liên quan đến việc tìm kiếm và xếp hạng ảnh theo nội dung bao gồm: phương pháp pageRank cho tìm kiếm ảnh sản phẩm [30], phương pháp CueFlik xếp hạng lại ảnh dựa trên các luật người dùng [14], phương pháp tìm kiếm ảnh dựa vào nội dung kết hợp các thuộc tính màu sắc, kết cấu, hình dạng[4] và phương pháp tìm kiếm ảnh với mẫu truy vấn là các phân vùng của ảnh [20]. Trong chương 4, khóa luận sẽ giới thiệu phương pháp lựa chọn đặc trưng của ảnh và mô hình tìm kiếm k lảng giềng gần nhất .

Chương 4. Mô hình k lảng giềng gần nhất sử dụng bộ lượng tử hóa

4.1. Đặt vấn đề

Bài toán tìm kiếm K lảng giềng gần nhất là một bài toán đơn giản và rất phổ biến. Bài toán có thể được định nghĩa như sau : Cho một tập n phần tử, xây dựng một cấu trúc dữ liệu sao cho khi đưa vào một truy vấn, hệ thống trả về K phần tử gần nhất với truy vấn. Các phần tử dữ liệu thường được biểu diễn trong không gian Oclit nhiều chiều. Tìm kiếm K lảng giềng gần nhất là bài toán quan trọng và được áp dụng trong nhiều lĩnh vực như nén dữ liệu, tìm kiếm thông tin, học máy, thống kê và phân tích dữ liệu, tìm kiếm ảnh và video,...

Trong khóa luận này, bài toán tìm kiếm K lảng giềng gần nhất được hiểu là từ ảnh dữ liệu đầu vào hệ thống sẽ tìm ra và trả về K ảnh tương đồng nhất với ảnh đầu vào từ cơ sở dữ liệu. Trong quá trình tính toán độ tương đồng, ảnh thường được biểu diễn dưới dạng các vector đặc trưng nhiều chiều. Việc tính toán độ tương đồng giữa các ảnh được quy về tính khoảng cách giữa các vector đặc trưng sử dụng độ đo Oclit. Tuy nhiên, việc tính toán khoảng cách giữa các vector đặc trưng nhiều chiều này tốn nhiều thời gian và tài nguyên máy. Nhiều phương pháp đánh chỉ mục đa chiều phổ biến như KD-tree hay những hướng kỹ thuật khác đã được đề xuất để giảm thời gian tìm kiếm. Tuy nhiên các phương pháp này vẫn chưa đạt được kết quả như mong muốn.

Khóa luận trình bày phương pháp lựa chọn các đặc trưng và tìm kiếm lảng giềng gần nhất dựa trên mô hình tìm kiếm sử dụng lượng tử hóa tích của Hervé Jégou và cộng sự [12] kết hợp với độ đo tương đồng về khoảng cách giữa các vector đặc trưng.

4.2. Cơ sở lý thuyết

4.2.1. Các ký hiệu và khái niệm

4.2.1.1. Lượng tử hóa vector

Về mặt hình thức, một bộ lượng tử hóa vector (quantization) là một hàm q ánh xạ một vector $x \in R^D$ thành một vector $q(x) \in C = \{c_i; i \in I\}$ với tập chỉ mục I được giả định là hữu hạn: $I = \{0, \dots, k-1\}$, D là số chiều của không gian vector đang xét. Các giá trị c_i gọi là “*trọng tâm (centroids)*” và tập các giá trị C gọi là *codebook*. Chúng ta giả thiết rằng các chỉ số nhận giá trị là các số nguyên liên tiếp từ 0 đến $k-1$.

Tập các vector V_i được ánh xạ tới một chỉ số i được gọi là một “ô Voronoi” (Voronoi cell), được định nghĩa:

$$V_i \triangleq \{x \in R^D : q(x) = c_i\} \quad (23)$$

k “ô” của bộ lượng tử xác định một phân vùng của R^D . Theo định nghĩa, tất cả các vector nằm trong cùng một “ô” V_i được đặt trong cùng một trọng tâm c_i . Chất lượng của một bộ lượng tử hóa được đo bằng giá trị lỗi bình phương trung bình (Mean Square error MSE) giữa vector đầu vào x và giá trị sau khi được lượng tử hóa của nó $q(x)$:

$$MSE(q) = E_x [d(\tilde{x}, x)^2] = \int p(x) d(q(x), x)^2 dx \quad (24)$$

Với $d(x, y) = \|x - y\|$ là khoảng cách Oclit giữa x và y , và $p(x)$ là hàm phân phối xác xuất tương ứng với biến ngẫu nhiên chung X .

Bộ lượng tử hóa tối ưu khi nó mãn hai thuộc tính còn gọi là điều kiện Lloyd:

- Thứ nhất: các vector x phải được lượng tử hóa tới trọng tâm codebook gần nhất của nó, thể hiện qua khoảng cách Oclit:

$$q(x) = \arg \min_{i \in I} d(x, c_i) \quad (25)$$

- Thứ hai: Giá trị biến đổi phải là một vector nằm trong ô Voronoi:

$$c_i = E_x [x | i] = \int_{V_i} p(x) x \quad (26)$$

Bộ lượng tử hóa Lloyd tương ứng với giải thuật phân cụm K-Means. Phương pháp của Hervé Jégou và cộng sự sử dụng sê học bộ lượng tử sử dụng K-Means.

4.2.1.2. Bộ lượng tử hóa tích

Vector đầu vào x được chia thành m vector con phân biệt nhau $\{x_1, x_2, \dots, x_j\}$ $1 \leq j \leq m$ với số chiều $D^* = D/m$, với D là bội số của m . Những vector con này được được lượng tử hóa một cách riêng biệt sử dụng m bộ lượng tử hóa khác nhau. Vì vậy, vector đầu vào x được ánh xạ như sau:

$$\underbrace{x_1, \dots, x_{D^*}}_{u_1(x)}, \dots, \underbrace{x_{D-D^*+1}, \dots, x_D}_{u_m(x)} \rightarrow q_1(u_1(x)), \dots, q_m(u_m(x)) \quad (27)$$

Với q_j là bộ lượng tử thấp (low-complexity quantizer) tương thích với vector con thứ j . Với bộ lượng tử q_j tương ứng với tập chỉ mục I_j , codebook C_j và giá trị biến đổi tương ứng $c_{j,i}$.

Giá trị biến đổi của bộ lượng tử tích được xác định bởi một phần tử của tập chỉ mục tích $I = I_1 \times I_2 \times \dots \times I_m$. Vì vậy, codebook được định nghĩa như là tích Đécác:

$$C = C_1 \times C_2 \times \dots \times C_m \quad (28)$$

Trọng tâm của tập này chính là các trọng tâm của các bộ lượng tử hóa con ghép lại với nhau. Giả sử rằng, tất cả các bộ lượng tử hóa con có k^* hữu hạn các giá trị biến đổi, như vậy, tổng số trọng tâm sẽ là:

$$k = (k^*)^m \quad (29)$$

Với $m = D$, thì tất cả các thành phần của vector x được lượng tử hóa một cách riêng biệt, với trường hợp $m=1$, bộ lượng tử hóa tích trở thành codebook k-means thông thường. Ưu điểm của bộ lượng tử hóa tích là tập trọng tâm lớn được sinh ra từ tập các trọng tâm nhỏ hơn tương ứng với bộ lượng tử con. Trong khi học bộ lượng tử con sử dụng giải thuật Lloyd có sử dụng một tập giới hạn các vector nhưng codebook vẫn thích nghi được với sự phân bố dữ liệu biểu diễn.

4.2.2. Tìm kiếm sử dụng lượng tử hóa

Việc tìm kiếm lảng giàn nhất phụ thuộc duy nhất vào khoảng cách giữa vector truy vấn và vector của cơ sở dữ liệu, hay tương đương với bình phương khoảng cách. Phần này trình bày phương pháp tìm kiếm sử dụng lượng tử hóa (Searching with quantization)[12] để so sánh khoảng cách giữa hai vector dựa vào chỉ số lượng tử hóa của chúng.

4.2.2.1. Tính toán khoảng cách sử dụng mã được lượng tử hóa

Hervé Jégou và cộng sự[12] sử dụng hai phương pháp để tính khoảng cách Oclit xấp xỉ giữa hai vector truy vấn x và vector cơ sở dữ liệu y : phương pháp tính toán đối xứng (Symmetric distance computation SDC) và phương pháp tính toán bất đối xứng. (Asymmetric distance computation ADC)

- Phương pháp tính toán đối xứng : Vector x và y được biểu diễn thành các trọng tâm riêng biệt $q(x)$ và $q(y)$. khoảng cách $d(x,y)$ được tính xấp xỉ bằng khoảng cách giữa $q(x)$ và $q(y)$:

$$\tilde{d}(x, y) = d(q(x), q(y)) = \sqrt{\sum_j d(q_j(x), q_j(y))^2} \quad (30)$$

Với khoảng cách $d(q_j(x), q_j(y))^2$ được lấy từ bảng tìm kiếm liên kết với bộ lượng tử hóa con thứ j

- Phương pháp tính toán bất đối xứng : Vector truy vấn x được giữ nguyên, vector trong cơ sở dữ liệu được lượng tử hóa thành $q(y)$, khoảng cách giữa x và y được tính xấp xỉ bằng :

$$\tilde{d}(x, y) = d(x, q(y)) = \sqrt{\sum_j d(u_j(x), q_j(u_j(y)))^2} \quad (31)$$

Với $d(u_j(x), c_{j,i})^2$: $j=1, \dots, m$ và $i=1, \dots, k^*$ được tính trước để tìm kiếm.

4.2.3. Tìm kiếm không toàn bộ

Trong phần trên trình bày phương pháp tìm kiếm láng giềng gần nhất sử dụng bộ lượng tử hóa với hai phương pháp tính toán đối xứng và tính toán bất đối xứng. Phương pháp này sử dụng m phép tính cộng cho mỗi lần tính khoảng cách. Tuy nhiên, việc tìm kiếm vẫn phải diễn ra trên toàn bộ tập vector đặc trưng. Đối với các hệ thống đa truy vấn, và sử dụng tập vector đặc trưng cục bộ cho để biểu diễn ảnh, thì số lượng vector đặc trưng là rất nhiều, việc tìm kiếm tồn rất nhiều bộ nhớ và thời gian. Hervé Jégou và cộng sự giới thiệu phương pháp tính toán khoảng cách bất đối xứng dựa trên chỉ mục ngược (inverted file asymmetric distance computation (IVFADC)) để tránh việc tìm kiếm trên toàn bộ tập vector đặc trưng.

4.2.3.1. Lượng tử hóa thô, danh sách chỉ mục ngược

Bộ lượng tử hóa thô là một mảng các danh sách : L_1, \dots, L_k . Nếu Y là tập vector cần đánh chỉ mục, danh sách L_i sẽ tương ứng với trọng tâm c_i của q_c lưu trữ tập: $\{y \in Y : q_c(y) = c_i\}$. Với một danh sách chỉ mục ngược, truy vấn được gán vào w vị trí chỉ mục phù hợp với w láng giềng gần nhất của x trong codebook q_c . Tất cả các danh sách chỉ mục ngược phù hợp được xem xét.

4.2.3.2. Xác định cục bộ mã lượng tử tích

Gọi $r(x)$ là vector dư (residual vector) được lượng tử hóa từ trọng tâm $q_c(x)$ tương ứng với vector x , $r(x)$ được định nghĩa :

$$r(x) = x - q_c(x) \quad (32)$$

Gọi q_p là bộ lượng tử hóa tích để mã hóa vector dư $r(x)$, vector x được biểu diễn bởi tuple $(q_c(x), q_p(r(x)))$, với $q_p(r(x))$ được lưu trong danh sách chỉ mục ngược tương ứng với x . Khoảng cách giữa vector truy vấn x và vector trong cơ sở dữ liệu y được tính xấp xỉ:

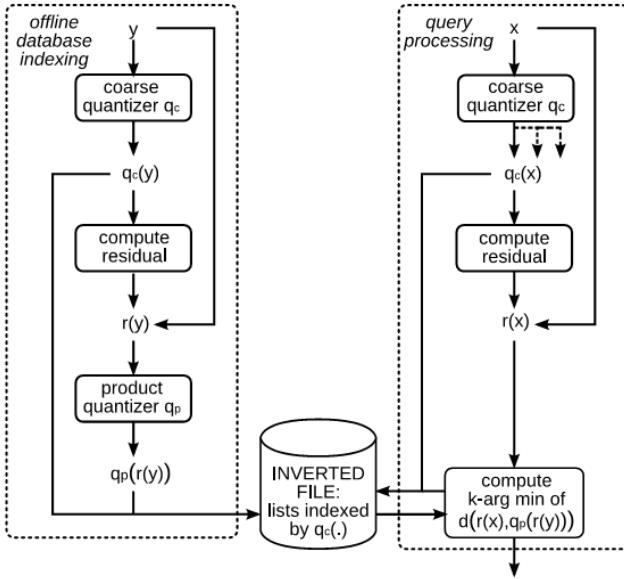
$$\ddot{d}(x, y)^2 = \sum_j d(u_j(x - q_c(x)), q_{p_i}(u_j(y - q_c(y))))^2 \quad (33)$$

Với q_{p_i} là bộ lượng tử hóa con thứ i. Cũng giống như phương pháp ADC, với mỗi bộ lượng tử hóa con q_{p_i} , khoảng cách giữa vector dư $u_j(x - q_c(x))$ và tất cả các trọng tâm $c_{j,i}$ của q_{p_i} được tính toán sơ bộ và lưu trữ lại.

4.2.3.3. Cấu trúc chỉ mục và thuật toán tìm kiếm

- Đánh chỉ mục vector y :
 1. Lượng tử hóa y thành $q_c(y)$
 2. Tính toán vector dư $r(y) = y - q_c(y)$
 3. Lượng tử hóa vector dư $r(y)$ thành $q_p(r(y))$ để thực hiện lượng tử hóa tích bằng cách gán $u_j(y)$ thành $q_j(u_j(y))$ với $j = 1, \dots, m$
 4. Lưu trữ các vector đặc trưng và mã biểu diễn chỉ mục lượng tử hóa tích trong một mục của danh sách chỉ mục ngược.
- Tìm kiếm các láng giềng gần nhất của truy vấn x :
 1. Lượng tử hóa x thành w láng giềng gần nhất trong codebook q_c
 2. Tính toán bình phương khoảng cách $d(u_j(r(x)), c_{j,i})^2$ cho mỗi lượng tử hóa con j và mỗi trọng tâm $c_{j,i}$ của nó.
 3. Tính bình phương khoảng cách giữa $r(x)$ và tất cả các vector chỉ mục trong danh sách chỉ mục ngược.
 4. Lựa chọn k lân cận gần nhất của x dựa vào đánh giá khoảng cách sử dụng giải thuật Maxheap

Mô hình hệ thống IVFADC :



Hình 14. Mô hình hệ thống IVFADC

Hệ thống bên trái: chèn một vector vào danh sách chỉ mục ngược; hệ thống bên phải: tìm kiếm k láng giềng gần nhất.

4.3. Mô hình bài toán

Trong phần (4.2), khóa luận trình bày hệ thống tìm kiếm k láng giềng gần nhất sử dụng tính khoảng cách bất đối xứng trong danh sách chỉ mục ngược (IVFADC) của Hervé Jégou và cộng sự [12]. Mô hình bài toán được xây dựng dựa trên cách tính toán khoảng cách bất đối xứng của hệ thống này kết hợp với độ đo tương đồng về khoảng cách giữa các vector đặc trưng.

4.3.1. Trích chọn đặc trưng ảnh

Đặc trưng cục bộ bất biến SIFT bất biến với việc thay đổi tỉ lệ ảnh, quay ảnh, đôi khi là thay đổi điểm nhìn và thêm nhiều ảnh hay thay đổi cường độ chiếu sáng của ảnh. Các đặc trưng SIFT này thường được sử dụng trong nhận dạng và tìm kiếm đối tượng[17] Yushi Jing cũng đã dùng đặc trưng SIFT của ảnh trong nghiên cứu về tìm kiếm ảnh sản phẩm sử dụng phương pháp PageRank[30].

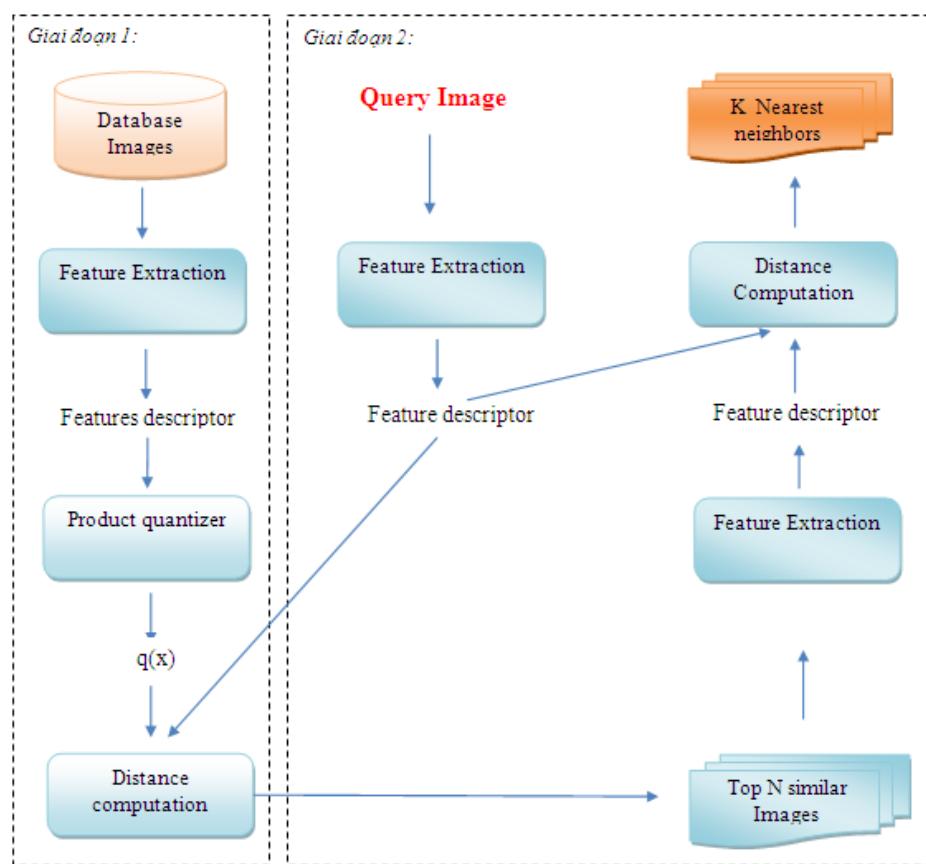
Khóa luận sử dụng đặc trưng SIFT trong bài toán tìm kiếm K láng giềng gần nhất và ứng dụng trong tìm kiếm ảnh sản phẩm. Mỗi ảnh được đặc trưng bởi các vector đặc trưng SIFT 128 chiều.

Quá trình và phương pháp trích chọn các đặc trưng SIFT này đã được khóa luận trình bày chi tiết trong phần 2.6

4.3.2. Tìm kiếm K láng giềng gần nhất

Sau khi trích chọn đặc trưng ảnh, khóa luận đưa ra mô hình tìm kiếm k láng giềng gần nhất dựa trên đặc trưng vừa trích chọn được. Mô hình này dựa trên phương pháp tìm kiếm k láng giềng gần nhất sử dụng bộ lượng tử hóa của Hervé Jégou sử dụng phương pháp ADC và kết hợp thêm độ đo về khoảng cách Oclit giữa các vector đặc trưng.

Mô hình giải quyết bài toán :



Hình 15. Mô hình giải quyết bài toán

Mô hình bài toán gồm 2 giai đoạn chính

Giai đoạn 1-Tìm N ảnh tương đồng với ảnh truy vấn : Giai đoạn này tiến hành việc trích chọn các vector đặc trưng của ảnh truy vấn và ảnh trong cơ sở dữ liệu (vector đặc trưng SIFT), sau đó tìm top N ảnh tương đồng với ảnh truy vấn từ tập ảnh trong cơ sở

dữ liệu theo phương pháp tìm kiếm sử dụng bộ lượng tử hóa với phương pháp tính toán khoảng cách bất đối xứng được trình bày trong phần 4.2. Các vector trong cơ sở dữ liệu được lượng tử hóa trong khi tập vector truy vấn được giữ nguyên. Khoảng cách giữa các vector truy vấn và vector trong cơ sở dữ liệu được tính theo công thức (30). Tập N ảnh tương đồng nhất được trả về theo độ đo khoảng cách giữa các vector truy vấn và các vector cơ sở dữ liệu. Tập N ảnh này là đầu vào cho giai đoạn 2.

Giai đoạn 2 – Tìm K láng giềng gần nhất với ảnh truy vấn: Sau khi tiến hành trích chọn các đặc trưng từ tập N ảnh tương đồng trả về từ giai đoạn 1, sẽ tính toán độ tương đồng giữa ảnh truy vấn và từng ảnh trả về dựa trên độ đo Oclit giữa các vector đặc trưng của ảnh. Khoảng cách Oclit giữa 2 vector đặc trưng x và y được tính :

$$d(x, y) = \sqrt{\sum_{i=1}^n |x_i - y_i|^2} \quad (34)$$

Tập K láng giềng gần nhất với ảnh truy vấn được trả về dựa trên độ đo tương đồng này. Ảnh gần nhất là ảnh có độ khoảng cách giữa các vector đặc trưng với ảnh truy vấn ngắn nhất.

Tổng kết chương 4

Chương 4 khóa luận đã trình bày phương pháp tìm kiếm k láng giềng gần nhất sử dụng lượng tử hóa của Hervé Jégou và cộng sự [12], đồng thời đưa ra mô hình bài toán tìm kiếm k láng giềng gần nhất dựa theo mô hình trên sử dụng phương pháp tính khoảng cách bất đối xứng (ADC) kết hợp với độ đo tương đồng về khoảng cách giữa các vector đặc trưng. Trong chương 5, khóa luận trình bày mô hình thử nghiệm bài toán, các kết quả đạt được và những nhận xét, đánh giá về kết quả thực nghiệm.

Chương 5. Thực nghiệm và đánh giá

Dựa vào cơ sở lý thuyết và mô hình đề xuất trong chương 4, khóa luận tiến hành thực nghiệm việc trích chọn các vector đặc trưng SIFT từ ảnh truy vấn và ảnh trong cơ sở dữ liệu, áp dụng mô hình k láng giềng gần nhất với tập đặc trưng vừa trích chọn được để tìm ra tập k ảnh gần nhất với ảnh truy vấn.

- Đầu vào của hệ thống : Một ảnh truy vấn do người dùng nhập vào
- Đầu ra của hệ thống : Tập k ảnh gần nhất với ảnh truy vấn

5.1. Môi trường và các công cụ sử dụng cho thực nghiệm

- *Cấu hình phần cứng*

Bảng 1. Cấu hình phần cứng sử dụng trong thực nghiệm

Thành phần	Chỉ số
CPU	1 Pentium IV 3.06 GHz
RAM	1 GB
OS	WindowsXP Service Pack 2
Bộ nhớ ngoài	80GB

- *Công cụ phần mềm sử dụng*

Bảng 2. Công cụ phần mềm sử dụng trong thực nghiệm

STT	Tên phần mềm	Tác giả	Nguồn
1	Matlab R2009b		http://www.mathworks.com/products/

- **Một số thư viện sử dụng**

Bảng 3. Một số thư viện sử dụng trong thực nghiệm

STT	Tên phần mềm	Tác giả	Nguồn
1	SiftDemoV4	David Lowe	http://people.cs.ubc.ca/~lowe/keypoints/
2	Pqsearch_matlab	Hervé Jegou, Matthij Douze	http://lear.inrialpes.fr/~jegou/ann.php
3	Kmeans_fast.tar	Hervé Jegou, Matthij Douze	http://lear.inrialpes.fr/~jegou/ann.php#matlla
4	FlickrSearcher	Nguyễn Cẩm Tú	http://www.dais.is.tohoku.ac.jp/%7Enamtu/software.htm

Ngoài các công cụ trên, chúng tôi còn tiến hành xây dựng các module xử lý dựa trên ngôn ngữ Matlab bao gồm các file sau:

- Similar_Euclidean: tính toán khoảng cách Euclidean giữa tập vector đặc trưng
- Pq_test: Kết hợp các module con, tìm kiếm và trả về k láng giềng gần nhất với truy vấn từ tập dữ liệu.

5.2. Xây dựng tập dữ liệu ảnh

Trong khóa luận này, chúng tôi thực nghiệm với tập dữ liệu ảnh liên quan đến sản phẩm, sử dụng kết quả từ Flickr và Google product Search.

- **Ảnh truy vấn:** Do người dùng nhập vào. Trong khóa luận này, chúng tôi chú trọng đến một số truy vấn có sự nhập nhằng giữa nội dung ảnh và văn bản đi kèm ảnh.
- **Tập ảnh cơ sở dữ liệu:** Với mỗi truy vấn, tập dữ liệu ảnh gồm 30 ảnh được trộn từ tập các ảnh lấy từ Google Product Search và Flickr. Chúng tôi tiến hành thu thập các ảnh bằng truy vấn text tương ứng với ảnh truy vấn từ Google Product Search. Sau đó bổ sung nhiều bằng các ảnh thu thập được từ Flickr theo truy vấn text tương ứng với ảnh truy vấn.

- Tập ảnh huấn luyện: Trong quá trình lượng tử hóa vector, cần một tập dữ liệu ảnh huấn luyện để xác định các tham số trong bộ lượng tử hóa con (4.2). Tập ảnh huấn luyện gồm 20 ảnh khác nhau được lấy từ kết quả trả về của Google Product Search.
- Tập ảnh trả về: Gồm k ảnh gần giống nhất với ảnh truy vấn. Các ảnh được sắp xếp giảm dần theo mức độ gần với truy vấn. Chúng tôi thử nghiệm với giá trị $k=10$.

5.3. Quy trình, phương pháp thực nghiệm

Quy trình thực nghiệm được tiến hành như sau:

Thực hiện truy vấn: Người dùng nhập vào truy vấn dưới dạng tên và đường dẫn đầy đủ đến ảnh truy vấn.

Trích chọn đặc trưng và tìm kiếm k ảnh tương đồng nhất: Quá trình này trải qua hai giai đoạn chính:

Giai đoạn 1: Giai đoạn này tiến hành trích chọn các đặc trưng của ảnh truy vấn và ảnh trong cơ sở dữ liệu sử dụng bộ công cụ SiftDemoV4[39] và trả về N ảnh tương đồng nhất sử dụng lượng tử hóa với phương pháp ADC. Tập các đặc trưng SIFT sau khi được trích chọn được lưu dưới dạng ma trận $nx128$ với n là số vector đặc trưng. Sau đó, các đặc trưng này được lượng tử hóa sử dụng bộ công cụ pqsearch_matlab[40] và tính khoảng cách giữa các vector sử dụng phương pháp ADC. N ảnh tương đồng nhất được trả về dựa trên độ đo khoảng cách này. Trong đó, ảnh gần nhất là ảnh có khoảng cách nhỏ nhất đến ảnh truy vấn. N ảnh này được lấy làm đầu vào cho giai đoạn 2.

Giai đoạn 2: Giai đoạn này nhận đầu vào là N ($N=20$) ảnh tương đồng trả về từ giai đoạn 1. Sử dụng các vector đặc trưng của các ảnh này đã được trích xuất trong giai đoạn 1 để tính toán khoảng cách giữa cách Oclit giữa các vector đặc trưng này với vector đặc trưng của ảnh truy vấn. K ảnh gần nhất với ảnh truy vấn được trả về theo khoảng cách được tính, trong đó ảnh gần nhất là ảnh có khoảng cách ngắn nhất đến truy vấn.

5.4. Kết quả thực nghiệm

Chúng tôi sử dụng độ chính xác trung bình (Average Precision) [1] để đánh giá kết quả xếp hạng của hệ thống.

Giả sử ta có 5 đối tượng là: a, b, c, d, e

Trong đó **a, b, c** là các đối tượng phù hợp và **d, e** là các đối tượng không phù hợp.

Một xếp hạng của các đối tượng cần đánh giá là: **c, a, d, b, e**

Độ chính xác trung bình được định nghĩa như sau:

$$AP = \frac{\sum_{k=1}^n P @ K \times I(K)}{\sum_{j=1}^n I(J)} \quad (35)$$

Trong đó:

n là số đối tượng được xét.

$P @ K = \frac{Match @ K}{K}$ ($Match @ K$ = số các đối tượng phù hợp ở K vị trí đầu tiên)

$I(K) = 1$ nếu đối tượng ở vị trí K , ngược lại $I(K) = 0$

Ví dụ: $P@1 = 1/1$, $P@2 = 2/2$, $P@3 = 2/3$, $P@4 = 3/4$. Thì độ chính xác trung bình là:

$$AP = \frac{\frac{1}{1} \times 1 + \frac{2}{2} \times 1 + \frac{3}{4} \times 1}{3} = 0.92 \quad (36)$$

Ngoài ra chúng tôi còn sử dụng Mean Average Precision (MAP) để đánh giá hệ thống.

Giá trị trung bình trên m xếp hạng:

$$MAP = \frac{\sum_{i=1}^m AP_i}{m} \quad (37)$$

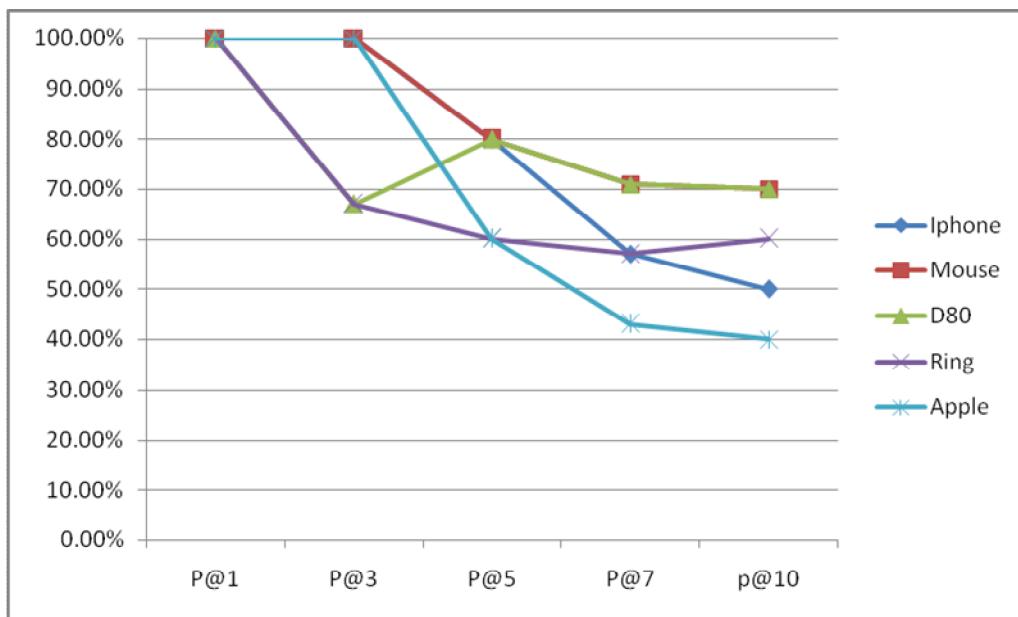
Chúng tôi thử nghiệm hệ thống với 10 truy vấn trên bộ dữ liệu thử nghiệm và đánh giá kết quả trả về đối với 10 kết quả trả về đầu tiên.

Kết quả về độ chính xác trung bình cho 10 ảnh trả về đầu tiên của 10 truy vấn:

Bảng 4. Kết quả độ chính xác trung bình của 10 truy vấn

STT	Truy vấn	AP
1	Apple	0.875
2	Coca cola	0.747
3	D80	0.804
4	CD-Rom	0.737
5	Iphone	0.885
6	Mouse	0.869
7	Nokia N97	0.883
8	Cooker	0.748
9	Ring	0.746
10	Printer	0.753

Bảng 5. Độ chính xác mức k của một số truy vấn



Từ các kết quả thống kê trên, chúng tôi tính toán được độ chính xác trung bình đối với 10 truy vấn của hệ thống là: MAP=0.804. Có thể thấy rằng, độ chính xác trung bình đối với 10 truy vấn của hệ thống là khá cao, ví dụ Iphone là 0.885, Nokia N97 là 0.883. Đặc biệt, theo khảo sát của thực nghiệm, hệ thống cho kết quả rất chính xác với kết quả đầu tiên trả về. Độ chính xác mức 1 của các truy vấn thường là 1. Đối với tập dữ liệu có chứa ảnh giống hệt so với ảnh truy vấn, thì khả năng ảnh thứ nhất được trả

về giống hệt với ảnh truy vấn là rất cao. Trong 10 truy vấn thực nghiệm thì 8 truy vấn trả về ảnh đầu tiên giống hệt so với ảnh truy vấn. Ví dụ tốp 10 kết quả đầu tiên với truy vấn Iphone:

Ảnh truy vấn	1	2	3	4	5
	6	7	8	9	10

Hình 16. 10 kết quả trả về đầu tiên của hệ thống với truy vấn Iphone

Tổng kết chương 5

Chương 5, Khóa luận trình bày về mô hình thực nghiệm của hệ thống. Các công cụ, phần mềm, mã nguồn hệ thống sử dụng. Khóa luận cũng trình bày quá trình tiến hành thực nghiệm, các kết quả đạt được của hệ thống với 10 truy vấn và một số nhận xét về độ chính xác của hệ thống đạt được. Từ những kết quả ban đầu đạt được đó cho thấy tính khả thi và đúng đắn của hệ thống.

Kết luận

Lượng ảnh số trên web tăng lên một cách nhanh chóng đòi hỏi phải có các hệ thống tìm kiếm ảnh hiệu quả và tiện lợi. Tuy các công cụ tìm kiếm ảnh theo văn bản đi kèm ảnh cho phép người dùng tìm kiếm ảnh với thời gian đáp ứng nhanh nhưng chưa giải quyết được vấn đề nhập nhằng giữa văn bản đi kèm và nội dung hiển thị của ảnh trả về. Khóa luận tập trung nghiên cứu một số phương pháp trích chọn đặc trưng ảnh và xây dựng hệ thống tìm kiếm láng giềng gần nhất với ảnh truy vấn dựa theo nội dung ảnh.

Khóa luận đã đạt được những kết quả sau :

Khóa luận đã tìm hiểu các đặc trưng của ảnh bao gồm đặc trưng văn bản đi kèm ảnh và đặc trưng nội dung ảnh. Đồng thời, tìm hiểu các phương pháp trích chọn đặc trưng nội dung ảnh cũng như một số độ đo tương đồng tương ứng với các đặc trưng.

Khóa luận cũng đi tìm hiểu một số phương pháp tìm kiếm và xếp hạng ảnh theo nội dung ảnh. Dựa theo mô hình tìm kiếm láng giềng sử dụng bộ lượng tử hóa của Hervé Jégou và cộng sự [12], khóa luận đưa ra mô hình tìm kiếm láng giềng gần nhất sử dụng bộ lượng tử hóa và phương pháp tính khoảng cách bất đối xứng kết hợp với độ đo tương đồng giữa các vector đặc trưng.

Khóa luận tiến hành thử nghiệm mô hình với 10 truy vấn. Kết quả có độ chính xác trung bình là 80.4% cho 10 kết quả trả về đầu tiên của hệ thống đối với 10 truy vấn. Từ những kết quả bước đầu cho thấy tính khả quan và đúng đắn của mô hình.

Một số vấn đề hạn chế và hướng nghiên cứu tiếp theo :

Do hạn chế về mặt thời gian và kiến thức sẵn có, khóa luận mới chỉ dừng lại ở mức thử nghiệm của mô hình trên đặc trưng SIFT của ảnh với tập dữ liệu nhỏ và ít truy vấn.

Trong thời gian tới, chúng tôi sẽ tiến hành thử nghiệm mô hình với các đặc trưng nội dung khác của ảnh. Đồng thời, mở rộng tập dữ liệu và truy vấn trên nhiều miền khác nhau để xây dựng mô hình tìm kiếm láng giềng gần nhất theo nội dung ảnh hoàn thiện.

Tài liệu tham khảo

Tài liệu tiếng Việt :

- [1]. Nguyễn Thu Trang (2009). Học xếp hạng trong tính hạng đối tượng và phân cụm tài liệu, *Luận văn Thạc sỹ, Trường Đại Học Công Nghệ*.

Tài liệu tiếng Anh :

- [2]. Alex Holub, Pierre Moreels, Pietro Perona (2008). Unsupervised clustering for google searches of celebrity images, *IEEE International Conference on Automatic Face and Gesture Recognition*, 2008
- [3]. Alexandre Noma, Ana Beatriz V. Graciano, Luís Augusto Consularo, Roberto M. Cesar, Isabelle Bloch (2008). A New Algorithm for Interactive Structural Image Segmentation, *CoRR abs/0805.1854*
- [4]. D. N. F. Awang Iskandar James A. Thom S. M. M. Tahaghoghi (2008). Content-based Image Retrieval Using Image Regions as Query Examples. *CRPIT Volume 75- Database technologies*.
- [5]. Deselaers T, Keysers D, Ney H (2005). Discriminative Training for Object Recognition using Image Patches. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 05)*. 2:157-162 San Diego, CA; 2005.
- [6]. Florian Schroff, Antonio Criminisi, Andrew Zisserman (2007). Harvesting Image Databases from the Web, *ICCV 2007*: 1-8
- [7]. G. Shakhnarovich, T. Darrell, and P. Indyk(2006). Nearest-Neighbor Methods in Learning and Vision: Theory and Practice, *MIT Press*, March 2006 ISBN 0-262-19547-X
- [8]. Hao Zhang Alexander C. Berg Michael Maire Jitendra Malik (2007). SVM-KNN: Discriminative Nearest Neighbor Classification for Visual Category Recognition. *Computer Science Division, EECS Department Univ. of California, Berkeley, CA 94720*
- [9]. Herve' Jégou, Matthijs Douze, and Cordelia Schmid (2008). Hamming embedding and weak geometric consistency for large scale image search. *The 10th European Conference on Computer Vision: Part I*.
- [10]. Hervé Jégou, Matthijs Douze, Cordelia Schmid(2009). Product quantization for nearest neighbor search, *IEEE Transactions on Pattern Analysis & Machine Intelligence – 2010*

- [11]. Hervé Jégou, Matthijs Douze, Cordelia Schmid(2008). Recent Advances in Large Scale Image Search, *ETVC 2008*: 305-326. (2008)
- [12]. Hervé Jégou, Matthijs Douze, Cordelia Schmid(2009). Searching with quantization: approximate nearest neighbor search using short codes and distance estimators. *Technical Report RR-7020, INRIA*
- [13]. J. Friedman, J. L. Bentley, and R. A. Finkel(). An algorithm for finding best matches in logarithmic expected time. *ACM Transaction on Mathematical Software, vol. 3, no. 3, pp. 209–226, 1977*
- [14]. James Fogarty, Desney S. Tan, Ashish Kapoor, Simon A. J. Winder(2008). CueFlik: interactive concept learning in image search. *The twenty-sixth annual SIGCHI conference on Human factors in computing system*
- [15]. Jun Zhao, Guo-Yin Wang, Hong Tang, Hua Li – the study on technologies for feature selection. *The 1st Int. Nat. Conf. On Machine Learning and Cybernetics (ICMLC02), 2002, Beijing, 689-693.*
- [16]. Kamarul Hawari Ghazali(2007). Feature Extraction technique using SIFT keypoints descriptors. *The International Conference on Electrical and Engineering and Informatics Institut technology Bandung, Indonesia, june 17-19, 2007*
- [17]. Lowe David(2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision 2004;60(2):91–110.*
- [18]. Michele Saad (2008). Low-Level Color and Texture Feature Extraction for Content-Based Image Retrieval . *EE 381K: Multi-Dimensional. Digital Signal Processing*
- [19]. Mitsuru Ambai Denso(2009). Multiclass VisualRank: Image Ranking Method in Clustered Subsets Based on Visual Features. *SIGIR'09, July 19–23, 2009, Boston, Massachusetts, USA.*
- [20]. P.S. Hiremath, Jagadeesh Pujari (2007). Content base image retrieval base on color, texture and shape feature using Image and its complement. *IJCSS, International journal of computer science and security, vol 1, issue 4, Dec 2007,pp. 25-35.*
- [21]. Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z. Wang (2008): Image Retrieval: Ideas, Influences, and Trends of the New Age. *ACM Computing Surveys, 40 (2).*
- [22]. Shuhui Wang, Qingming Huang, Shuqiang Jiang(2009). Visual ContextRank for Web Image Re-ranking. *The First ACM workshop on Large-scale multimedia retrieval and mining*

- [23]. Tee Cheng Siew(2008). Feature selection for content-based image retrieval using statistical discriminant analysis. *PhD thesis Faculty of Computer Science and Information System Universiti Teknologi Malaysia. 2008*
- [24]. Thomas Deselaers¹, Daniel Keysers², and Hermann Ney¹: Features for Image Retrieval: An Experimental Comparison. *Information Retrieval vol 11, issue 2, Kluwer Academic Publishers Hingham, MA, USA*
- [25]. W. Jiang, G. Er, Q. Dai and J. Gu. (2006). Similarity-Based Online Feature Selection In Content-Based Image Retrieval. *IEEE Trans. Image Processing, 15 (3), pp.702-712.*
- [26]. W. Jiang. M. Li, H. Zhang, J. Gu. (2004. Online feature Selection based on Generalized Feature Contrast Model. *IEEE International Conference on Multimedia and Expo(ICME). pp. 1995-1998*
- [27]. Yossi Rubner, an Puzicha,Carlo Tomasi and Joachim M. Buhmann Empirical: Evaluation of Dissimilarity Measures for Color and Texture. *Computer Vision and Image Understanding, vol 84, issue 1. Elsivier Science Ins.*
- [28]. Yushi Jing, Shumeet Baluja, Henry A. Rowley(2007). Canonical image selection from the web, *CIVR 2007: 280-287*
- [29]. Yushi Jing(2008). VisualRank: Applying PageRank to Large-Scale Image Search. *IEEE Trans Pattern Anal Mach Intell.*
- [30]. Yushi Jing(2008). PageRank for images products search. *Reafered Track: Rich media, April 21-25, 2008. Beijing, China.*
- [31]. V. Shiv Naga Prasad. A.G. Faheema, Subrata Rakshi(2002). Feature Selection in Example-Based Image Retrieval Systems. *Indian Conference on Vision Graphics and Image Processing*
- [32]. C. V. Jawahar, P. J. Narayanan, and S. Rakshit(2000). A flexible scheme for representation, matching, and retrieval of images. *ICVGIP 2000, pages 271–277. Allied Publishers Ltd., 2000.*
- [33]. Mohamed Aly(2006). Face Recognition using SIFT Features. *AlyCNS186 Term Project Winter*
- [34]. Globerson, A. and Roweis, S. (2005). Metric Learning by Collapsing Classes. *Conference on Neural Information Processing Systems (NIPS 2005), 451-458.*

Website tham khảo :

- [35]. Website: <http://www.searchenginejournal.com/7-similarity-based-image-search-engines/8265/>
- [36]. Website: <http://www.thongtincongnghe.com/article/9703>
- [37]. Website: <http://www.hori.ecei.tohoku.ac.jp>
- [38]. Website: <http://www.flickr.com/search/>
- [39]. Website: <http://people.cs.ubc.ca/~lowe/keypoints/>
- [40]. Website: <http://lear.inrialpes.fr/~jegou/ann.php>
- [41]. Website: http://www.freeseamlesstextures.com/texture_gallery/index.htm