

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

NGUYỄN THỊ DƯƠNG

**GÁN NHÃN ẢNH TỰ ĐỘNG SỬ DỤNG KẾT HỢP
NHIỀU NGUỒN THÔNG TIN**

LUẬN VĂN THẠC SĨ ĐIỆN TỬ - VIỄN THÔNG

HÀ NỘI - 2017

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

NGUYỄN THỊ DƯƠNG

**GÁN NHÃN ẢNH TỰ ĐỘNG SỬ DỤNG KẾT HỢP
NHIỀU NGUỒN THÔNG TIN**

Ngành: Điện tử truyền thông

Chuyên ngành: Kỹ thuật điện tử

Mã số: 60530203

LUẬN VĂN THẠC SĨ ĐIỆN TỬ- VIỄN THÔNG

NGƯỜI HƯỚNG DẪN KHOA HỌC: TS. LÊ VŨ HÀ

HÀ NỘI - 2017

LỜI CẢM ƠN

Đầu tiên tôi xin chân thành cảm ơn đến Ban Giám Hiệu và tập thể thầy cô trong khoa Điện Tử - Viễn Thông, Trường Đại Học Công Nghệ , Đại Học Quốc Gia Hà Nội đã tổ chức giảng dạy và tạo môi trường thuận lợi cho tôi được học tập và nghiên cứu chuyên sâu về lĩnh vực Điện tử viễn thông.

Xin chân thành cảm ơn các anh chị đồng nghiệp trong cơ quan công tác đã tạo điều kiện thuận lợi, đóng góp những ý kiến quý báu. Xin cảm ơn bạn bè đã chỉ bảo tôi trong suốt quá trình học tập và hoàn thành luận văn này.

Đặc biệt xin gửi lời cảm ơn sâu sắc đến Thầy hướng dẫn khoa học **TS. Lê Vũ Hà** và **TS. Nguyễn Hồng Thịnh** đã tận tình hướng dẫn, chỉ bảo tôi trong suốt quá trình nghiên cứu và thực hiện luận văn này.

Cuối cùng, tôi xin biết ơn gia đình đã tạo điều kiện thuận lợi cho tôi yên tâm trong suốt thời gian học tập cũng như thời gian nghiên cứu và hoàn thành luận văn cao học.

Nghiên cứu này được tài trợ bởi trung tâm nghiên cứu Châu Á, Đại học Quốc Gia Hà Nội trong đề tài mã số CA.15.10A

Hà Nội, ngày 05 tháng 9 năm 2017

Tác giả luận văn

Nguyễn Thị Dương

LỜI CAM ĐOAN

Tôi xin cam đoan những nội dung trong luận văn này là do tôi thực hiện dưới sự hướng dẫn của **TS. Lê Vũ Hà**.

Các số liệu, kết quả trình bày trong luận văn là trung thực. Những tư liệu được sử dụng trong luận văn có nguồn gốc và trích dẫn rõ ràng, đầy đủ.

Mọi sao chép không hợp lệ, vi phạm quy chế đào tạo, hay gian trá tôi xin chịu hoàn toàn trách nhiệm.

Tác giả luận văn

Nguyễn Thị Dương

MỤC LỤC

PHẦN MỞ ĐẦU	1
1. Tính cấp thiết của luận văn	1
2. Mục tiêu của luận văn	2
3. Các đóng góp của luận văn	2
4. Bố cục của luận văn	3
Chương 1. TỔNG QUAN VỀ PHƯƠNG PHÁP GÁN NHÃN TỰ ĐỘNG CHO ẢNH.....	4
1.1. Tổng quan.....	4
1.2. Các nghiên cứu trong lĩnh vực gần đây	7
1.3. Bài toán phân loại và gán nhãn ảnh tự động.....	8
1.4. Biểu diễn ảnh dưới dạng các vector đặc trưng (Low level features).....	10
1.4.1. Tại sao phải biểu diễn ảnh	10
1.4.2. Đặc trưng của ảnh.....	11
1.4.2.1. Đặc trưng toàn cục.....	12
1.4.2.2. Đặc trưng cục bộ.....	12
1.4.3. Các vector đặc trưng.....	12
1.4.3.1. Lược đồ màu CH.....	12
1.4.3.2. Đặc trưng tương quan màu CORR	13
1.4.3.3. Lược đồ hệ số góc EDH.....	14
1.4.3.4. Cấu trúc sóng con WT	15
1.4.3.5. Mô-men màu CM55.....	16
1.4.3.6. Phương pháp túi từ điển BOW	16
1.5. Các nguồn thông tin phụ trợ.....	18
1.6. Phương pháp huấn luyện dùng để phân loại và gán nhãn.....	20
1.6.1. Phương pháp K - láng giềng gần nhất (K-Nearest Neighbors).....	20

1.6.2. Phương pháp SVM - Support Vector Machine	21
1.6.3. Đánh giá	23
Chương 2. THUẬT TOÁN GÁN NHÃN ẢNH CHỈ SỬ DỤNG CÁC VECTOR ĐẶC TRƯNG	25
2.1. Ảnh và chú giải (Tags).....	25
2.2. Giới thiệu cách tạo ra 81 nhãn trong tập dữ liệu NUS-WIDE.....	26
2.3. Biểu diễn ảnh dưới dạng vector đặc trưng.....	27
2.4. Thuật toán K-NN.....	29
2.4.1. Phân tích dữ liệu thực nghiệm với K-NN.....	29
2.4.2. Kết quả recall của K-NN.....	31
2.5. Thuật toán SVM.....	32
2.5.1. Phân tích dữ liệu thực nghiệm với SVM.....	32
2.5.2. Kết quả recall của SVM.....	33
Chương 3: THUẬT TOÁN SỬ DỤNG KẾT HỢP THÔNG TIN PHỤ TRỢ.....	35
3.1. Sử dụng thông tin chú giải	35
3.1.1. Phân tích mối quan hệ giữa chú giải và nhãn.....	35
3.1.2. Thực nghiệm sử dụng thông tin chú giải của người dùng trên tập kiểm thử.....	37
3.2. Sử dụng vector đặc trưng và chú giải.....	38
3.2.1. Dựa trên thông tin chú giải của ảnh kiểm thử.....	38
3.2.2. Dựa trên thông tin chú giải của ảnh huấn luyện.....	39
KẾT LUẬN	42
TÀI LIỆU THAM KHẢO.....	44
PHỤ LỤC :	

DANH MỤC HÌNH ẢNH

Hình 1. 1 Một hệ thống CBIR điển hình [Manal and Nordin 2009].	4
Hình 1. 2 Sơ đồ khối của một khung công việc gán nhãn ảnh tự động.	5
Hình 1. 3 Phương pháp tổng quát cách thức thực hiện bài toán gán nhãn ảnh sử dụng các nguồn thông tin khác nhau [Li et al. 2016].	7
Hình 1. 4 Ví dụ tìm kiếm ảnh với từ khóa “mountain” trên Flickr.com.	9
Hình 1. 5 Ví dụ minh họa không thể dùng lược đồ màu để phân biệt con bò và con ngựa trên thảo nguyên	11
Hình 1. 6 Minh họa hai ảnh giống nhau nhưng góc nhìn khác nhau	11
Hình 1. 7 Minh họa hai ảnh có đặc trưng tương quan màu giống nhau	14
Hình 1. 8 Minh họa phương pháp túi từ điển [Fei et al. 2005]	17
Hình 1. 9 Túi từ - miêu tả đối tượng với biểu đồ của các từ xuất hiện[Fei et al. 2005].	18
Hình 1. 10 Ví dụ về ảnh và các tag người dùng gán cho ảnh	18
Hình 1. 11 Thông tin về các thông số của máy ảnh và EXIF của ảnh	19
Hình 1. 12 Ví dụ phân lớp cho X bằng thuật toán K-NN	21
Hình 1. 13 Ví dụ phân lớp giữa hai tập hợp bóng Xanh - Đỏ sử dụng SVM.	23
Hình 2. 1 81 nhãn trong tập dữ liệu của NUS-WIDE	28
Hình 2. 2 Số hình ảnh có liên quan đến 81 nhãn của NUS-WIDE	29
Hình 2. 3 Sơ đồ tóm tắt các bước thực hiện quá trình phân loại và gán nhãn cho ảnh kiểm thử sử dụng K-NN	30
Hình 2. 4 Biểu đồ so sánh kết quả recall của 6 đặc trưng và kết hợp các đặc trưng với $K = 50$ và số nhãn gán cho ảnh (N) tăng dần từ 1 đến 10	31
Hình 2. 5 Sơ đồ tóm tắt các bước thực hiện quá trình phân loại và gán nhãn cho ảnh kiểm thử sử dụng SVM	33
Hình 2. 6 Biểu đồ so sánh kết quả recall của 6 đặc trưng sử dụng thuật toán SVM.	34

Hình 3. 1 Biểu đồ Recall khi sử dụng thông tin chú giải của người dùng trên tập kiểm thử.....	37
Hình 3. 2 Sơ đồ khôi quá trình kết hợp đặc trưng ảnh và chú giải của ảnh kiểm thử.....	38
Hình 3. 3 Biểu đồ so sánh kết quả recall của đặc trưng ảnh và đặc trưng kết hợp với chú giải của ảnh kiểm thử.....	39
Hình 3. 4 Sơ đồ khôi dựa trên thông tin chú giải của ảnh huấn luyện theo cách a.....	40
Hình 3. 5 Biểu đồ so sánh kết quả recall của đặc trưng ảnh và đặc trưng ảnh kết hợp với chú giải của ảnh huấn luyện theo cách a.....	40
Hình 3. 6 Sơ đồ khôi dựa trên thông tin chú giải của ảnh huấn luyện theo cách b.....	41
Hình 3. 7 Biểu đồ so sánh kết quả recall của đặc trưng ảnh và đặc trưng ảnh kết hợp với chú giải của ảnh huấn luyện theo cách b.....	41

DANH MỤC BẢNG

Bảng 1. 1 So sánh các kỹ thuật gán nhãn.....	6
Bảng 1. 2 So sánh ưu điểm và nhược điểm của các kỹ thuật gán nhãn.....	6
Bảng 2. 1 Chú giải phổ biến nhất trong tập NUS-WIDE	26
Bảng 2. 2 Kết quả recall của các đặc trưng khi sử dụng thuật toán K-NN.....	31
Bảng 2. 3 Kết quả recall của các đặc trưng khi sử dụng thuật toán SVM.....	34
Bảng 3. 1 Ví dụ một số kết quả khảo sát mối liên hệ giữa chú giải (tag) và nhãn (concept)	36

PHẦN MỞ ĐẦU

1. Tính cấp thiết của luận văn

Những năm gần đây, sự phát triển của khoa học và công nghệ đặc biệt là các thiết bị chụp ảnh và video cũng như của mạng Internet dẫn đến sự bùng nổ thông tin đa phương tiện. Lượng lớn thông tin hình ảnh, video với sự đa dạng chủ đề được lưu trữ chia sẻ và truy cập trên mạng. Điều này cũng đồng nghĩa với yêu cầu bức thiết trong vấn đề lưu trữ, quản lý và truy cập dữ liệu. Gán nhãn ảnh tự động đã và đang trở thành chủ đề nóng và thách thức lớn trong xử lý thông tin và khoa học máy tính. Mục tiêu chính của kỹ thuật này là thiết lập quá trình mà hệ thống máy tính tự động gán cho mỗi đối tượng trong ảnh một vài nhãn thông tin, căn cứ vào đó ta có thể dễ dàng tìm kiếm, phân loại và truy xuất ảnh.

Hiện nay, các kỹ thuật được sử dụng để phân loại và truy xuất ảnh thường sử dụng thông tin văn bản (text) từ nguồn thông tin phụ trợ (metadata) của dữ liệu ví dụ như nhãn, chú giải (tag), bình luận (comment) của người dùng (ví dụ như Google tìm kiếm văn bản). Tuy nhiên không phải lúc nào các thông tin này cũng có sẵn, cũng chính xác, đặc biệt lại phụ thuộc lớn vào quan điểm của người dùng cũng như ngôn ngữ mà người dùng sử dụng. Hơn nữa việc mô tả bức ảnh bằng một vài từ khóa không phải công việc dễ dàng. Việc gán nhãn ảnh bằng tay còn thực sự khó khăn tốn kém và trở lên bất khả thi khi số lượng ảnh tăng lên rất nhanh chóng như hiện nay.

Xu hướng hiện tại là gán nhãn ảnh một cách tự động sử dụng trực tiếp nội dung ảnh (image content), gán các nhãn cho ảnh đó, qua đó giúp dễ dàng quản lý ảnh cũng như tìm kiếm ảnh. Vấn đề này vẫn luôn nhận được sự quan tâm của các nhà nghiên cứu với hi vọng có thể cải thiện hơn độ chính xác, tốc độ xử lý, sự đa dạng và độ khó của các tập ảnh. Xu hướng chung của các nghiên cứu này là tìm ra các đặc trưng tốt hơn trong việc đặc tả ảnh, qua đó cải thiện được độ chính xác trong quá trình nhận dạng và gán nhãn ảnh. Một xu hướng nữa cũng được quan tâm là sử dụng khai thác triệt để tất cả các thông tin có thể có từ ảnh

(metadata, ví dụ như thời gian chụp, địa điểm chụp, tham số của máy ảnh, chú giải, bình luận của người dùng...) để cải tiến giải thuật. Trong đề tài này tôi cũng đi theo hai xu hướng trên, khai thác việc sử dụng đồng thời nhiều loại đặc trưng trong mô tả ảnh và thông tin phụ trợ có được từ internet gắn liền với ảnh đó để tăng độ chính xác trong quá trình nhận dạng và gán nhãn hình ảnh.

2. Mục tiêu của luận văn

Mục tiêu của luận văn là :

- (i) Nghiên cứu việc sử dụng kết hợp nhiều loại đặc trưng ảnh (low level image features) để mô tả và biểu diễn ảnh.
- (ii) Sử dụng kết hợp các nguồn thông tin phụ trợ như thông tin văn bản (nhãn, chú giải...), nhằm cải thiện được chất lượng/ hiệu quả của quá trình gán nhãn ảnh.

Ngoài ra, chúng tôi cũng đưa ra các khảo sát đánh giá hiệu quả sử dụng kết hợp nhiều thông tin, độ phức tạp cũng như mức độ cải tiến của thuật toán mới.

3. Các đóng góp của luận văn

Trong luận văn này, tôi nghiên cứu đề xuất phương pháp phân loại và gán nhãn tự động cho ảnh sử dụng các vector đặc trưng để mô tả ảnh. Ý tưởng quan trọng là gán nhãn ảnh bằng cách so sánh ảnh cần gán nhãn với các ảnh đã biết nhãn trong tập huấn luyện. Việc so sánh này không thể thực hiện trực tiếp trên hai ảnh bởi vì hình ảnh có thể khác nhau về kích thước, vị trí đối tượng trong ảnh, độ sáng tối, góc chụp vv ...

Mỗi loại đặc trưng đều có ưu điểm, nhược điểm riêng nên tôi tiến hành thực nghiệm với từng loại đặc trưng riêng để đánh giá đặc trưng nào cho kết quả tốt nhất, đặc trưng nào cho kết quả xấu nhất và kết hợp giữa các đặc trưng để biết được mức độ cải tiến so với việc chỉ sử dụng từng đặc trưng riêng biệt. Ngoài ra chúng tôi còn tiến hành việc phân tích và sử dụng thông tin phụ trợ kết hợp với các đặc trưng trên để giảm bớt thời gian và tăng độ chính xác của việc gán nhãn tự động cho ảnh.

4. Bố cục của luận văn

Luận văn được chia làm ba chương :

Chương 1: Tổng quan về phương pháp gán nhãn tự động cho ảnh. Chương này đưa ra hướng tiếp cận của bài toán phân loại và gán nhãn. Trình bày về sáu loại đặc trưng (**Low-level features**) dùng để biểu diễn ảnh như: Lược đồ màu CH (**64-D color histogram**); Đặc trưng tương quan màu CORR (**144-D color auto-correlogram**); Lược đồ hệ số góc EDH (**73-D edge direction histogram**); Cấu trúc sóng con WT (**128-D wavelet texture**); Mô men màu CM55 (**225-D block-wise color moments**) và phương pháp túi từ điển Bag of Word BOW (**500-D bag of visual words**) cũng như nguồn thông tin phụ trợ được sử dụng cho việc gán nhãn. Đồng thời cũng trình bày về hai phương pháp sử dụng để phân loại và gán nhãn đó là KNN và SVM. Và giới thiệu tham số dùng để đánh giá cho kết quả của thuật toán là Recall.

Chương 2: Thuật toán gán nhãn ảnh chỉ sử dụng các vector đặc trưng. Trong chương này sẽ trình bày về tập dữ liệu của NUS-WIDE sử dụng trong phần thực nghiệm gồm có tập ảnh huấn luyện (training), tập ảnh kiểm thử (testing), 81 nhãn (concept) và 1000 chú giải (tags) lấy từ thông tin phụ trợ của ảnh và đưa ra các bước tiến hành thực nghiệm và kết quả của thuật toán khi sử dụng các đặc trưng riêng biệt và kết hợp các đặc trưng với nhau.

Chương 3: Thuật toán sử dụng kết hợp thông tin phụ trợ. Chương này thực hiện việc kết hợp giữa các đặc trưng và các chú giải để cải thiện độ phức tạp và đưa ra kết quả cải tiến của thuật toán.

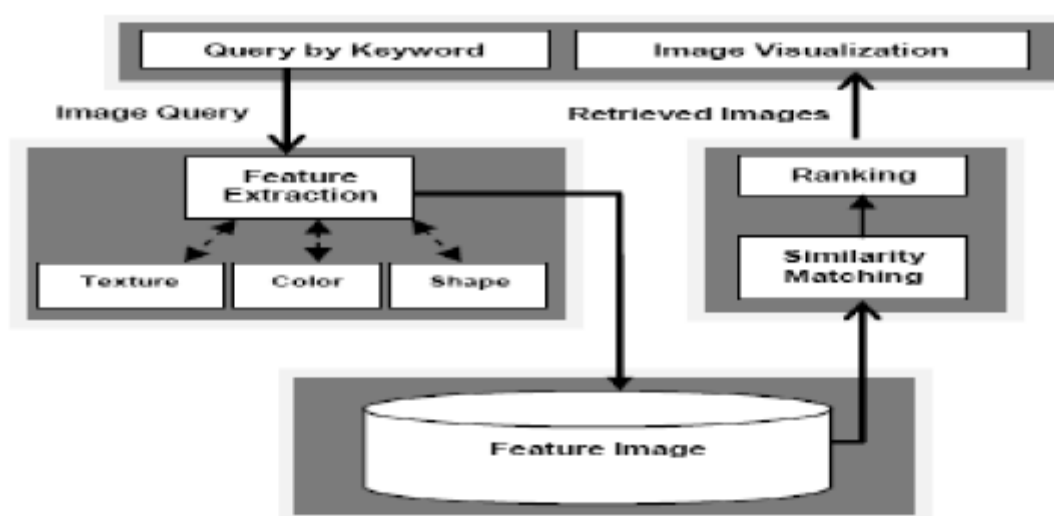
Cuối cùng, phần kết luận tóm lược các kết quả đã đạt được và đề xuất các nghiên cứu trong tương lai.

Chương 1. TỔNG QUAN VỀ PHƯƠNG PHÁP GÁN NHÃN TỰ ĐỘNG CHO ẢNH

1.1. Tổng quan

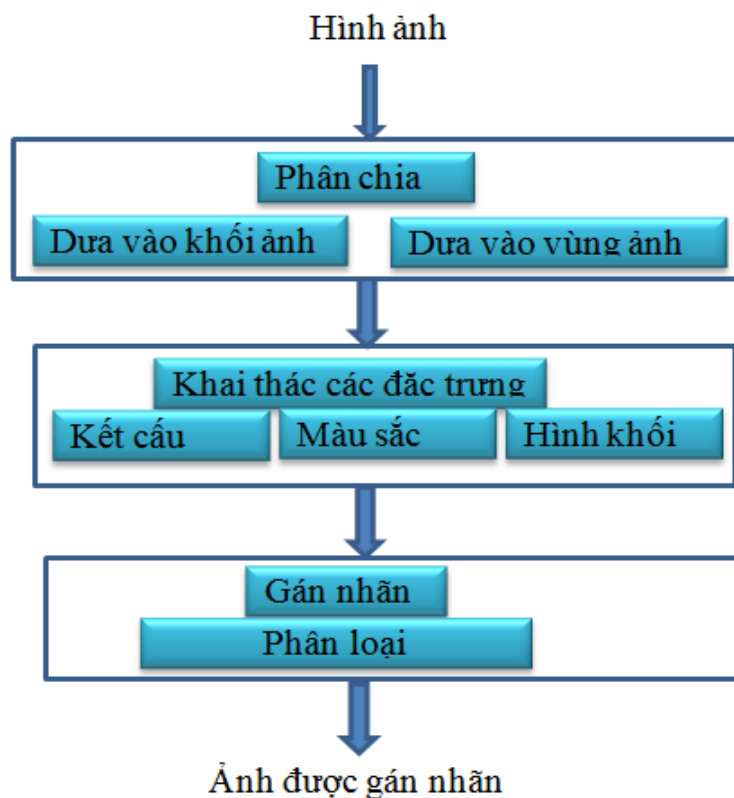
Gán nhãn ảnh tự động có thể được định nghĩa là quá trình mô hình hóa công việc của người thực hiện gán nhãn bằng tay khi gán các từ khóa cho hình ảnh dựa trên các thuộc tính của ảnh. Đến nay phần lớn các hệ thống gán nhãn ảnh được dựa trên sự kết hợp của việc phân tích hình ảnh và các kỹ thuật học máy. Để nâng cao tính chính xác của gán nhãn, những nghiên cứu tập trung đã được chuyển từ sự thiết kế phức tạp của thuật toán khai thác các đặc trưng đến giảm khoảng cách ngữ nghĩa giữa các đặc trưng và sự phong phú về ngữ nghĩa của con người.

Theo truyền thống có hai xu hướng chính trong quá trình tìm kiếm hình ảnh. Xu hướng đầu tiên được gọi là truy xuất hình ảnh dựa trên nội dung (CBIR: Content based image retrieval) cũng được biết đến như là truy vấn theo nội dung hình ảnh (QBIC: Query by image content) hoặc là truy xuất thông tin hình ảnh dựa trên nội dung (CBVIR: Content based visual information retrieval) có nghĩa là việc tìm kiếm sẽ phân tích nội dung thực tế của hình ảnh bằng cách sử dụng các kỹ thuật phân tích hình ảnh.



Hình 1. 1 Một hệ thống CBIR điển hình [Manal and Nordin 2009].

Gán nhãn ảnh tự động cũng được biết đến bao gồm một số kỹ thuật nhằm tìm ra mối tương quan giữa các đặc trưng mức thấp và ngữ nghĩa mức cao. Khó khăn chính trong việc gán nhãn ảnh tự động là tạo ra một mô hình có thể gán các từ khóa chính cho ảnh để có thể mô tả thành công nội dung của ảnh đó. Điểm khởi đầu cho hầu hết các thuật toán này là một tập ảnh huấn luyện đã được gán nhãn bằng tay. Thông tin phụ trợ bao gồm các từ khóa đơn giản mô tả lại nội dung của ảnh. Kỹ thuật phân tích hình ảnh được sử dụng để trích xuất các đặc trưng của ảnh như màu sắc, kết cấu, hình khối để tạo mô hình phân bố một thuật ngữ có mặt trong ảnh. Các đặc trưng có thể thu được từ toàn bộ hình ảnh (phương pháp tiếp cận toàn cục), hoặc từ các khối màu được phân chia của ảnh (phương pháp tiếp cận cục bộ). Bước tiếp theo là trích xuất thông tin đặc trưng từ hình ảnh chưa biết để so sánh nó với tất cả tập ảnh huấn luyện đã được tạo ra bằng tay. Kết quả của so sánh này mang lại một giá trị xác suất của mỗi từ khóa được gán trong ảnh. Sơ đồ khối của khung gán nhãn ảnh tự động được thể hiện ở Hình 1. 2.



Hình 1. 2 Sơ đồ khối của một khung công việc gán nhãn ảnh tự động.

Có ba phương pháp gán nhãn cho ảnh: Bằng tay, tự động và bán tự động [Wenyin et al. 2002]. Gán nhãn bằng tay cần người thực hiện nhập một vài từ khóa miêu tả khi thực hiện gửi ảnh. Ngược lại, gán nhãn tự động phát hiện và gán từ có ngữ nghĩa với nội dung của ảnh một cách tự động, không cần con người can thiệp. Với gán nhãn bán tự động, nó cần tương tác của người thực hiện để cung cấp truy vấn ban đầu và phản hồi cho gán nhãn ảnh trong khi trình duyệt. Bảng 1.1 và Bảng 1.2 đưa ra so sánh ba kỹ thuật gán nhãn trên về kỹ thuật, ưu điểm và nhược điểm riêng của chúng.

Bảng 1. 1 So sánh các kỹ thuật gán nhãn.

Kỹ thuật gán nhãn	Bằng tay	Bán tự động	Tự động
Tương tác của con người ban đầu	Nhập một vài từ khóa miêu tả	Cung cấp truy vấn ban đầu ngay từ khi bắt đầu	Không cần tương tác
Nhiệm vụ của máy	Cung cấp đĩa hoặc cơ sở dữ liệu để lưu trữ ảnh được gán nhãn	Phân tích truy vấn của con người và trích xuất thông tin ngữ nghĩa để thực hiện gán nhãn	Sử dụng công nghệ nhận dạng tự động phát hiện và gán từ có ngữ nghĩa cho ảnh
Nhiệm vụ của con người	Thực hiện đầy đủ thông tin ngữ nghĩa cho các mục đích tìm kiếm	Thực hiện một số gán nhãn và làm việc với đầu ra của máy	Xác nhận kết quả đầu ra và tính độ chính xác của việc gán nhãn tự động

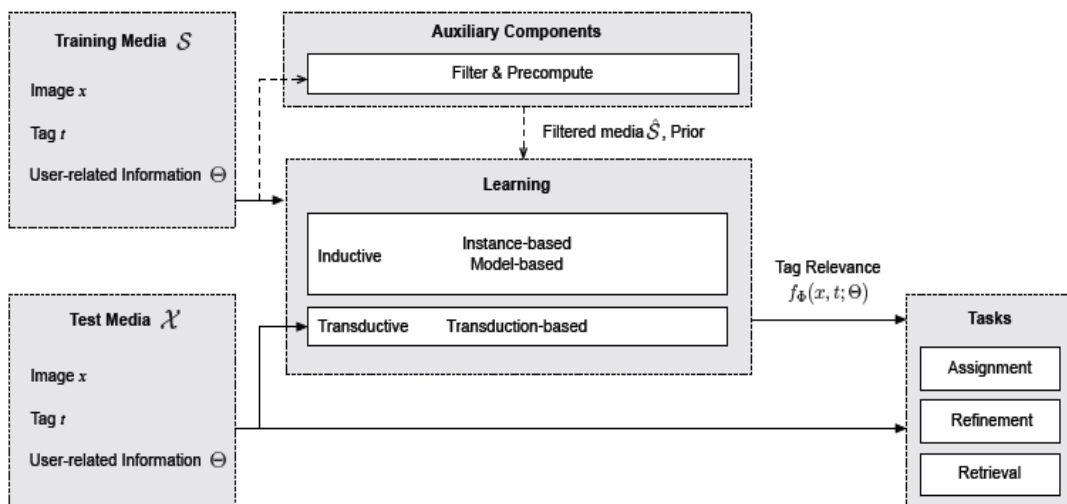
Bảng 1. 2 So sánh ưu điểm và nhược điểm của các kỹ thuật gán nhãn.

Kỹ thuật gán nhãn	Bằng tay	Bán tự động	Tự động
Ưu điểm	Độ chính xác cao	Chất lượng của gán nhãn cải thiện theo cách thức tương tác sau khi hiệu chỉnh	Hiệu quả nhất, tốn ít thời gian, thực hiện được với tập dữ liệu lớn
Nhược điểm	Mất thời gian, tốn kém, khó thực hiện với tập dữ liệu lớn, không thống nhất	Mất ít thời gian hơn so với gán nhãn bằng tay, nhưng vẫn nhiều hơn so với tự động gán nhãn	Dễ bị lỗi, độ chính xác thấp hơn so với gán nhãn bằng tay và bán tự động

1.2. Các nghiên cứu trong lĩnh vực gần đây

Nội dung chính của luận văn này là trình bày phương pháp gán nhãn sử dụng nhiều nguồn thông tin, do vậy trong phần này chúng tôi sẽ trình bày sơ lược về các nghiên cứu được công bố gần đây liên quan đến bài toán gán nhãn ảnh kết hợp các nguồn thông tin khác nhau.

Hình 1.3 trình bày một cách tổng quát cách thức thực hiện bài toán gán nhãn ảnh sử dụng các nguồn thông tin khác nhau. So sánh các phương pháp truyền thống chỉ dựa vào nội dung ảnh (visual information), hoặc chỉ dựa vào thông tin dạng text, phương pháp được trình bày ở phía dưới đây thông tin được sử dụng có thể là nội dung ảnh, các nhãn và cả thông tin liên quan đến người dùng. Để thu được tập huấn luyện chính xác, các bộ lọc có thể được sử dụng để loại bỏ các dữ liệu không mong muốn. Ngoài ra, các thông tin tiền xử lý như số lượng thống kê nhãn, tương quan nhãn và tương thích hình ảnh trong tập huấn luyện cũng được tính toán từ trước để đảm bảo cho hiệu quả về mặt thời gian cũng như tốc độ xử lý. Thông qua mô hình huấn luyện, với mỗi ảnh đầu vào ta sẽ xác định được nhãn thích hợp cho nó. Nhãn này có thể sử dụng trong bài toán gán nhãn, hoặc bài toán hiệu chỉnh nhãn, hoặc truy xuất dữ liệu.



Hình 1.3 Phương pháp tổng quát cách thức thực hiện bài toán gán nhãn ảnh sử dụng các nguồn thông tin khác nhau [Li et al. 2016].

Căn cứ vào nguồn thông tin sử dụng, có thể phân loại các phương pháp gán nhãn ảnh thành 3 loại như sau:

Dựa vào nhãn: Phương pháp này xây dựng hoàn toàn dựa trên thông tin nhãn. Ý tưởng cơ bản của phương pháp này là giả sử rằng ảnh kiểm thử đã được gán nhãn một vài nhãn trước đó, từ đó gán thêm những nhãn có giá trị liên quan hoặc có ngữ nghĩa gần với phần lớn các nhãn liên quan cho ảnh kiểm thử. Các phương pháp nổi bật [Sigurbjörnsson and Van Zwol 2008; Zhu et al. 2012, Xu et al. 2009].

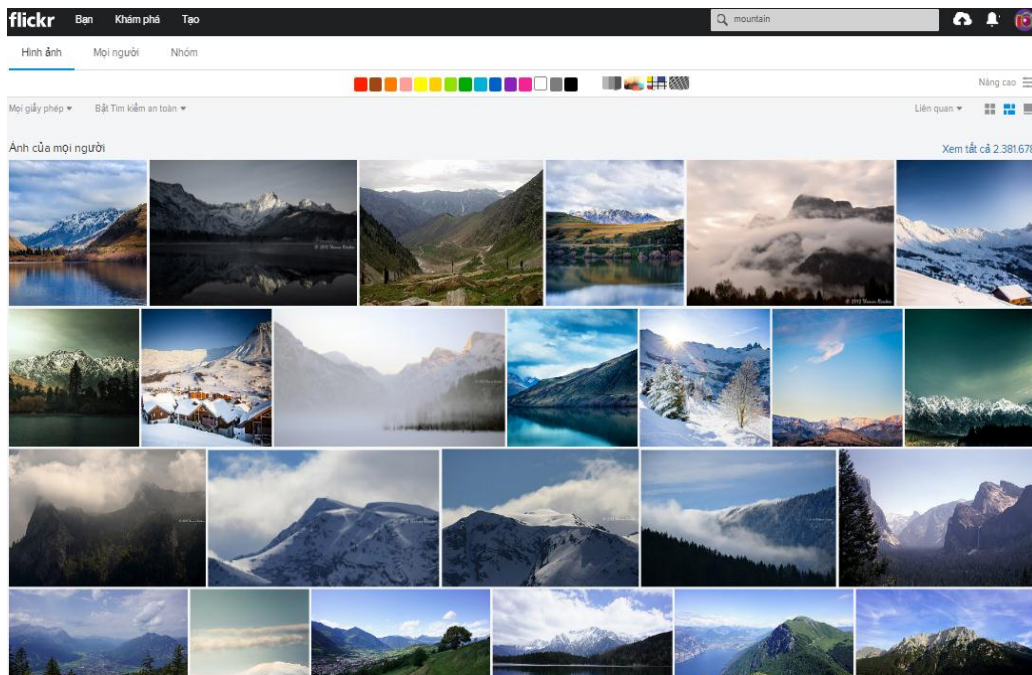
Dựa vào nhãn và ảnh: Phương pháp này sử dụng thông tin hình ảnh và các nhãn liên kết có sẵn. Bởi vì các ảnh có cùng nội dung cần phải/nên được gán các nhãn giống nhau. Do đó, ta có thể so sánh về mặt nội dung giữa ảnh kiểm thử và các ảnh huấn luyện, từ đó chuyển giao nhãn giữa ảnh test và ảnh huấn luyện. Có thể chia phương pháp này thành ba nhóm chính. Một là, sử dụng nội dung ảnh gần với ảnh thực nghiệm [Li et al. 2009b; Li et al. 2010; Verbeek et al. 2010; Ma et al. 2010; Wu et al. 2011; Feng et al. 2012]. Hai là, khai thác mối quan hệ giữa các ảnh được gán nhãn với nhãn giống nó [Liu et al. 2009; Richter et al. 2012; Liu et al. 2011b; Kuo et al. 2012; Gao et al. 2013]. Ba là, tìm kiếm phân loại nội dung từ ví dụ được gán nhãn xã hội [Wang et al. 2009; Chen et al. 2012; Li and Snoek 2013; Yang et al. 2014].

Dựa vào nhãn và ảnh và thông tin khác: Ngoài phương pháp nhãn và ảnh, công việc trong nhóm này khai thác thêm thông tin khác. Các thông tin này có thể là thông tin người dùng [Li et al. 2009b], nhãn yêu thích [Sawant et al. 2010], đến độ tin cậy người dùng [Ginsca et al. 2014], thành viên nhóm ảnh [Johnson et al. 2015]. Tương tự, việc xây dựng lại dữ liệu dựa trên tensor trước được thực hiện trong [Qian et al. 2015], để khám phá mối liên hệ tiềm ẩn giữa người dùng, ảnh và nhãn. Yếu tố thời gian, GPS được sử dụng [Kim and Xing 2013, McParlane et al. 2013].

1.3. Bài toán phân loại và gán nhãn ảnh tự động

Ảnh kỹ thuật số ngày càng trở lên dễ dàng sau những tiến bộ nhanh chóng trong kỹ thuật chụp ảnh số, mạng và công nghệ lưu trữ. Những hình ảnh được chia sẻ trên website như Flickr và Picasa là rất phổ biến trong đời sống hàng ngày. Ví dụ như có hơn 2000 ảnh được tải lên Flickr mỗi phút [<http://www.flickr.com/>]. Trong thời gian cao điểm, lên tới 12000 ảnh được tải lên mỗi giây, và kỷ lục về số lượng hình ảnh được tải lên mỗi ngày có thể vượt quá 2 triệu ảnh [F. Blog. <http://blog.flickr.net/en/2007/05/29/were-going-down/>]. Khi người dùng chia sẻ những hình ảnh của họ, họ thường gán một vài nhãn để

miêu tả nội dung bức ảnh của họ ví dụ như hình 1.10. Trong quá trình tìm kiếm ảnh, người dùng có thói quen tìm kiếm ảnh bằng từ khóa và việc tìm kiếm bằng từ khóa thì đơn giản hơn, nhanh hơn, dễ dàng thực hiện hơn so với việc tìm kiếm bằng hình ảnh. Căn cứ vào đó, tôi tiến hành nghiên cứu và thực hiện các kỹ thuật gán nhãn ảnh tự động. Nhiệm vụ chính là tự động gán cho mỗi ảnh một vài từ khóa. Các từ khóa này có thể được lấy từ tên của ảnh, các từ khóa xung quanh ảnh hay bằng các phân tích nội dung của ảnh cần gán nhãn với các ảnh đã biết (ảnh huấn luyện), nhằm mục đích chủ yếu là đơn giản hóa quá trình tìm kiếm và truy cập dữ liệu. Để thực hiện bài toán phân loại và gán nhãn ảnh tự động trong nội dung của luận văn này chúng tôi chủ yếu nghiên cứu các đặc trưng của ảnh và thực hiện biểu diễn chúng dưới dạng các vector để máy tính có thể dễ dàng xử lý. Thêm vào đó, chúng tôi khai thác thêm thông tin phụ trợ để cải thiện độ chính xác của quá trình gán nhãn. Hai thuật toán phân loại được sử dụng trong luận văn là K-NN(K-Nearest Neighbors) và SVM (Support Vector Machine) [Cortes and Vapnik 1995]. Chúng tôi sử dụng cả hai thuật toán trên nhằm mục đích so sánh kết quả của chúng để chọn ra thuật toán tốt hơn cho quá trình phân loại và gán nhãn. Cuối cùng chúng tôi đánh giá sự ảnh hưởng của thuật toán đến các đặc trưng.



Hình 1. 4 Ví dụ tìm kiếm ảnh với từ khóa “mountain” trên Flickr.com.

1.4. Biểu diễn ảnh dưới dạng các vector đặc trưng (Low level features)

1.4.1. Tại sao phải biểu diễn ảnh

Để cho phép người dùng sử dụng, truy xuất, tìm kiếm ... trong các tập ảnh siêu lớn và tiếp tục ngày càng tăng (ví dụ như để tìm kiếm các hình ảnh có chứa đối tượng người cụ thể, hoặc để sắp xếp các ảnh vào các chủ đề phân biệt) thì các hình ảnh cần được lập chỉ mục (index) hay gán nhãn bằng các từ khóa có ngữ nghĩa. Nhưng đối với người dùng, công việc gán nhãn hình ảnh là công việc tẻ nhạt, phức tạp và rất khó thực hiện với các tập dữ liệu lớn. Do đó, cần thiết có các kỹ thuật tự động gán nhãn ảnh trực tiếp thực hiện bằng máy tính. Mặc dù vậy nhưng thực tế thì không thể trực tiếp sử dụng các nội dung hình ảnh. Ta không thể so sánh trực tiếp hai ảnh với nhau được do kích thước ảnh khác nhau, vị trí kích thước đối tượng trong ảnh khác nhau, độ sáng tối, góc nhìn của ảnh khác nhau. Điều này dẫn đến cần thiết phải biểu diễn ảnh lại thành một dạng nào đó hay chuyển nó về cùng một không gian để so sánh bằng các đặc trưng ảnh như các vector cơ sở biểu diễn lại ảnh thành các vector - của các đặc trưng ảnh - mà máy tính có thể xử lý dễ dàng và hiệu quả. Một phép biểu diễn ảnh tốt cho phép mã hóa tất cả các thông tin có liên quan về các nội dung trực quan của ảnh. Những thông tin trong ảnh được coi là có liên quan phụ thuộc vào loại dữ liệu, ứng dụng và mục đích mà chúng ta mong muốn. Ví dụ để phân biệt các bức ảnh thành phố và các bãi biển sử dụng thuộc tính màu sắc sẽ rất hiệu quả. Tuy nhiên, sử dụng màu sắc để phân biệt con bò và con ngựa trên thảo nguyên lại rất khó khăn (Hình 1.5). Hoặc với những bức ảnh giống nhau nhưng bị xoay đi theo các góc nhìn khác nhau, chúng ta cũng không thể sử dụng lược đồ hệ số góc để phân loại chúng (Hình 1.6). Nhưng sử dụng lược đồ màu thì lại hiệu quả đối với các ảnh bị quay hay phóng. Điều này cũng khẳng định với các tập dữ liệu lớn, phức tạp các phép biểu diễn ảnh đơn giản sử dụng chỉ một loại đặc trưng ảnh sẽ không thể đáp ứng được. Chúng ta cần thiết nghiên cứu các phép biểu diễn ảnh phức tạp hơn, kết hợp được nhiều đặc trưng của ảnh hơn để khắc phục được nhược điểm của mỗi đặc trưng và phát huy được ưu điểm của các đặc trưng.



Hình 1.5 Ví dụ minh họa không thể dùng lược đồ màu để phân biệt con bò và con ngựa trên thảo nguyên

Ví dụ Hình 1.5 ta có thể thấy con bò và con ngựa có màu sắc tương đối giống nhau. Nếu sử dụng lược đồ màu để biểu diễn hai bức ảnh trên chúng ta khó có thể phân biệt được đâu là lược đồ màu biểu diễn cho con bò và đâu là lược đồ màu biểu diễn cho con ngựa. Trong trường hợp này sử dụng lược đồ màu là không hiệu quả cho việc phân loại ảnh.



Hình 1.6 Minh họa hai ảnh giống nhau nhưng góc nhìn khác nhau

1.4.2. Đặc trưng của ảnh

Đặc trưng của ảnh là các thuộc tính cơ bản của ảnh mà có thể sử dụng để phân biệt các ảnh với nhau. Có nhiều loại đặc trưng của ảnh: màu sắc, độ tương

phản, hình khối, cấu trúc, thông tin tần số cao / tần số thấp... Các đặc trưng của ảnh thường được dùng trong phép biểu diễn ảnh. Có thể chia đặc trưng làm 2 loại là đặc trưng toàn cục và đặc trưng cục bộ.

1.4.2.1. Đặc trưng toàn cục

Đặc trưng ảnh toàn cục cho phép mô tả lại cả bức ảnh, các thuộc tính của toàn ảnh bằng một vector biểu diễn duy nhất. Ưu điểm của phương pháp này là gọn, duy nhất, dễ tính toán và không thay đổi theo cấu trúc của ảnh. Tuy vậy phép biến đổi này lại rất khó áp dụng để phân loại hay nhận diện các đối tượng trong ảnh bởi vì thông tin về các đối tượng đều được đồng hóa cùng với thông tin nền thành một thể thống nhất.

1.4.2.2. Đặc trưng cục bộ

Đặc trưng cục bộ biểu diễn bởi một tập hợp các đặc trưng trích xuất từ các khối ảnh rời rạc. Ưu điểm của phương pháp là phân biệt rõ ràng đặc trưng liên quan đến đối tượng ảnh, đặc trưng liên quan đến thông tin nền, do đó có thể sử dụng hiệu quả trong các thuật toán phân loại và nhận diện ảnh. Hạn chế của phương pháp này là sai số: các khối ảnh có thể chứa thông tin nhiều thành phần của nhiều đối tượng khác nhau, do đó việc trích xuất đặc trưng cũng mang tính xấp xỉ. Như vậy mỗi loại đặc trưng đều có ưu điểm, nhược điểm riêng. Việc chọn sử dụng đặc trưng nào, sẽ tùy thuộc vào loại dữ liệu ảnh (ảnh phong cảnh/ ảnh đối tượng), loại đặc trưng sử dụng màu sắc/ cấu trúc/ hình khối ..., mục đích áp dụng.

1.4.3. Các vector đặc trưng

1.4.3.1. Lược đồ màu CH

Lược đồ màu biểu thị phân bố của các điểm ảnh cho mỗi dải được lượng hóa [Shapiro and Stockman 2003]. Lược đồ màu được tính toán dễ dàng và hiệu quả trong mô tả phân bố màu toàn cục và cục bộ trong ảnh. Hơn nữa, lược đồ màu không nhạy cảm với quay và dịch chuyển về trục quan sát và thay đổi chậm với tỷ lệ và vị trí quan sát. Do mọi điểm ảnh trong ảnh có thể được mô tả bởi ba thành phần màu trong một không gian màu nào đó (ví dụ các thành phần đỏ,

xanh lam, xanh lơ trong không gian RGB, hoặc sắc màu, độ nét và giá trị trong không gian HSV), một lược đồ có thể được định nghĩa cho mỗi thành phần. Một lược đồ màu chứa nhiều dải hơn sẽ có khả năng phân biệt các ảnh tốt hơn. Tuy nhiên, điều này sẽ tăng độ phức tạp tính toán và khó khăn cho cơ chế đánh chỉ số cơ sở dữ liệu ảnh. Lược đồ màu là đại lượng đặc trưng cho phân bố màu cục bộ của ảnh được định lượng bởi công thức :

$$h(i) = \frac{n_i}{N} \quad , \quad i=1,2,\dots, K \quad (1-1)$$

Trong đó: n_i số điểm ảnh với giá trị i , N là tổng số điểm ảnh trong ảnh, K là kích thước của bin lượng tử (với $K=4$). Kết quả là lược đồ màu có số chiều là 64 ($4 \times 4 \times 4$).

Hạn chế chính của lược đồ màu là chưa tận dụng được thông tin không gian của các vùng ảnh. Điều này có thể dẫn đến sai số không mong muốn; trong gán nhãn ảnh sử dụng lược đồ màu là không thể phân biệt hai ảnh khác nhau nhưng có lược đồ màu giống nhau.

Lược đồ màu bất biến đối với phép quay và tịnh tiến ảnh, và nếu chuẩn hoá lược đồ màu sẽ bất biến đối với phép co giãn.

1.4.3.2. Đặc trưng tương quan màu CORR

Đặc trưng tương quan màu [Huang et al. 1997] mô tả các phân bố màu của các điểm ảnh và chỉ ra tương quan không gian của các cặp màu. Chiều thứ nhất và thứ hai của lược đồ màu ba chiều là các màu của các cặp điểm ảnh và chiều thứ ba là khoảng cách không gian của chúng. Một tương quan màu là một mảng được đánh chỉ số bởi các cặp màu, ở đây mục thứ k cho (i, j) chỉ rõ xác suất tìm được một điểm ảnh có màu j tại một khoảng cách k từ một điểm ảnh có màu i trong ảnh. Cho I biểu diễn toàn bộ tập các điểm ảnh và $I_{c(j)}$ biểu diễn tập các điểm ảnh có màu $c(i)$. Tương quan màu được định nghĩa như sau:

$$\gamma_{i,j}^k = Prob_{p_1 \in I_{c(i)}, p_2 \in I_{c(j)}} [p_2 \in I_{c(j)} \mid |p_1 - p_2| = k] \quad (1-2)$$

Ở đây $i, j \in \{1, 2, \dots, N\}$, $k \in \{1, 2, \dots, d\}$, và $|p_1 - p_2|$ là khoảng cách giữa các điểm ảnh p_1 và p_2 . Nếu xét tất cả các kết hợp có thể của các cặp màu, thì số

chiều của tương quan màu rất lớn, tương quan màu chỉ thu thập được sự tương quan về không gian giữa các cặp màu giống hệt nhau và do đó làm giảm kích thước từ $O(N^2d)$ tới $O(Nd)$. Tác giả đã lượng tử hóa các thành phần màu HSV (H: Hue/ Vùng màu; S: Saturation/ Độ bão hòa màu; V: Value/ Độ sáng) thành 36 bins và thiết lập khoảng cách metric thành bốn khoảng lẻ là $d = \{1, 3, 5, 7\}$. Do vậy tương quan màu có số chiều là 144 (36x4).

So sánh với lược đồ màu, tương quan màu cho các kết quả tra cứu màu tốt hơn. Tuy nhiên tương quan màu có độ phức tạp tính toán cao, do vector đặc trưng có số chiều cao.



Hình 1.7 Minh họa hai ảnh có đặc trưng tương quan màu giống nhau

1.4.3.3. Lược đồ hệ số góc EDH

Lược đồ hệ số góc mã hóa sự phân bố liên kết của các góc [Park 2000]. Lược đồ gồm 73 phần tử trong đó: 72 phần tử đầu chứa số điểm ảnh có hệ số góc từ 0 -355 độ, các hệ số góc này cách nhau 5 độ. Phần tử cuối chứa số phần tử không nằm trên biên cạnh. Cần chuẩn hóa các đặc trưng này để thích hợp với kích thước khác nhau của ảnh.

Ta có biểu thức:

$$H_i = \frac{H(i)}{M_e}, \text{ nếu } i \in [0, \dots, 71] \quad (1-3)$$

$$H_i = \frac{H(i)}{M}, \text{ nếu } i=72 \quad (1-4)$$

Trong đó: H_i là số điểm ảnh thuộc biên cạnh có hệ số góc là $\alpha_i = i*5$

M_e là tổng các điểm ảnh thuộc biên cạnh

M là tổng số điểm ảnh của ảnh

Để tìm kiếm một ảnh mà ta quan tâm nhiều đến khía cạnh hình dáng của các đối tượng trong ảnh thì nên sử dụng các đặc trưng về hình dáng. Đặc trưng cơ bản nhất về hình dáng là lược đồ hệ số góc. Đây là đặc trưng chỉ thể hiện được hình dáng chung của các đối tượng trong ảnh.

1.4.3.4. Cấu trúc sóng con WT

Sóng con cung cấp cách tiếp cận đa độ phân giải để phân tích cấu trúc [Manjunath and Ma 1996]. Về cơ bản biến đổi sóng con phân tách một tín hiệu với tập hợp các tín hiệu dựa trên hàm $\Psi_{mn}(x)$ thu được thông qua sự dịch chuyển và giãn nở của sóng mẹ $\Psi(x), i.e.,$

$$\Psi_{mn}(x) = 2^{-\frac{m}{2}} \Psi(2^{-m}x - n) \quad (1-5)$$

Ở đây m và n là hai thông số giãn nở và dịch chuyển. Một tín hiệu $f(x)$ có thể được biểu diễn như sau:

$$f(x) = \sum_{m,n} c_{mn} \Psi_{mn}(x) \quad (1-6)$$

Biến đổi sóng con thực hiện trên hình ảnh liên quan đến lọc đệ quy và lấy mẫu phụ. Ở mỗi mức độ, hình ảnh được chia thành dải bốn tần số phụ LL, LH, HL, HH, trong đó L biểu thị cho tần số thấp, H biểu thị cho tần số cao. Hai loại biến đổi sóng con chính thường được sử dụng để phân tích cấu trúc là biến đổi sóng con theo cấu trúc kim tự tháp (PWT: Pyramid-structured wavelet transform) và biến đổi sóng con theo cấu trúc hình cây (TWT: Tree-structured wavelet transform). PWT phân tách đệ quy theo băng LL còn TWT phân tách các băng tần còn lại LH, HL, HH để giữ lại những thông tin quan trọng nhất xuất hiện trong kênh trung tần.

Sau khi phân tách, vector đặc trưng có thể được xây dựng bằng cách sử dụng độ lệch trung bình và độ lệch chuẩn của sự phân bố năng lượng của mỗi băng phụ tại mỗi mức độ. Đối với ba mức độ phân tách, kết quả PWT là một vector đặc trưng 24 ($3 \times 4 \times 2$) thành phần. Còn với TWT, vector đặc trưng sẽ phụ thuộc vào băng phụ tại mỗi mức độ được phân tách như thế nào. Một cây phân tách cố định có thể thu được theo thứ tự các băng LL, LH và HL. Kết quả thu được một vector đặc trưng là 104 (52×2) thành phần.

1.4.3.5. Mô-men màu CM55

Mô-men màu là các mô-men thống kê của các phân bố xác suất của các màu [Stricker and Orengo 1995]. Các mô-men màu được sử dụng trong nhiều hệ thống tra cứu ảnh như QBIC [Biblack et al. 1993; Flickner et al. 1995]. Các mô-men màu bậc nhất (trung bình), bậc hai (phương sai) và bậc ba (độ lệch), đã được chứng minh là hiệu quả trong biểu diễn các phân bố màu của ảnh [Stricker and Orengo 1995].

Về mặt toán học, ba mô-men đầu tiên được định nghĩa như sau:

$$\mu_i = \frac{1}{N} \sum_{j=1}^N f_{ij} \quad (1-7)$$

$$\sigma_i = \left(\frac{1}{N} \sum_{j=1}^N (f_{ij} - \mu_i)^2 \right)^{\frac{1}{2}} \quad (1-8)$$

$$S_i = \left(\frac{1}{N} \sum_{j=1}^N (f_{ij} - \mu_i)^3 \right)^{\frac{1}{3}} \quad (1-9)$$

Trong đó f_{ij} là giá trị của thành phần màu thứ i của điểm ảnh j và N là tổng số các điểm ảnh trong ảnh.

Mô-men màu diễn tả nhỏ gọn nội dung của ảnh so với các đặc trưng màu khác. Đối với việc sử dụng ba mô-men màu như miêu tả ở trên, chỉ 9 thành phần (ba mô-men màu, mỗi mô-men màu có ba thành phần màu) được sử dụng để biểu diễn đặc trưng màu của mỗi ảnh. Do tính chặt chẽ này, các mô-men màu có thể giảm khả năng phân biệt ảnh. Thông thường, các mô-men màu có thể được sử dụng như sơ duyệt lần đầu để giảm không gian tra cứu trước khi các đặc trưng màu phức tạp khác được sử dụng. Vì vậy, đối với bộ dữ liệu NUS-WIDE, tác giả đã trích xuất mô-men màu thành các block-wise thông qua lưới ngăn cố định 5x5, tạo ra một mô-men màu block-wise có số chiều là 225.

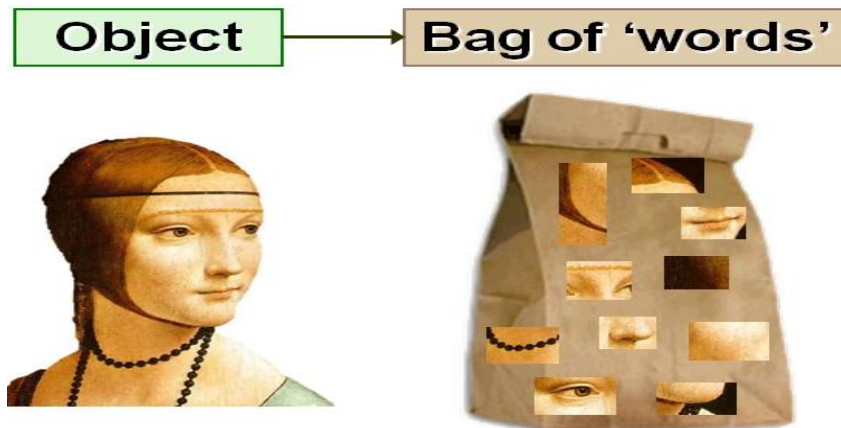
1.4.3.6. Phương pháp túi từ điển BOW

Như tên gọi, khái niệm BOW thực sự được lấy từ cách phân tích văn bản. Ý tưởng của phương pháp này là trình bày văn bản như một “túi” các từ khóa quan trọng mà không có thứ tự của các từ (đó là lý do tại sao gọi là “túi từ” thay vì gọi là một danh sách các ví dụ) [Fei et al. 2005].

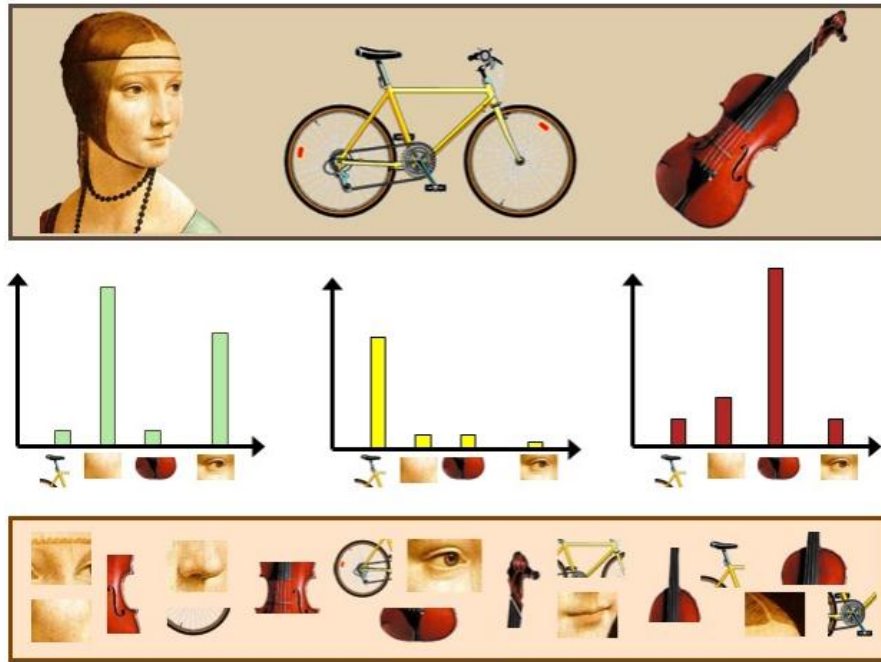
Trong thị giác máy tính, ý tưởng của phương pháp này là tương tự. Tác giả miêu tả một đối tượng như một túi từ đặc trưng - đối tượng được chia nhỏ ra và được mô tả lại bằng các khối nhỏ (Hình 1.8).

Chúng ta có thể sử dụng mô hình túi từ điển để phân loại đối tượng bằng cách xây dựng một từ vựng lớn với nhiều từ trực quan và mô tả từng hình ảnh như là một biểu đồ thể hiện tần số xuất hiện của các từ trong hình ảnh. Hình 1.9 minh họa ý tưởng này.

Phương pháp từ điển [Lowe 2004]: Việc tạo ra các từ ngữ trực quan bao gồm 3 bước chính: (a) Tác giả vận dụng sự khác biệt của hàm lọc Gaussian trên tỷ lệ xám của ảnh để tìm ra các điểm mấu chốt và tỉ lệ tương đối; (b) Tác giả tính toán SIFT (Scale Invariant Feature Transform) [Lowe 2004] qua vùng cục bộ được định nghĩa bởi điểm mấu chốt và tỉ lệ; (c) Tác giả thực hiện lượng tử hóa vector trên vùng miêu tả SIFT để xây dựng từ điển trực quan bằng cách khai thác cụm k-means [Flickner et al. 1995; Li et al. 2009]. Ở đây, tác giả tạo ra 500 cụm và vì vậy số chiều của phương pháp từ điển là 500.



Hình 1. 8 Minh họa phương pháp túi từ điển [Fei et al. 2005]



Hình 1. 9 Túi từ - miêu tả đối tượng với biểu đồ của các từ xuất hiện[Fei et al. 2005]

1.5. Các nguồn thông tin phụ trợ

Các nguồn thông tin phụ trợ (metadata) của ảnh là tất cả những thông tin liên quan đến ảnh khai thác được từ internet. Thông tin phụ trợ cung cấp thông tin chi tiết về ảnh giúp người dùng phân loại chúng vào những nhóm khác nhau. Thông tin phụ trợ cũng cấp các thông tin như lý lịch mô tả nguồn gốc, định dạng, tác giả, máy chụp, thời gian, vị trí GPS, tags của ảnh. Nó là nguồn thông tin quan trọng liên quan trực tiếp đến nội dung của ảnh. Ý nghĩa của mỗi thông tin phụ trợ có thể khai thác và sử dụng trong quá trình phân loại và gán nhãn như sau:

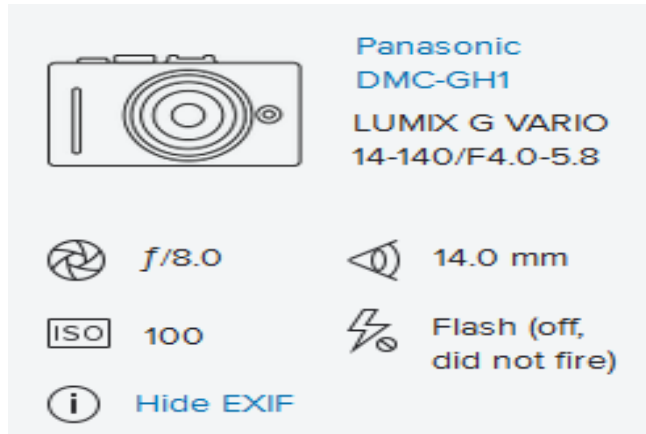


Hình 1. 10 Ví dụ về ảnh và các tag người dùng gán cho ảnh

- Những mô tả/ chú giải ảnh (tags) mà người dùng gán cho ảnh liên quan đến trực tiếp đến nội dung của ảnh - Dựa vào những chú giải được người dùng

gắn cho ảnh ta có thể biết được bức ảnh thuộc chủ đề nào, chụp ở đâu, đối tượng trong ảnh gồm những gì, ảnh chụp phong cảnh hay thành phố.

- Các thông số liên quan đến máy ảnh :



Hình 1. 11 Thông tin về các thông số của máy ảnh và EXIF của ảnh

EXIF (Exchangeable image file format) của hình ảnh: Dựa vào EXIF của ảnh ta có thể thu được rất nhiều thông tin liên quan đến ảnh như: Vị trí GPS, thời gian. Tùy vào thuộc tính của ảnh có thể chọn một số loại thông tin. Ví dụ, dựa vào thời gian có thể được sử dụng để phát hiện sự kiện, phân loại cảnh ban ngày hay ban đêm, phân loại theo mùa ... Vị trí GPS thường được sử dụng để phân loại và gán nhãn các cảnh ngoài trời chẳng hạn như phong cảnh khi ảnh chụp ở những cảnh giống nhau hoặc các di tích có vị trí GPS trong một vùng. Vị trí địa lý này là một đặc trưng tốt cho việc phát hiện và công nhận những di tích. Trong trường hợp ảnh hay đối tượng ảnh được chụp ở trong nhà thì thông tin vị trí địa lý là không hữu ích.

- Tác giả của bức ảnh cũng có thể được sử dụng trong quá trình phân loại dựa vào các mục đích phân loại khác nhau.

- Những bình luận (comments) của người xem cũng liên quan đến nội dung của ảnh.

Hầu hết những hình ảnh lấy từ web đều mang thông tin phụ trợ. Ý tưởng sử dụng thông tin phụ trợ để phân loại và gán nhãn cho ảnh không phải là mới. Những nghiên cứu trước đó đã cho thấy lợi ích của việc sử dụng nhãn để phân loại và truy xuất ảnh [Chen et al. 2010; Guillaumin et al. 2010; Hwang and

Grauman 2012; Niu et al. 2014], sử dụng dữ liệu GPS [Hays and Efros 2008; Li et al. 2009; Zamir et al. 2014]] để cải thiện việc phân loại ảnh, và sử dụng thời gian [Biblack et al. 1993] để cải thiện sự nhìn nhận và nghiên cứu sự tiến triển của chủ đề theo thời gian. Trong đề tài này, chúng tôi cũng tiến hành phân tích dữ liệu thông tin phụ trợ để cải thiện kết quả gán nhãn tự động cho ảnh. Qua quá trình phân tích tập dữ liệu của NUS-WIDE chúng tôi nhận thấy tất cả các ảnh trong tập dữ liệu NUS-WIDE đều có chứa thông tin chú giải. Do vậy, chúng tôi chỉ sử dụng thông tin chú giải kết hợp với các đặc trưng cho quá trình phân loại và gán nhãn ảnh. Đối với các thông tin khác như GPS hay thời gian không phải tất cả các ảnh đều có chứa thông tin đó nên không đủ dữ liệu vì vậy chúng tôi không tiến hành thực nghiệm với nó.

Như vậy, Thông tin phụ trợ có chứa rất nhiều nội dung liên quan trực tiếp đến ảnh. Nó đưa ra thông tin khá chi tiết và đầy đủ về ảnh. Vì vậy, chúng tôi sử dụng chúng kết hợp với các đặc trưng của ảnh với mong muốn mang lại kết quả cao nhất cho quá trình gán nhãn tự động cho ảnh.

1.6. Phương pháp huấn luyện dùng để phân loại và gán nhãn

1.6.1. Phương pháp K - láng giềng gần nhất (K-Nearest Neighbors)

K-Nearest Neighbor algorithm (K-NN) [Altman 1992] được sử dụng rất phổ biến trong lĩnh vực Data Mining. K-NN là phương pháp để phân lớp các đối tượng dựa vào khoảng cách gần nhất giữa đối tượng cần xếp lớp (Query point) và tất cả các đối tượng trong tập dữ liệu huấn luyện.

Một đối tượng được phân lớp dựa vào K láng giềng của nó. K là số nguyên dương được xác định trước khi thực hiện thuật toán. Người ta thường dùng khoảng cách Euclidean để tính khoảng cách giữa các đối tượng.

Thuật toán K-NN được mô tả như sau:

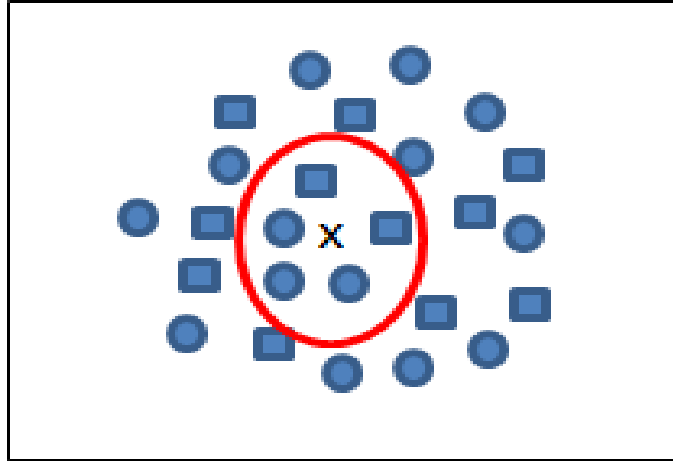
Xác định giá trị tham số K (số láng giềng gần nhất).

Tính khoảng cách giữa đối tượng cần phân lớp (Query point) với tất cả các đối tượng trong tập dữ liệu huấn luyện (thường sử dụng khoảng cách Eucidean).

Sắp xếp khoảng cách theo thứ tự tăng dần và xác định K láng giềng gần nhất với Query Point.

Lấy tất cả các lớp của K láng giềng gần nhất đã xác định.

Dựa vào phần lớn lớp láng giềng gần nhất để xác định lớp cho Query Point.



Hình 1. 12 Ví dụ phân lớp cho X bằng thuật toán K-NN

Ví dụ Hình 1.12 Với K=5, 5 phần tử gần nhất sẽ được lấy, có 3 phần tử là hình tròn, 2 phần tử là hình vuông. Suy ra, quyết định phần tử X là hình tròn.

Ta có công thức tính khoảng cách Minkowski:

$$d(i,j) = \sqrt[q]{(|x_{i_1} - x_{j_1}|^q + |x_{i_2} - x_{j_2}|^q + \dots + |x_{i_p} - x_{j_p}|^q)} \quad (2-1)$$

với $i = (x_{i_1}, x_{i_2}, \dots, x_{i_p})$ và $j = (x_{j_1}, x_{j_2}, \dots, x_{j_p})$ là 2 phần tử dữ liệu, trong đó p là số chiều, q là số nguyên dương.

• nếu $q = 1$, d là khoảng cách Manhattan.

$$d(i,j) = |x_{i_1} - x_{j_1}| + |x_{i_2} - x_{j_2}| + \dots + |x_{i_p} - x_{j_p}| \quad (2-2)$$

• nếu $q = 2$, d là khoảng cách Eucidean.

$$d(i,j) = \sqrt{|x_{i_1} - x_{j_1}|^2 + |x_{i_2} - x_{j_2}|^2 + \dots + |x_{i_p} - x_{j_p}|^2} \quad (2-3)$$

1.6.2. Phương pháp SVM - Support Vector Machine

SVM- Máy vector hỗ trợ [Cortes and Vapnik 1995] xây dựng một siêu phẳng hoặc tập hợp các điểm siêu phẳng trong một không gian nhiều chiều hoặc

vô hạn chiều, có thể được sử dụng cho phân loại, hồi quy, hoặc các nhiệm vụ khác. Một cách trực giác, để phân loại tốt thì các siêu phẳng nằm ở càng xa các điểm dữ liệu của tất cả các lớp (gọi là hàm lề) càng tốt, vì nói chung lề càng lớn thì sai số tổng quát của thuật toán càng bé.

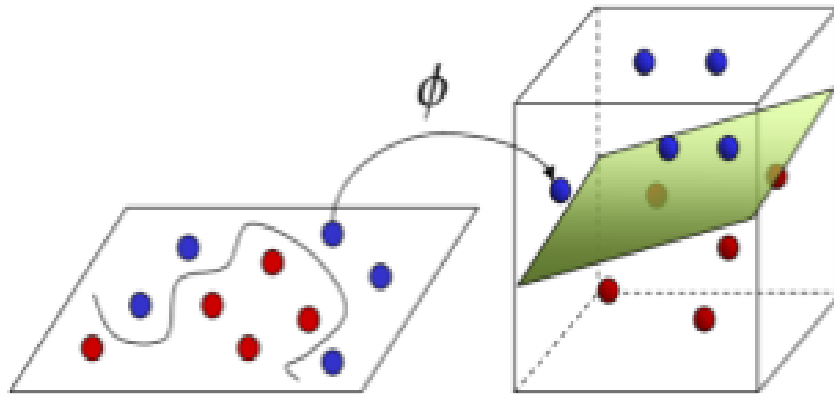
Trong nhiều trường hợp, không thể phân chia các lớp dữ liệu một cách tuyến tính trong một không gian ban đầu được dùng để mô tả một vấn đề. Vì vậy, nhiều khi cần phải ánh xạ các điểm dữ liệu trong không gian ban đầu vào một không gian mới nhiều chiều hơn, để việc phân tách chúng trở lên dễ dàng hơn trong không gian mới. Để việc tính toán được hiệu quả, ánh xạ sử dụng trong thuật toán SVM chỉ đòi hỏi tích vô hướng của các vector dữ liệu trong không gian mới có thể được tính dễ dàng từ các tọa độ trong không gian cũ. Tích vô hướng này được xác định bằng một hàm Kernel $K(x,y)$ phù hợp. Một siêu phẳng trong không gian mới được định nghĩa là tập hợp các điểm có tích vô hướng với một vector cố định trong không gian đó là một hằng số. Vector xác định một siêu phẳng sử dụng trong SVM là một tổ hợp tuyến tính của các vector dữ liệu luyện tập trong không gian mới với các hệ số. Với siêu phẳng lựa chọn như trên, các điểm x trong không gian đặc trưng được ánh xạ vào một siêu phẳng là các điểm thỏa mãn :

$$\sum_i \alpha_i K(x_i, x) = \text{hằng số} \quad (2-4)$$

Ghi chú rằng nếu $K(x,y)$ nhận giá trị ngày càng nhỏ khi y xa dần khỏi x thì mỗi số hạng của tổng trên được dùng để đo độ tương tự giữa x với điểm x_i tương ứng trong dữ liệu huấn luyện. Như vậy, tác dụng của tổng trên chính là so sánh khoảng cách giữa điểm cần dự đoán với các điểm dữ liệu đã biết. Lưu ý là tập các điểm x được ánh xạ vào một siêu phẳng có thể có độ phức tạp tùy ý trong không gian ban đầu, nên có thể phân tách các tập hợp thậm chí không lồi trong không gian ban đầu.

Ví dụ: Ta có tập các quả bóng đỏ và xanh ở trên bàn. Nếu các quả bóng không đan xen nhau, ta có thể dùng một cây que để chia các quả bóng thành hai

tập hợp đỏ và xanh. Khi ta đặt một quả bóng mới lên bàn, bằng cách xác định xem nó nằm ở phía bên nào ta có thể dự đoán được màu sắc của quả bóng đó.



Hình 1.13 Ví dụ phân lớp giữa hai tập hợp bóng Xanh - Đỏ sử dụng SVM

Nhưng dữ liệu thực tế rất phức tạp. Nếu các quả bóng đan xen vào nhau thì không thể dùng 1 cây que để phân lớp được.

Vì vậy, ta cần đưa các dữ liệu về không gian đa chiều và dùng một siêu phẳng để phân lớp dữ liệu.

Giả sử ở ví dụ Hình 1.13 các quả bóng xanh và đỏ nằm trên mặt phẳng Oxy lần lượt có tọa độ là:

Tọa độ bóng xanh $(x_1, x_2, x_3, \dots, x_i)$ và bóng đỏ $(y_1, y_2, y_3, \dots, y_j)$.

Để phân lớp cho các quả bóng trên ta đưa chúng về không gian đa chiều có dạng:

Bóng xanh $(f(x_1), f(x_2), \dots, f(x_i))$ và bóng đỏ $(f(y_1), f(y_2), \dots, f(y_j))$.

Thay vì phân lớp cho các x_i và y_j ta chuyển sang bài toán phân lớp cho $f(x_i)$ và $f(y_j)$.

Với điều kiện $f(x_i)$ và $f(y_j)$ phải là hàm Kernel.

1.6.3. Đánh giá

Recall được định nghĩa bởi tỉ lệ giữa số lượng các kết quả chính xác (true positive) và số lượng thực có trong tham chiếu. Recall được sử dụng để đo khả năng hệ thống tìm được chính xác đối tượng trong toàn tập dữ liệu, vì vậy đôi khi còn được gọi là độ nhạy (sensitivity).

$$\text{Recall} = \frac{\text{Số lượng kết quả chính xác}}{\text{Tất cả thực có}} = \frac{\text{Relevant} \cap \text{Retrieved}}{\text{Relevant}} \quad (2-5)$$

Chúng tôi cũng áp dụng công thức trên để đánh giá số ảnh được gán đúng nhãn trong quá trình thực nghiệm.

Chương 2. THUẬT TOÁN GÁN NHÃN ẢNH CHỈ SỬ DỤNG CÁC VECTOR ĐẶC TRƯNG

2.1. Ảnh và chú giải (Tags)

Để xây dựng tập dữ liệu NUS-WIDE các giả đã lấy ngẫu nhiên hơn 300.000 hình ảnh cùng với chú giải của nó được chia sẻ trên trang Flickr.com. Bằng cách loại bỏ các ảnh có kích thước quá nhỏ hoặc hình ảnh trùng lặp, cuối cùng tác giả đã giữ lại 269,648 ảnh. Tất cả các ảnh trên được chia làm 81 nhãn (concept). Một ảnh có thể được gán cho một hoặc nhiều nhãn. Danh sách các nhãn được mô tả ở Hình 2.1. Bên cạnh đó, tác giả cũng xây dựng tập thông tin phụ trợ với thông tin chú giải và định vị GPS. Đi kèm với 269,648 ảnh là 425,059 chú giải có nghĩa được gắn kèm. Các chú giải này chính là các từ khoá nằm trong EXIF header của ảnh. Với số lượng chú giải rất lớn (425,059), nhóm tác giả thực hiện việc lọc các chú giải bằng cách loại bỏ các chú giải xuất hiện rất ít hoặc rất nhiều, chú giải là tên riêng, là số, những chú giải không tồn tại trong WordNet. Cuối cùng, chỉ còn lại 1000 chú giải (tham khảo danh sách 1000 chú giải được trình bày ở phần cuối của luận văn), là các danh từ chung chung mức độ xuất hiện vừa phải, được sử dụng làm thông tin phụ trợ dùng cho mục đích phân loại và gán nhãn. Những nhãn này có thể tìm thấy chúng tại <http://lms.comp.nus.edu.sg/research/NUS-WIDE.htm>. Bảng 2.1 đưa ra 20 chú giải (trong 1000 chú giải trên) xuất hiện nhiều nhất cùng với tần suất xuất hiện của chúng.

Một vấn đề quan trọng trong gán nhãn ảnh và lập chỉ mục là mối tương quan giữa ngữ nghĩa trong các khái niệm. Khái niệm ngữ nghĩa không tồn tại độc lập. Thay vào đó, chúng xuất hiện mối tương quan và tương tác tự nhiên với nhau. Ví dụ, “sunset” thường xuất hiện với chú giải ở “sea” nhưng “airplane” và “animal” thì không. Một vài nghiên cứu đã được thực hiện làm thế nào để khai thác mối tương quan ngữ nghĩa để cải thiện việc gán nhãn cho ảnh và video [Fei et al. 2004; Li et al. 2009].

Bảng 2. 1 Chú giải phổ biến nhất trong tập NUS-WIDE

Chú giải	Tần suất	Chú giải	Tần suất
nature	20142	sunset	10962
sky	18935	light	10869
blue	17822	white	10082
water	17646	people	9324
clouds	14201	sea	9016
red	13172	night	8806
green	13169	art	8759
bravo	12003	architecture	8589
landscape	11667	yellow	8191
explore	11144	portrain	8139

Trong trường hợp này, mỗi tương quan ngữ nghĩa có thể dễ dàng thu được bằng cách tính toán ma trận cùng xuất hiện giữa các chú giải. Chúng ta thấy rằng, các ma trận cùng đồng xảy ra là khá đầy đủ có mỗi tương quan chặt chẽ với 5,081 chú giải có nghĩa trong tập dữ liệu ta sử dụng.

2.2. Giới thiệu cách tạo ra 81 nhãn trong tập dữ liệu NUS-WIDE

Để đánh giá hiệu quả của quá trình nghiên cứu được tiến hành trên tập dữ liệu của NUS-WIDE, tác giả đã mời một nhóm sinh viên (được gọi là người chú giải-annotators) xây dựng lên 81 nhãn tham chiếu với các chủ đề khác nhau được gán nhãn thủ công được trình bày ở Hình 2.1. Nhóm tác giả tạo ra 81 nhãn tham chiếu đến từ trường trung học và đại học quốc gia Singapore. 81 nhãn được lựa chọn cẩn thận bằng cách: (a) Chúng phù hợp với những khái niệm được định nghĩa trong nhiều văn cảnh khác nhau [Barnard et al. 2003; Fei et al. 2004; Naphade et al. 2006; Snoek et al, 2006]; (b) Hầu hết chúng tương ứng với các chú giải thường xuất hiện trên Flickr; (c) Chúng có cả các khái niệm chung như "animal" và các khái niệm cụ thể như "dog" hay "flowers"; (d) Chúng thuộc các thể loại khác nhau bao gồm cảnh, đối tượng, sự kiện, chương trình, con người và đồ họa.

Hướng dẫn gán nhãn được thực hiện theo cách sau: Nếu người gán nhãn thấy một nhãn nào đó tồn tại trong hình ảnh, nhãn đó sẽ được gán cho ảnh; nếu

nhãn đó không tồn tại trong ảnh hoặc chú thích không chắc chắn về việc liệu hình ảnh có chứa các nhãn đó, thì nhãn đó sẽ không được gán cho ảnh. Hình 2.2 cho thấy số lượng hình ảnh có liên quan đến 81 nhãn.

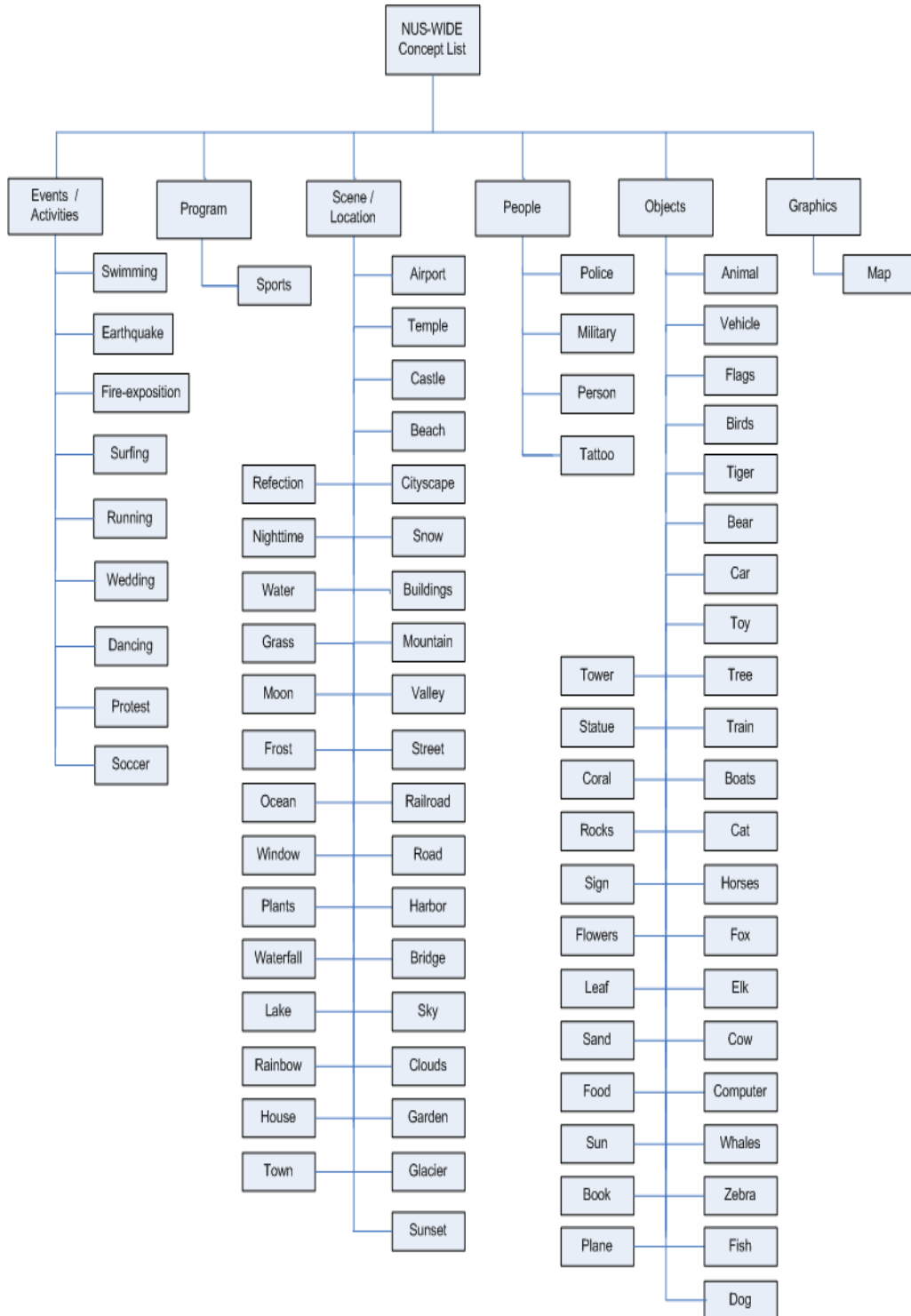
Vì có 269,648 hình ảnh trong tập dữ liệu, nó gần như không thể được gán nhãn thủ công cho tất cả các hình ảnh với 81 nhãn tham chiếu. Do đó tác giả đã xây dựng một hệ thống để tìm ra càng nhiều càng tốt hình ảnh liên quan đến mỗi nhãn để hỗ trợ việc gán nhãn thủ công. Gán nhãn thủ công được tiến hành từng cái một đối với tất cả các nhãn. Ở đây, tác giả giới thiệu ngắn gọn quy trình gán nhãn cho một ảnh. Đầu tiên, tất cả các ảnh đã được gán từ chú giải với nhãn được hiển thị cho người chú giải xác nhận thủ công. Sau bước này, tác giả thu được tham chiếu cho một phần nhỏ của tập dữ liệu. Thứ hai, tác giả sử dụng phần tham chiếu này như dữ liệu huấn luyện để thực hiện thuật toán K-NN cho các ảnh chưa được gán nhãn. Các hình ảnh chưa được gán nhãn này được sắp xếp theo giá trị thu được bởi K-NN. Thứ ba, tác giả tiến hành sắp xếp danh sách hình ảnh để người chú giải gán nhãn thủ công cho đến khi người chú giải không thể tìm thấy bất kỳ hình ảnh liên quan nào trong 200 ảnh liên tiếp. Trung bình, người chú giải tự xem và chú thích thủ công khoảng một phần tư của tất cả ảnh. Tuy nhiên, đối với một số nhãn phổ biến nào đó như “sky” và “animal”, người chú giải có thể gán nhãn gần như toàn bộ tập dữ liệu. Tác giả tin rằng giá trị tham chiếu được tạo ra là hợp lý vì phần còn lại là ba phần tư hình ảnh không nhìn thấy rất có thể sẽ không chứa những nhãn theo tiêu chí lựa chọn của tác giả.

Tác giả ước tính rằng nỗ lực tổng thể cho việc bán gán nhãn thủ công của 81 nhãn tham chiếu mất khoảng 3,000 giờ làm việc. Để thuận lợi cho việc đánh giá tác giả chia tập dữ liệu thành hai phần. Phần thứ nhất chứa 161,789 ảnh sử dụng làm ảnh huấn luyện và phần thứ hai chứa 107,859 ảnh sử dụng làm ảnh kiểm thử.

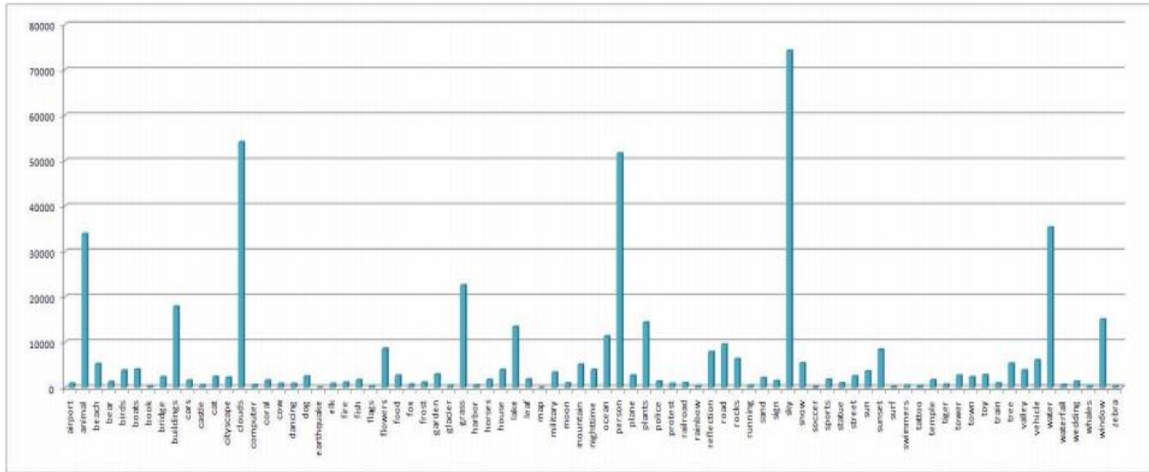
2.3. Biểu diễn ảnh dưới dạng vector đặc trưng

Để thực hiện nhiệm vụ phân loại và gán nhãn ảnh, các bức ảnh trong tập dữ liệu trên được biểu diễn lại thành các vector đặc trưng. 6 loại đặc trưng được

sử dụng để biểu diễn ảnh là đặc trưng màu sắc (Color Histogram, Color Moment, Color Correlation), Wavelet texture, Edge direction và BOW (500 words). Để tiết kiệm thời gian, chúng tôi sử dụng trực tiếp các vector đặc trưng được cung cấp bởi nhóm tác giả. Link download tập biểu diễn các đặc trưng trên có thể tìm thấy tại: <http://lms.comp.nus.edu.sg/research/NUS-WIDE.htm>.



Hình 2. 1 81 nhãn trong tập dữ liệu của NUS-WIDE



Hình 2. 2 Số hình ảnh có liên quan đến 81 nhãn của NUS-WIDE

2.4. Thuật toán K-NN

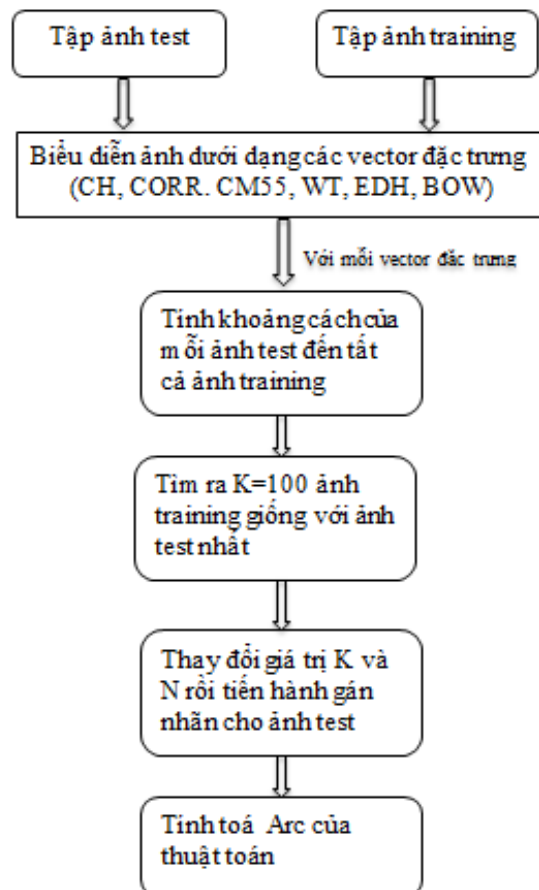
2.4.1. Phân tích dữ liệu thực nghiệm với K-NN

Bộ dữ liệu của NUS-WIDE gồm 269,648 ảnh thuộc 81 nhãn (concepts). Trong phân thực nghiệm lấy ngẫu nhiên 161,789 ảnh dùng làm tập huấn luyện và 107,859 ảnh dùng làm tập kiểm chứng thuật toán. Với bộ dữ liệu này độ phức tạp của thuật toán chủ yếu ở phần tính toán khoảng cách của mỗi ảnh kiểm thử với toàn tập huấn luyện. Nếu thực hiện trên cả tập ảnh này, chúng ta sẽ phải tính $161,789 \times 107,859$ khoảng cách cho mỗi loại đặc trưng. Đây là khối lượng tính toán khổng lồ rất khó xử lý, tốn rất nhiều thời gian và bộ nhớ. Vì vậy, chúng tôi tiến hành giảm bớt lượng ảnh huấn luyện xuống còn 50 ảnh/ nhãn bằng phương pháp K-means. Như vậy tập ảnh huấn luyện chúng tôi dùng để tiến hành thực nghiệm chỉ còn 4,032 ảnh. Tập ảnh để kiểm thử bao gồm 107,859, với mỗi đặc trưng ta tiến hành tính khoảng cách của mỗi ảnh kiểm thử đến 4,032 ảnh huấn luyện và tìm ra 100 ảnh huấn luyện giống nó nhất. Tiếp theo, chúng tôi chọn giá trị $K = 50$ và tính khoảng cách từ ảnh kiểm thử đến 100 ảnh huấn luyện giống nó nhất để tìm ra 50 ảnh huấn luyện giống với ảnh kiểm thử nhất. Sau khi tính khoảng cách và sắp xếp theo thứ tự tăng dần sẽ chọn ra 10 ảnh huấn luyện cuối cùng giống nó nhất và tiến hành gán nhãn của ảnh huấn luyện cho ảnh kiểm thử với số nhãn thay đổi từ 1 đến 10. Cuối cùng, chúng tôi tiến hành đánh giá kết quả của thuật toán bằng cách tính giá trị recall trung bình khi số nhãn gán cho ảnh kiểm thử thay đổi từ 1 đến 10.

Một số từ viết tắt cần lưu ý:

- D là số chiều của vector đặc trưng.
- CH - Vector lược đồ màu - 64-D color histogram.
- CORR - Vector tương quan màu - 144-D color auto-correlogram.
- CM55 - Vector mô men màu - 225-D block-wise color moments.
- WT - Vector 128-D wavelet texture.
- EDH - Vector lược đồ hệ số góc -73-D edge direction histogram.
- BOW - Vector 500-D Bag of word.
- N là số nhãn (concept) được gán cho ảnh test.
- Arc giá trị recall trung bình của nhãn.

Sơ đồ các bước thực hiện quá trình phân loại và gán nhãn cho ảnh kiểm thử bằng thuật toán KNN:



Hình 2. 3 Sơ đồ tóm tắt các bước thực hiện quá trình phân loại và gán nhãn cho ảnh kiểm thử sử dụng K-NN

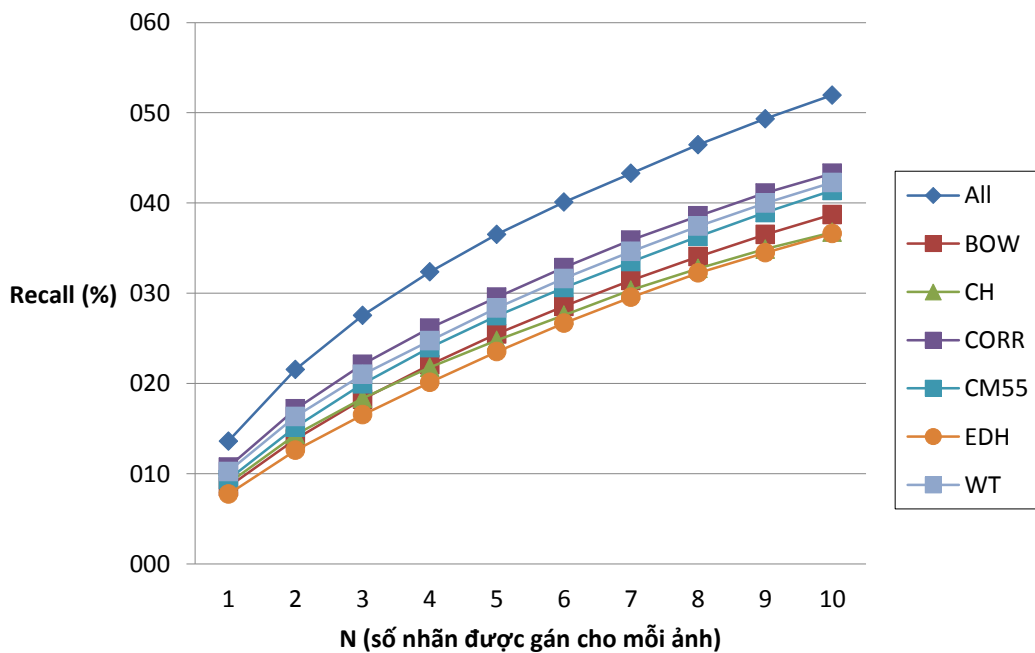
2.4.2. Kết quả recall của K-NN

Chúng tôi tiến hành đánh giá thuật toán bằng cách tính giá trị recall trung bình của thuật toán theo công thức (2-5).

Kết quả recall của K-NN được thể hiện ở Bảng 2.2

Bảng 2. 2 Kết quả recall của các đặc trưng khi sử dụng thuật toán K-NN

N	All	BOW	CH	CORR	CM55	EDH	WT
1	13,62	8,57	8,95	10,74	9,35	7,74	10,25
2	21,56	13,84	14,34	17,18	15,17	12,60	16,34
3	27,56	18,26	18,37	22,13	19,97	16,53	21,01
4	32,36	22,07	21,79	26,13	23,98	20,13	24,74
5	36,54	25,50	24,78	29,56	27,50	23,53	28,38
6	40,10	28,55	27,57	32,84	30,63	26,69	31,63
7	43,29	31,42	30,35	35,88	33,50	29,55	34,62
8	46,46	34,07	32,75	38,56	36,31	32,25	37,43
9	49,34	36,51	34,90	41,11	38,95	34,48	39,98
10	51,96	38,73	36,77	43,28	41,39	36,61	42,27



Hình 2. 4 Biểu đồ so sánh kết quả recall của 6 đặc trưng và kết hợp các đặc trưng với $K = 50$ và số nhãn gán cho ảnh (N) tăng dần từ 1 đến 10

Nhận xét: Căn cứ vào kết quả trên cho thấy, với tập dữ liệu của NUS-WIDE thì việc kết hợp các đặc trưng cho kết quả tốt hơn hẳn việc chỉ sử dụng từng đặc trưng riêng biệt. Do mỗi đặc trưng có ưu nhược điểm riêng nên việc kết hợp này có thể phát huy được ưu điểm và khắc phục nhược điểm của mỗi loại đặc trưng làm cho kết quả cải thiện hơn.

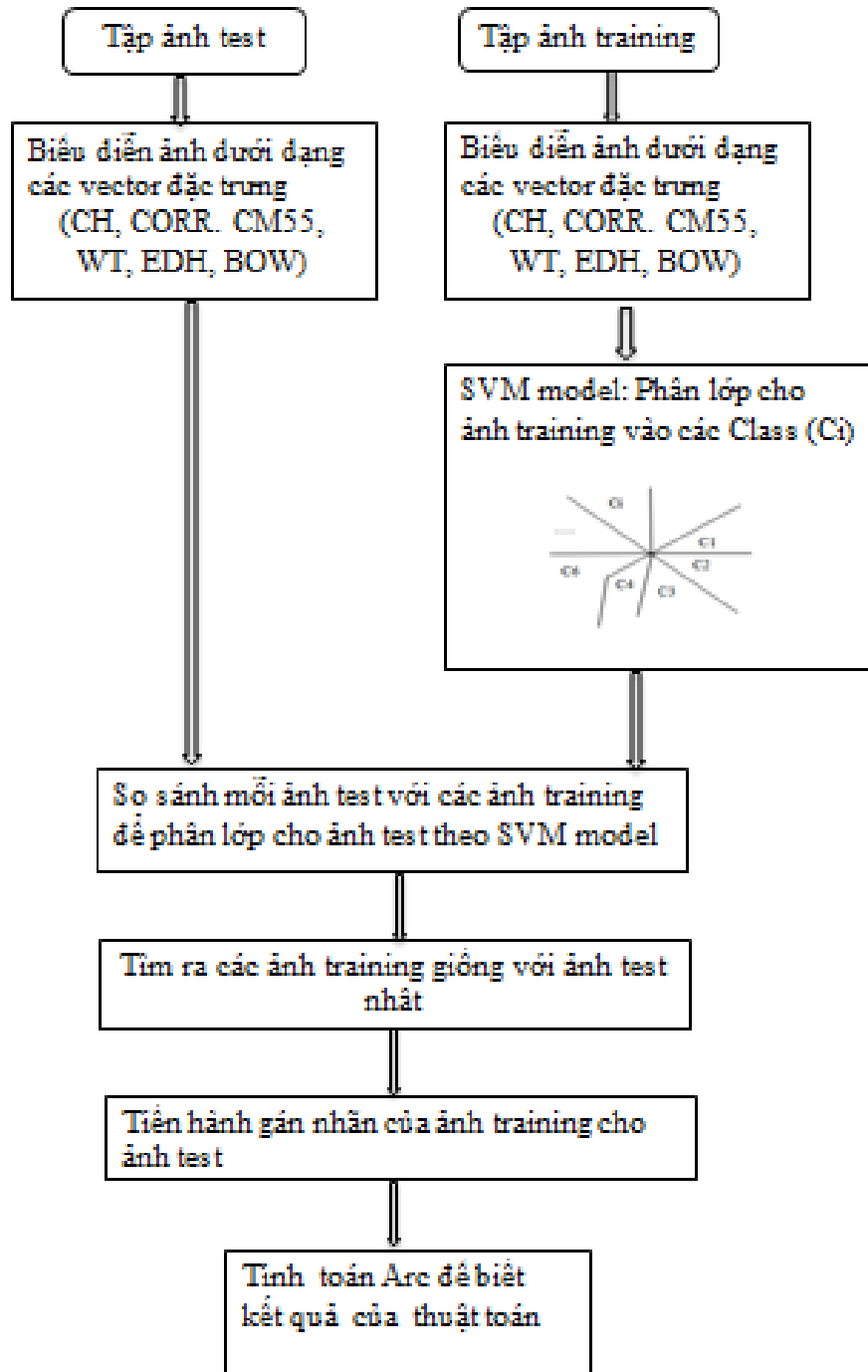
Về thời gian xử lý của hệ thống: Do số lượng ảnh nhiều, tất cả các loại đặc trưng và ghép các loại đặc trưng đều được tính toán từ trước (tiền xử lý); ma trận khoảng cách giữa các đặc trưng cũng được chia để tính toán song song. Thuật ngữ chỉ sử dụng thông tin về khoảng cách giữa các đặc trưng và cho ra thông tin nhãn được gán nên yếu tố thời gian tính toán xử lý không được xem xét.

2.5. Thuật toán SVM

2.5.1. Phân tích dữ liệu thực nghiệm với SVM

Thuật toán SVM được thực hiện dựa trên Matlab toolbox LIB-SVM sử dụng Gaussian kernel. Ban đầu chúng tôi tiến hành huấn luyện trên tập đã được scales thông qua K-means gồm 4,032 huấn luyện và kiểm thử trên toàn tập 107,859 ảnh. Kết quả không cải thiện nhiều so với thuật toán K-NN. Điều này có thể giải thích là do khi thực hiện scales tập huấn luyện thông qua thuật toán K-means ta đã tiến hành lấy trung bình các ảnh trong cùng một cluster. Quá trình này làm cho việc thực hiện ước lượng mô hình phân loại trong thuật toán SVM gặp khó khăn. Do đó, chúng tôi quay lại tiến hành SVM trên chính các ảnh huấn luyện. Tuy nhiên để giảm bớt lượng ảnh huấn luyện, chúng tôi chọn ngẫu nhiên tối đa 200 ảnh/ nhãn để làm tập huấn luyện, và kiểm thử trên toàn bộ ảnh.

Sơ đồ các bước thực hiện gán nhãn ảnh với mô hình SVM:



Hình 2. 5 Sơ đồ tóm tắt các bước thực hiện quá trình phân loại và gán nhãn cho ảnh kiểm thử sử dụng SVM

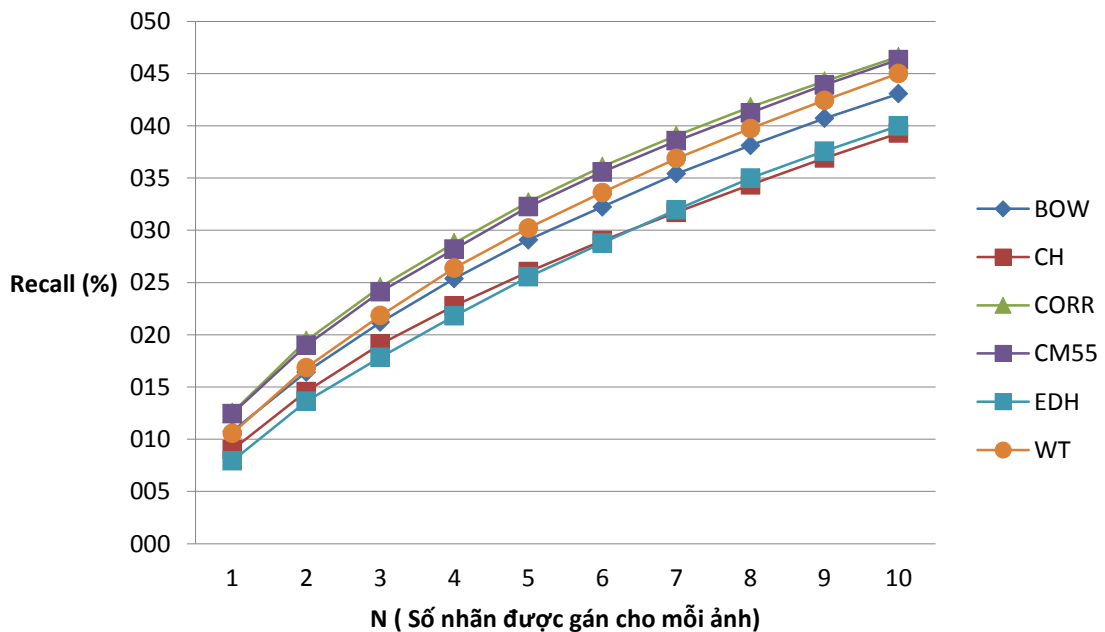
2.5.2. Kết quả recall của SVM

Chúng tôi cũng dùng công thức tính giá trị recall trung bình như thuật toán KNN (2-5) để đánh giá kết quả của SVM.

Kết quả recall của SVM cũng được thể hiện ở bảng 2.3

Bảng 2. 3 Kết quả recall của các đặc trưng khi sử dụng thuật toán SVM

N	BOW	CH	CORR	CM55	EDH	WT
1	10,77	9,00	12,55	12,45	7,93	10,57
2	16,43	14,56	19,45	18,99	13,63	16,85
3	21,19	19,11	24,58	24,12	17,85	21,84
4	25,38	22,77	28,79	28,19	21,81	26,38
5	29,08	26,03	32,70	32,26	25,55	30,22
6	32,24	29,01	36,10	35,58	28,78	33,62
7	35,41	31,72	39,07	38,57	31,97	36,87
8	38,13	34,35	41,82	41,25	35,00	39,74
9	40,71	36,91	44,26	43,92	37,58	42,44
10	43,07	39,30	46,60	46,33	40,00	45,01

**Hình 2. 6 Biểu đồ so sánh kết quả recall của 6 đặc trưng sử dụng thuật toán SVM.**

Nhận xét: Mô hình máy vector hỗ trợ- SVM có độ phức tạp hơn so với K-NN, nhưng với tập dữ liệu NUS-WIDE thì kết quả thực nghiệm cũng không cao hơn nhiều so với K-NN. Qua kết quả thực nghiệm cho thấy, với cả hai phương pháp K-NN và SVM thì đặc trưng CORR cho kết quả tốt nhất và đặc trưng EDH cho kết quả kém nhất. Điều này chứng tỏ, kết quả của quá trình phân loại và gán nhãn không phụ thuộc vào phương pháp thực hiện chúng khi thực nghiệm với tập dữ liệu NUS-WIDE.

Chương 3: THUẬT TOÁN SỬ DỤNG KẾT HỢP THÔNG TIN PHỤ TRỢ

3.1. Sử dụng thông tin chú giải

Chú giải (tags), là các từ ngữ được người dùng gán kèm ảnh khi đăng tải ảnh trên mạng (VD như trên trang flickr.com) hoặc được lọc tại các vùng xung quanh xuất hiện của ảnh tại các website chứa ảnh. Đối với tập dữ liệu NUS-WIDE, sau khi tải về và lọc; nhóm tác giả đã giữ lại 1,000 từ hay chú giải sử dụng làm thông tin phụ trợ. Phần này sẽ trình bày cách sử dụng thông tin chú giải này trong việc cải thiện độ chính xác khi gán nhãn ảnh.

3.1.1. Phân tích mối quan hệ giữa chú giải và nhãn

Chú giải và nhãn (Tags và Concepts đều là các từ Tiếng Anh có ngữ nghĩa riêng và giữa chúng đều có mối liên hệ với nhau). Ví dụ như các chú giải: “airplane, aircraft, plane, jet...” đều có mối liên hệ rất lớn đến nhãn chủ đề “airport”, nhưng các chú giải “mountain” hay “animal” thì lại không hề liên hệ đến nhãn “airport”. Để xác định được mối liên hệ này có thể sử dụng các kỹ thuật xử lý ngôn ngữ tự nhiên hoặc hệ thống từ điển WordNet. Tuy nhiên dù đánh giá đúng mối liên hệ này thì chúng ta cũng đã bỏ qua ngữ cảnh của bức ảnh. Ví dụ, bức ảnh có thể chụp một “mountain” bên cạnh “airport”. Việc chú giải “mountain” không sai; nhưng nếu ta thấy chú giải “mountain” không liên quan đến “airport” mà bỏ “airport” đi thì sẽ gây ra sai sót. Hơn nữa quá trình tính toán trên lại bỏ qua việc phân tích thông tin của bản thân tập dữ liệu. Ở đây, chúng tôi đề xuất việc tính toán mối liên hệ giữa chú giải và nhãn dựa trên chính tập huấn luyện của NUS-WIDE.

Qua phân tích tập huấn luyện hơn 160,000 ảnh của NUS-WIDE chúng tôi thấy:

- Mỗi ảnh đều có một vài chú giải, nằm trong danh sách 1000 chú giải của thông tin phụ trợ.

- Các ảnh thuộc cùng nhãn thường sẽ được gán bằng các chú giải giống nhau. Do đó, thông qua việc tìm kiếm các chú giải chung, xuất hiện thường xuyên; ta có thể tìm ra mối liên hệ đến nội dung của ảnh. Qua mức độ xuất hiện của các chú giải này đối với mỗi nhãn cụ thể; ta hoàn toàn có thể xác định được

giá trị (score) - chỉ ra mức độ liên hệ của mỗi chú giải với mỗi nhãn. Thuật toán tính giá trị $score_tag_concept$ như sau:

for each concept_i

- Xác định tập S_i chứa tất cả các ảnh huấn luyện có nhãn là $concept_i$

- Xác định tổng $tag(i,:)$ tất cả các tag của tập S_i ; $tag(i,j)$ là số lần xuất hiện của tag_j của tất cả các ảnh trong tập S_i .

- $Score_tag_concept(i,j) = normalized(tag(i,j))$;

End

Sau khi thực hiện thuật toán, ta có thể xác định được mối liên hệ giữa các chú giải và nhãn. Bảng 3.1 chỉ ra một số ví dụ về kết quả khảo sát mối liên hệ giữa chú giải - nhãn:

Bảng 3.1 Ví dụ một số kết quả khảo sát mối liên hệ giữa chú giải (tag) và nhãn (concept)

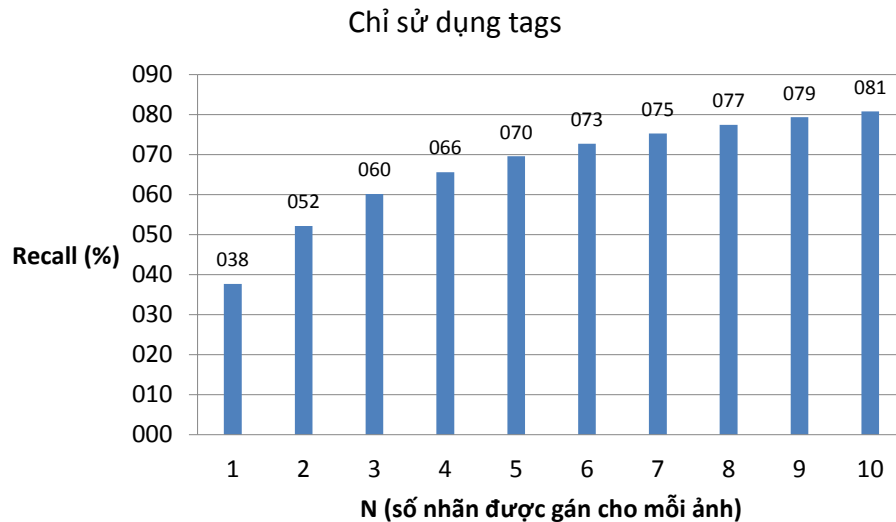
Concepts	Top most relevance tags
1-Airport	Airport, Airplane, Aircraft, Aviation, Plane, Jet, Flying, Airplanes, Wings
2-Animal	Nature, Animal, Wildlife, Animals, Zoo, Bird, Cat, Dog, Bravo, Cute
6.Boat	Boats, Boat, Water, Sea, Sunset, Sky, Clouds, Ship, Blue, Beach
8.Bridge	Bridge, Water, River, Night, Sky, Bridges, City, Clouds, Blue, Reflection.
23. Fish	Fish, Underwater, Scuba, Aquarium, Diving, Water, Sea, Ocean, Coral
40. Moutain	Moutain, Lanscape, Moutains, Nature, Clouds, Snow, Sky, Blue, Water

Có thể dễ dàng thấy kết quả thu được thông qua việc khảo sát trên toàn bộ tập huấn luyện cho ra kết quả tương đối chính xác mối liên hệ giữa chú giải và nhãn của bức ảnh. Ví dụ như, nhãn "Person" các chú giải "People, Girl, Wonman, Man, Child, Boy" đều có giá trị đánh giá mức liên quan đến chủ đề lớn.

Điều này có nghĩa là: Nếu một ảnh kiểm thử cũng có các chú giải này, nó sẽ có xác suất lớn thuộc nhãn "*Person*". Hay đối với nhãn "*Fish*" các chú giải "*Fish, Underwater, Ocean, Aquarium, Sea*" đều có liên hệ trực tiếp đến chủ đề của bức ảnh và được gán thông số giá trị - score lớn. Một bức ảnh kiểm thử có các chú giải trên khả năng cao sẽ thuộc nhóm Fish.

3.1.2. Thực nghiệm sử dụng thông tin chú giải của người dùng trên tập kiểm thử

Chúng tôi tiến hành đánh giá kết quả khi sử dụng thông tin chú giải của người dùng trên tập kiểm thử. Với mỗi ảnh kiểm thử, một danh sách các chú giải được lấy ra và tính giá trị của chúng đối với mỗi nhãn. Giá trị của ảnh kiểm thử với từng chủ đề sẽ bằng tổng của giá trị của tất cả các chú giải. Ở đây, chúng tôi không tiến hành chuẩn hoá cho số lượng chú giải vì nếu ảnh có nhiều chú giải liên quan đến nhãn thì nên có giá trị lớn hơn ảnh chỉ có 1 hoặc 2 chú giải có liên hệ đến nhãn đó. Sau đó, bằng việc sắp xếp theo giá trị giảm dần, dễ dàng có thể chọn ra 10 nhãn có giá trị lớn nhất hay có xác suất cao nhất là chủ đề của ảnh kiểm thử. Kết quả thể hiện ở hình 3.1.



Hình 3. 1 Biểu đồ Recall khi sử dụng thông tin chú giải của người dùng trên tập kiểm thử.

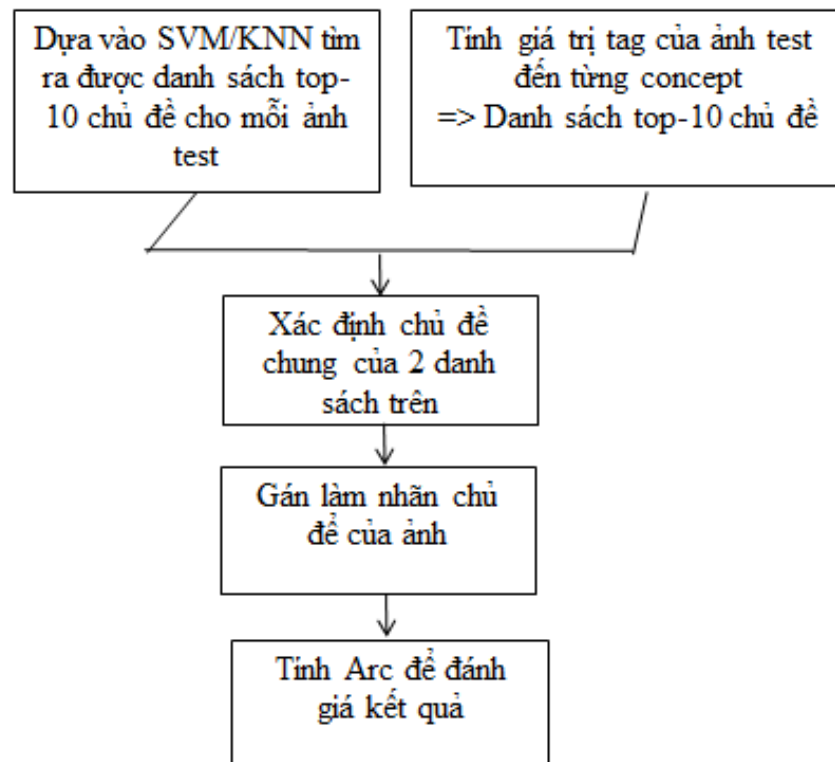
Nhận xét: Kết quả recall khi sử dụng chú giải của người dùng trên tập kiểm thử rất cao. Điều đó chứng tỏ, với tập dữ liệu NUS-WIDE, những chú giải người dùng cho ảnh có chất lượng rất tốt cho quá trình gán nhãn cho ảnh.

3.2. Sử dụng vector đặc trưng và chú giải

Kết quả ở trên gợi ý việc sử dụng thông tin chú giải trong hiệu chỉnh kết quả khi kết hợp với các đặc trưng của ảnh. Ta tiến hành thực hiện trên hai cách đó là dựa trên thông tin chú giải của ảnh kiểm thử và dựa trên chú giải của ảnh huấn luyện.

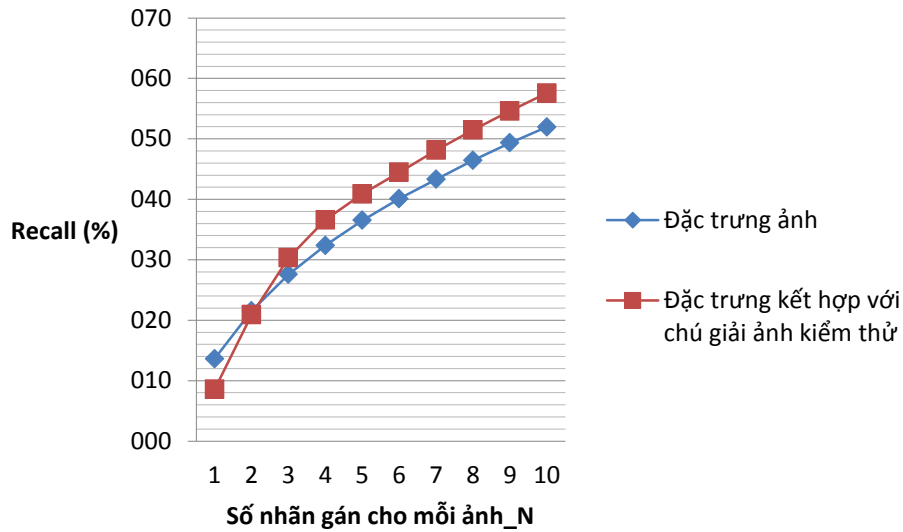
3.2.1. Dựa trên thông tin chú giải của ảnh kiểm thử

Với mỗi ảnh kiểm thử, dựa trên thông tin hình ảnh ta đã xác định được danh sách nhóm 10 nhãn (căn cứ theo thuật toán K-NN hoặc SVM) có xác suất cao nhất là chủ đề. Dựa theo danh sách các chú giải của chính ảnh kiểm thử đó, ta cũng tính được score_concept_tag của ảnh đó với từng nhãn cụ thể (như đã trình bày ở phần trước). Kết hợp 2 danh sách này, ta có thể xác định được những nhãn chung, đó là nhãn chủ đề của ảnh. Trong trường hợp số lượng nhãn chung $< N$ hoặc không có nhãn nào chung; ta lựa chọn ngẫu nhiên ($N -$ số nhãn chung) hoặc cả N nhãn (khi không có nhãn chung nào) trong cả 2 danh sách và gán làm các nhãn chủ đề của ảnh.



Hình 3.2 Sơ đồ khối quá trình kết hợp đặc trưng ảnh và chú giải của ảnh kiểm thử.

Kết quả recall dựa trên thông tin hình ảnh kiểm thử thể hiện ở Hình 3.3.

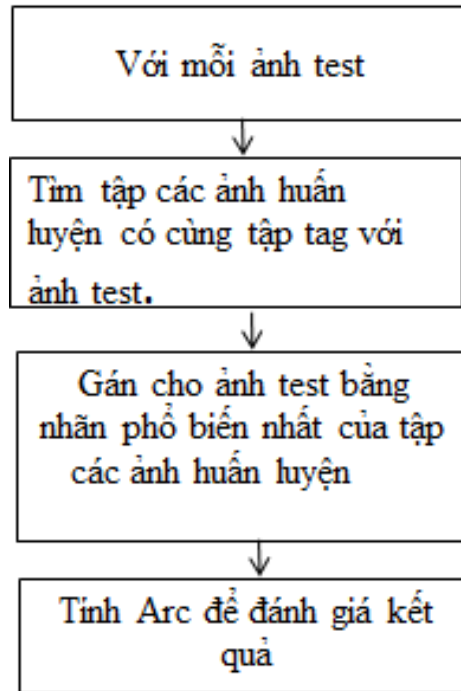


Hình 3.3 Biểu đồ so sánh kết quả recall của đặc trưng ảnh và đặc trưng kết hợp với chú giải của ảnh kiểm thử.

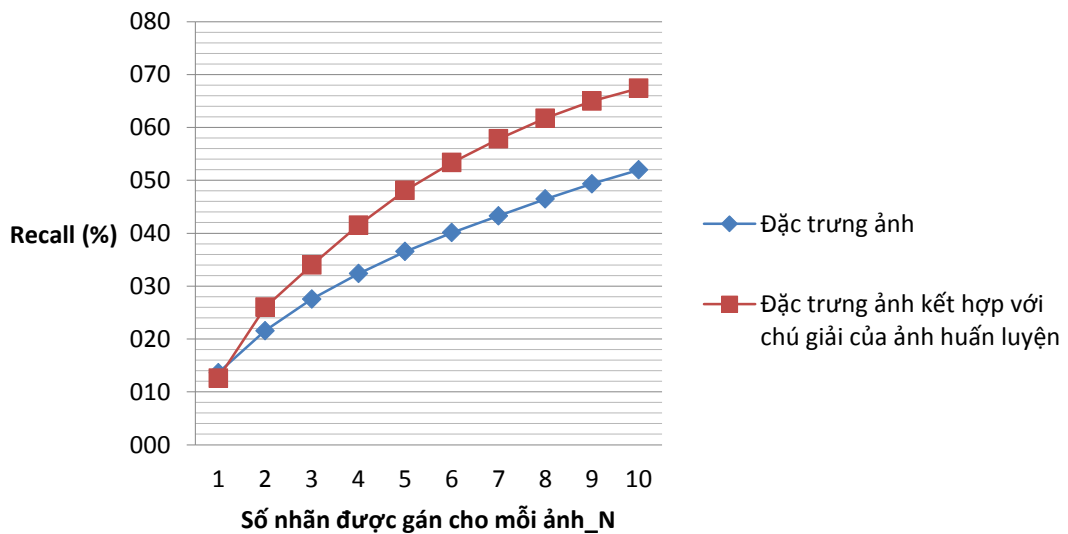
Nhận xét: Từ kết quả Hình 3.3 cho thấy, việc kết hợp các đặc trưng ảnh với thông tin chú giải của chính ảnh kiểm thử cho kết quả recall cao hơn chỉ sử dụng các đặc trưng ảnh với số điều kiện số nhãn gán cho mỗi ảnh phải lớn hơn 3. Còn với số nhãn được gán cho ảnh bằng 1 hoặc 2 thì kết quả kết hợp lại thấp hơn do số lượng nhãn chung $< N$ hoặc không có nhãn nào chung; ta lựa chọn ngẫu nhiên ($N -$ số nhãn chung) hoặc cả N nhãn (khi không có nhãn chung nào) trong cả 2 danh sách và gán làm các nhãn chủ đề của ảnh.

3.2.2. Dựa trên thông tin chú giải của ảnh huấn luyện

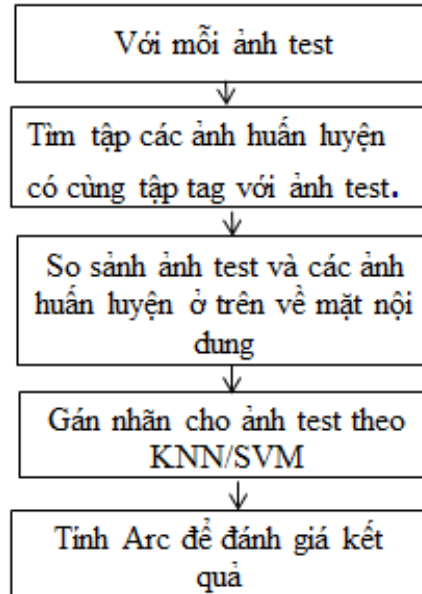
Với mỗi ảnh kiểm thử: (a) tìm tập gồm K -NN các ảnh huấn luyện có cùng chung (hoặc chung nhiều nhất) tập chú giải giống với nó. Đến đây ta có thể gán cho ảnh kiểm thử bằng nhãn phổ biến nhất của tập các ảnh huấn luyện kết quả thực nghiệm như Hình 3.5 hoặc (b) từ tập K -NN ảnh ảnh huấn luyện trên, so sánh chúng về mặt các vector đặc trưng; chọn ra các top K -NN/2 giống nó nhất, xác định nhãn phổ biến nhất của tập trên và gán nó cho ảnh kiểm thử kết quả thể hiện ở Hình 3.7.



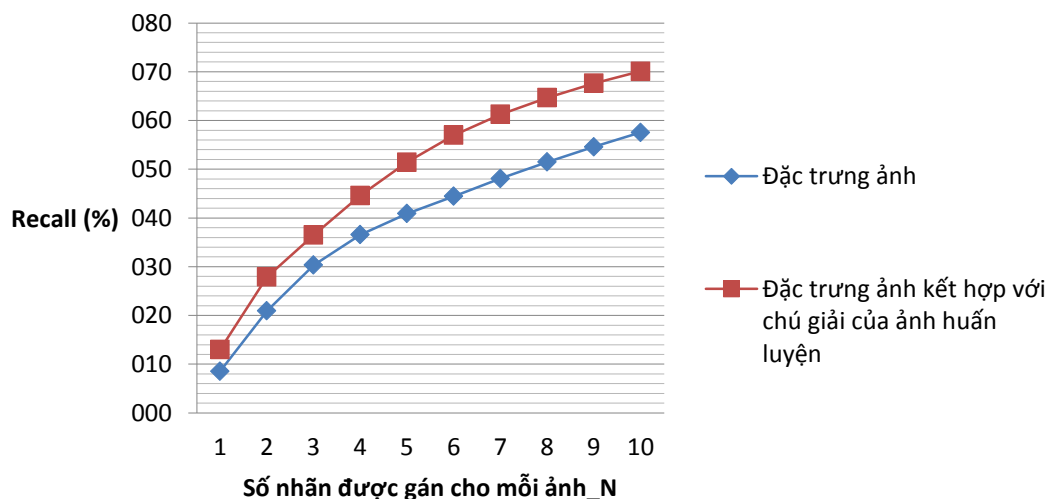
Hình 3. 4 Sơ đồ khối dựa trên thông tin chú giải của ảnh huấn luyện theo cách a



Hình 3. 5 Biểu đồ so sánh kết quả recall của đặc trưng ảnh và đặc trưng ảnh kết hợp với chú giải của ảnh huấn luyện theo cách a



Hình 3. 6 Sơ đồ khối dựa trên thông tin chú giải của ảnh huấn luyện theo cách b



Hình 3. 7 Biểu đồ so sánh kết quả recall của đặc trưng ảnh và đặc trưng ảnh kết hợp với chú giải của ảnh huấn luyện theo cách b

Nhận xét: Từ kết quả recall của Hình 3.5 và Hình 3.7 cho thấy phương pháp kết hợp các vector đặc trưng của ảnh với chú giải của ảnh huấn luyện cho kết quả cải thiện hơn rất nhiều so với việc chỉ sử dụng vector đặc trưng để gán nhãn tự động cho ảnh. Việc sử dụng chú giải của ảnh huấn luyện để tìm ra nhãn chung nhất để gán cho ảnh kiểm thử cũng cho kết quả cao hơn việc sử dụng chính chú giải của ảnh kiểm thử để tìm nhãn chung nhất. Tuy nhiên, với phương pháp này thì số nhãn gán cho ảnh kiểm thử bằng 1 vẫn chưa cho kết quả tốt.

KẾT LUẬN

Với số lượng dữ liệu ảnh đa dạng và phong phú trên Internet như ngày nay, nhu cầu về tìm kiếm ảnh để phục vụ cho mục đích của từng người dùng là rất cần thiết. Người dùng vẫn có thói quen tìm kiếm ảnh bằng từ khóa. Trước nhu cầu đó chúng tôi tiến hành nghiên cứu phương pháp tự động gán nhãn cho ảnh. Việc thực hiện những nghiên cứu về tự động gán nhãn cho ảnh đã được quan tâm từ lâu, nhưng lĩnh vực này vẫn còn nhiều vấn đề cần phải cải tiến để đạt được kết quả tốt nhất. Để cải tiến vấn đề này chúng tôi tiến hành nghiên cứu sử dụng các vector đặc trưng kết hợp với thông tin phụ trợ của ảnh.

Thứ nhất, chúng tôi đã tiến hành khảo sát kết quả khi sử dụng từng đặc trưng để đánh giá xem đặc trưng nào cho kết quả tốt nhất và đặc trưng nào cho kết quả xấu nhất trên tập dữ liệu NUS-WIDE. Mục đích của thực nghiệm là gán cho tập ảnh kiểm thử các nhãn có liên quan trực tiếp đến nội dung của ảnh nhất và số nhãn được gán cho ảnh là ít nhất nhưng lại có mức độ liên quan đến các ảnh cao nhất để có thể tìm kiếm được tối đa ảnh mà người dùng mong muốn tìm kiếm. Qua kết quả khảo sát thì đặc trưng tương quan màu (CORR) cho kết quả tốt nhất và đặc trưng về lược đồ hệ số góc (EDH) cho kết quả kém nhất. Kết quả

này cho thấy sự tương quan giữa các cặp màu trong ảnh có vai trò rất quan trọng trong việc phân loại và gán nhãn.

Thứ hai, chúng tôi đã tiến hành thực nghiệm với việc kết hợp các vector đặc trưng và kết quả của nó cũng tăng đáng kể so với việc chỉ sử dụng các đặc trưng riêng biệt. Do mỗi đặc trưng có ưu điểm và nhược điểm riêng, nên việc kết hợp chúng với nhau có thể dùng ưu điểm của đặc trưng này để khắc phục nhược điểm của đặc trưng khác. Ví dụ như với lược đồ màu (CH) ưu điểm của nó là bất biến với phép quay và dịch nhưng nó lại không thể phân biệt các cặp màu được như đặc trưng tương quan màu CORR. Hai đặc trưng này có thể khắc phục nhược điểm của nhau để quá trình phân loại ảnh được tốt hơn.

Thứ ba, chúng tôi không chỉ dừng lại ở việc kết hợp các đặc trưng với nhau, chúng tôi còn dùng kết quả của việc kết hợp các đặc trưng ở trên kết hợp với thông tin phụ trợ khai khác được từ internet như chú giải để cải thiện thuật toán. Với việc phân tích mối liên hệ giữa 1,000 chú giải và 81 nhãn trong thực nghiệm. Làm cho thuật toán đơn giản hơn, rút ngắn thời gian tính toán hơn thay vì phải thực hiện từng ảnh kiểm thử đối với toàn bộ tập ảnh huấn luyện, chúng tôi chỉ cần so sánh ảnh kiểm thử với K-NN các ảnh huấn luyện có chung tập chú giải nhất với nó để chọn ra nhãn chúng nhất gán cho ảnh kiểm thử. Thuật toán kết hợp với chú giải đơn giản hơn, thời gian thực nghiệm nhanh hơn, kết quả tốt hơn so với việc chỉ sử dụng các vector đặc trưng.

Tuy nhiên, việc kết hợp thông tin chú giải với các đặc trưng của ảnh cho kết quả cải thiện hơn việc chỉ sử dụng các đặc trưng. Nhưng theo như kết quả của khảo sát ở trên (Hình 3.1) thì nó không cao bằng việc chỉ sử dụng thông tin chú giải. Điều đó chứng tỏ với tập dữ liệu của NUS-WIDE, thông tin chú giải mang lại hiệu quả rất cao cho quá trình gán nhãn tự động. Đồng nghĩa với kết quả đó, chúng tôi kết luận rằng việc kết hợp giữa các đặc trưng của ảnh với chú giải chỉ hiệu quả cho các ảnh có chất lượng chú giải thấp.

Đề xuất một số vấn đề cần nghiên cứu tiếp trong tương lai:

- Sử dụng thêm các thông tin phụ trợ khác khác thác được từ internet như định vị GPS, thời gian, bình luận ... vào quá trình phân loại và gán nhãn tự động cho ảnh.

- Xây dựng hệ thống tìm kiếm ảnh với gợi ý là 81 nhãn trong tập dữ liệu của NUS-WIDE để người dùng có thể tìm kiếm được hết nhưng ảnh có liên quan với từ khóa tương ứng.

- Thực nghiệm trên nhiều tập ảnh có kích thước lớn hơn và đa dạng hơn.

TÀI LIỆU THAM KHẢO

1. S. N. Altman (1992), *An introduction to kernel and nearest-neighbor nonparametric regression*, The American Statistician, 46(3), pp. 175-185.
2. Pradeep K. Atrey, M. Anwar Hossain, Abdulmotaleb El Saddik, and Mohan S. Kankanhalli (2010), *Multimodal Fusion for Multimedia Analysis: A Survey*, Multimedia Systems 16, pp. 345-379.
3. Lamberto Ballan, Tiberio Uricchio, Lorenzo Seidenari, and Alberto Del Bimbo (2014), *A Cross-media Model for Automatic Image Annotation*, In Proc. of ACM ICMR, pp. 73-80.
4. K. Barnard, P. Duygulu, D. Forsyth, N. de Freitas, D. M. Blei, and M. I. Jordan (2003), *Matching words and pictures*, *Journal of Machine Learning Research*, pp. 1107 - 1135.
5. W. Biblack et al (1993), *Query image by content, using color, texture, and shape*, SPIE conference on Storage and Retrieval for Image and Video Database, Vol.1908, pp.173-187.

6. David M. Blei, Andrew Y. Ng, and Michael I. Jordan (2003), *Latent Dirichlet Allocation*, Journal of Machine Learning Research 3 , pp. 993-1022.
7. F. Blog.<http://blog.flickr.net/en/2007/05/29/were-going-down>.
8. L. Chen, D. Xu, I. W. Tsang, and J. Luo (2010), *Tag-based web photo retrieval improved by batch mode re-tagging*, In CVPR.
9. Lin Chen, Dong Xu, Ivor W. Tsang, and Jiebo Luo (2012), *Tag-Based Image Retrieval Improved by Augmented Features and Group-Based Refinement*, IEEE Transactions on Multimedia, pp. 1057-1067.
10. Cortes. C and Vapnik. V (1995), *Support vector machine*, Machine Learning, 20(3), pp. 273-297.
11. Kun Duan, David J. Crandall, and Dhruv Batra (2014), *Multimodal Learning in Loosely-organized Web Images*, In Proc. of CVPR, pp. 2465-2472.
12. L. Fei-Fei, R. Fergus, and P. Perona (2004), *Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories*, In CVPR Workshop on Generative-Model Based Vision.
13. L. Fei-Fei (Stanford), Rob Fergus (NYU), Antonio Torralba (MIT) (2005), *Recognizing and Learning Object Categories*, Awarded the Best Short Course Prize at ICCV.
14. Zheyun Feng, Songhe Feng, Rong Jin, and Anil K. Jain (2014), *Image Tag Completion by Noisy Matrix Recovery*, In Proc. of ECCV, pp. 424-438.
15. Songhe Feng, Congyan Lang, and Bing Li (2012), *Towards Relevance and Saliency Ranking of Image Tags*, In Proc. of ACM MM, pp. 917-920.
16. M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, I. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker (1995), *Query by image and video content: The QBIC project*, IEEE computer, 28(9), pp. 23-32.

17. Yue Gao, Meng Wang, Zheng-Jun Zha, Jialie Shen, Xuelong Li, and Xindong Wu (2013), *Visual-Textual Joint Relevance Learning for Tag-Based Social Image Search*, IEEE Transactions on Image Processing 22, 1 (2013), pp. 363-376.
18. Alexandru Lucian Ginsca, Adrian Popescu, Bogdan Ionescu, Anil Armagan, and Ioannis Kanellos (2014), *Toward an Estimation of User Tagging Credibility for Social Image Retrieval*, In Proc. of ACM MM, pp.1021-1024.
19. M. Guillaumin, J. Verbeek, and C. Schmid (2010), *Multimodal semi-supervised learning for image classification*, In CVPR.
20. J. Hays and A. A. Efros (2008), *IM2GPS: estimating geographic information from a single image*, In CVPR.
21. J. Huang, S. Kumar, M. Mitra, W.J. Zhu, and R. Zabih. June (1997), *Image indexing using color correlogram*, In IEEE Conf. on Computer Vision and Pattern Recognition, pp. 762-768.
22. S. J. Hwang and K. Grauman (2012), *Learning the relative importance of objects from tagged images for retrieval and cross-modal search*, IJCV, 100(2):134-153.
23. Thorsten Joachims (1999), *Transductive Inference for Text Classification using Support Vector Machines*, In Proc. of ICML, pp. 200-209.
24. Justin Johnson, Lamberto Ballan, and Li Fei-Fei (2015), *Love Thy Neighbors: Image Annotation by Exploiting Image Metadata*, In Proc. of ICCV.
25. Mahdi M. Kalayeh, Haroon Idrees, and Mubarak Shah (2014), *NMF-KNN: Image Annotation Using Weighted Multi-view Non-negative Matrix Factorization*, In Proc. of CVPR, pp. 184-191.
26. Lyndon S. Kennedy, Malcolm Slaney, and Kilian Weinberger (2009), *Reliable Tags Using Image Similarity: Mining Specificity and Expertise from Large-scale Multimedia Databases*, In Proc. of ACM MM Workshop on Web-scale Multimedia Corpus, pp. 17-24.

27. Gunhee Kim and Eric P Xing (2013), *Time-sensitive Web Image Ranking and Retrieval via Dynamic Multitask Regression*, In Proc. of ACM WSDM, pp. 163-172.
28. G. Kim, E. P. Xing, and A. Torralba (2010), *Modeling and analysis of dynamic behaviors of web image collections*, In ECCV.
29. Yin-Hsi Kuo, Wen-Huang Cheng, Hsuan-Tien Lin, and Winston H. Hsu (2012), *Unsupervised Semantic Feature Discovery for Image Object Retrieval and Tag Refinement*, IEEE Transactions on Multimedia 14, 4 (2012), pp. 1079-1090.
30. Sihyoung Lee, Wesley De Neve, and Yong Man Ro (2013), *Visually Weighted Neighbor Voting for Image Tag Relevance Learning*, Multimedia Tools and Applications 72, 2 (2013), pp. 1363-1386.
31. Xirong Li, Cees G.M. Snoek, and Marcel Worring (2010), *Unsupervised Multi-feature Tag Relevance Learning for Social Image Retrieval*, In Proc. of ACM CIVR, pp. 10-17.
32. Y. Li, D. Crandall, and D. Huttenlocher (2009), *Landmark classification in large-scale image collections*, In ICCV.
33. Xirong Li and Cees G.M. Snoek (2013), *Classifying Tag Relevance with Relevant Positive and Negative Examples*, In Proc. of ACM MM, pp. 485-488.
34. Xirong Li (2016), *Tag Relevance Fusion for Social Image Retrieval*, Multimedia Systems In press (2016). DOI:<http://dx.doi.org/10.1007/s00530-014-0430-9>.
35. Xirong Li, Cees G.M. Snoek, and Marcel Worring (2009b), *Learning Social Tag Relevance by Neighbor Voting*, IEEE Transactions on Multimedia 11, 7 (2009), pp. 1310-1322.
36. Xirong Li, Efstratios Gavves, Cees G.M. Snoek, Marcel Worring, and Arnold W.M. Smeulders (2011b), *Personalizing Automated Image Annotation using Cross-Entropy*, In Proc. of ACM MM, pp. 233-242.

37. Xirong Li and Cees G.M. Snoek (2013), *Classifying Tag Relevance with Relevant Positive and Negative Examples*, In Proc. of ACM MM, pp. 485-488.
38. Xirong Li, Cees G.M. Snoek, and Marcel Worring (2009b), *Learning Social Tag Relevance by Neighbor Voting*, IEEE Transactions on Multimedia 11, pp. 1310-1322.
39. Jing Liu, Yifan Zhang, Zechao Li, and Hanqing Lu (2013), *Correlation Consistency Constrained Probabilistic Matrix Factorization for Social Tag Refinement*, Neurocomputing 119, pp. 3-9.
40. Jing Liu, Zechao Li, Jinhui Tang, Yu Jiang, and Hanqing Lu (2014), *Personalized Geo-specific Tag Recommendation for Photos on Social Websites*, IEEE Transactions on Multimedia 16, pp. 588-600.
41. Dong Liu, Xian-Sheng Hua, Linjun Yang, Meng Wang, and Hong-Jiang Zhang (2009), *Tag Ranking*, In Proc. of WWW, pp. 351-360.
42. Dong Liu, Shuicheng Yan, Xian-Sheng Hua, and Hong-Jiang Zhang (2011b), *Image Retagging Using Collaborative Tag Propagation*, IEEE Transactions on Multimedia 13, pp. 702-712.
43. Jing Liu, Yifan Zhang, Zechao Li, and Hanqing Lu (2013), *Correlation Consistency Constrained Probabilistic Matrix Factorization for Social Tag Refinement*, Neurocomputing 119, pp. 3-9.
44. D. Lowe (2004), *Distinctive image features from scale-invariant keypoint*, Int'l J. Computer Vision, 2(60), pp. 91-110.
45. Hao Ma, Jianke Zhu, Michael Rung-Tsong Lyu, and Irwin King (2010), *Bridging the Semantic Gap Between Image Contents and Tags*, IEEE Transactions on Multimedia 12, pp. 462-473.
46. Syaifulnizam Abd Manal and Md Jan Nordin (2009), "Review on statistical approaches for automatic image annotation", international conference on electrical engineering and informatics 5-7, IEEE 978-1-4244-4913-2/09.

47. B. S. Manjunath and W. Y. Ma (1996), *Texture features for browsing and retrieval of image data*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 18(8), pp. 837-842.
48. Julian McAuley and Jure Leskovec (2012), *Image Labeling on a Network: Using Social-network Metadata for Image Classification*, In Proc. of ECCV, pp. 828-841.
49. Philip McParlane, Stewart Whiting, and Joemon Jose (2013b), *Improving Automatic Image Tagging Using Temporal Tag Co-occurrence*, In Proc. of MMM, pp. 251-262.
50. Philip J. McParlane, Yashar Moshfeghi, and Joemon M. Jose (2013a), *On Contextual Photo Tag Recommendation*, In Proc. of ACM SIGIR, pp. 965-968.
51. Ryszard S. Michalski (1993), *Readings in Knowledge Acquisition and Learning*, Morgan Kaufmann Publishers Inc., Chapter A Theory and Methodology of Inductive Learning, pp. 323-348.
52. M. Naphade, J. R. Smith, J. Tesic, S. Chang, W. Hsu, L. Kennedy, A. Hauptmann, and J. Curtis (2006), *A large-scale concept ontology for multimedia*. IEEE MultiMedia, 13, pp. 86-91.
53. Z. Niu, G. Hua, X. Gao, and Q. Tian (2014), *Semi-supervised relational topic model for weakly annotated image recognition in social media*, In CVPR.
54. Zhenxing Niu, Gang Hua, Xinbo Gao, and Qi Tian (2014), *Semi-supervised Relational Topic Model for Weakly Annotated Image Recognition in Social Media*, In Proc. of CVPR, pp. 4233-4240.
55. D. K. Park, Y. S. Jeon, and C. S. Won (2000), *Efficient use of local edge histogram descriptor*, In ACM Multimedia.
56. Jose Costa Pereira, Emanuele Coviello, Gabriel Doyle, Nikhil Rasiwasia, Gert R.G. Lanckriet, Roger Levy, and Nuno Vasconcelos (2014), *On the Role of Correlation and Abstraction in Cross-Modal Multimedia Retrieval*, IEEE Transactions on Pattern Analysis and Machine Intelligence 36, pp. 521-535

57. Guo-Jun Qi, Charu Aggarwal, Qi Tian, Heng Ji, and Thomas Huang (2012), *Exploring Context and Content Links in Social Media: A Latent Space Method*, IEEE Transactions on Pattern Analysis and Machine Intelligence 34, pp. 850-862.
58. Xueming Qian, Xian-Sheng Hua, Yuan Yan Tang, and Tao Mei (2014), *Social Image Tagging With Diverse Semantics*, IEEE Transactions on Cybernetics 44, 12 (2014), pp. 2493-2508.
59. Zhiming Qian, Ping Zhong, and Runsheng Wang (2015), *Tag Refinement for User-contributed Images via Graph Learning and Nonnegative Tensor Factorization*, IEEE Signal Processing Letters 22, 9 (2015), pp. 1302-1305.
60. Ba Quan Truong, Aixin Sun, and Sourav S. Bhowmick (2012), *Content is Still King: The Effect of Neighbor Voting Schemes on Tag Relevance for Social Image Retrieval*, In Proc. of ACM ICMR, 9:1-9:8.
61. Fabian Richter, Stefan Romberg, Eva Horster, and Rainer Lienhart (2012), *Leveraging Community Metadata for Multimodal Image Ranking*, Multimedia Tools and Applications 56, 1 (2012), pp. 35-62.
62. Jitao Sang, Changsheng Xu, and Jing Liu (2012a), *User-Aware Image Tag Refinement via Ternary Semantic Analysis*, IEEE Transactions on Multimedia 14, pp. 883-895.
63. Jitao Sang, Changsheng Xu, and Dongyuan Lu (2012b), *Learn to Personalized Image Search From the Photo Sharing Websites*, IEEE Transactions on Multimedia 14, pp. 963-974.
64. Jitao Sang, Changsheng Xu, and Jing Liu (2012a), *User-Aware Image Tag Refinement via Ternary Semantic Analysis*, IEEE Transactions on Multimedia 14, 3 (2012), pp. 883-895.
65. Neela Sawant, Ritendra Datta, Jia Li, and James Z. Wang (2010), *Quest for Relevant Tags Using Local Interaction Networks and Visual Content*, In Proc. of ACM MIR, pp. 231-240.
66. L. G. Shapiro and G. C. Stockman (2003), *Computer Vision*, Prentice Hall.

- 67.V. Shirahatti and K. Barnard (1996), “*Method for Comparing Content Based Image Retrieval Methods*” in Proceedings of the SPIE 2003, pp.1-8. (IJSIS '96), pp. 261.
- 68.Borkur Sigurbjornsson and Roelof Van Zwol (2008), *Flickr Tag Recommendation based on Collective Knowledge*, In Proc. of WWW, pp. 327-336.
- 69.C. G. M. Snoek, M. Worring, J. C. van Gemert, J.-M. Geusebroek, and A. W. M. Smeulders (2006), *The challenge problem for automated detection of 101 semantic concepts in multimedia*, In ACM Multimedia.
- 70.Nitish Srivastava and Ruslan R. Salakhutdinov (2014), Multimodal Learning with Deep Boltzmann Machines. *Journal of Machine Learning Research* 15, 1 (2014), pp. 2949-2980.
- 71.M. Stricker and M. Orengo (1995), Similarity of color images, In SPIE Storage and Retrieval for Image and Video Databases III.
- 72.Aixin Sun, Sourav S. Bhowmick, Nam Nguyen, Khanh Tran, and Ge Bai (2011), Tag-Based Social Image Retrieval: An Empirical Evaluation, *Journal of the American Society for Information Science and Technology* 62, pp. 2364-2381.
- 73.Ledyard R. Tucker (1966), Some Mathematical Notes on Three-mode Factor Analysis. *Psychometrika* 31, 3 (1966), pp. 279-311.
- 74.Jakob Verbeek, Matthieu Guillaumin, Thomas Mensink, and Cordelia Schmid (2010), Image Annotation with TagProp on the MIRFLICKR Set, In Proc. of ACM MIR, pp. 537-546.
- 75.Gang Wang, Derek Hoiem, and David Forsyth (2009), Building Text Features for Object Image Classification, In Proc. of CVPR, pp. 1367-1374, 462-473.
- 76.L. Wenyin, S. Dumais, Y. Sun, H. J. Zhang, M. Czerwinski and B.Field (2002), “Semi Automatic Image Annotation” in 8th IFIP T.C 13Conference on Human-Computer Interaction, pp. 326-333.

- 77.C. H. Wiener, N. Simou and Tzouvaras (2006), Image Annotation on the Semantic Web. [Online].Available: <http://www.w3.org/TR/2006/WD-swbp-image-annotation-20060322>.
- 78.Lei Wu, Linjun Yang, Nenghai Yu, and Xian-Sheng Hua (2009), Learning to tag, In Proc. of WWW, pp. 361-370.
- 79.Lei Wu, Rong Jin, and Anubhav K. Jain (2013), Tag Completion for Image Retrieval, IEEE Transactions on Pattern Analysis and Machine Intelligence 35, 3 (2013), pp. 716-727.
- 80.Hao Xu, Jingdong Wang, Xian-Sheng Hua, and Shipeng Li (2009), Tag Refinement by Regularized LDA, In Proc. of ACM MM, pp. 573-576.
- 81.Xing Xu, Akira Shimada, and Rin-ichiro Taniguchi (2014), Tag Completion with Defective Tag Assignments via Image-Tag Re-weighting, In Proc. of ICME, pp. 1-6.
- 82.Yang Yang, Yue Gao, Hanwang Zhang, Jie Shao, and Tat-Seng Chua (2014), Image Tagging with Social Assistance, In Proc. of ACM ICMR, pp. 81-88.
- 83.Zamir, S. Ardeshir, and M. Shah (2014), Gps-tag refinement using random walks with an adaptive damping factor, In CVPR.
- 84.Guangyu Zhu, Shuicheng Yan, and Yi Ma (2010), *Image Tag Refinement Towards Low-Rank, Content-Tag Prior and Error Sparsity*, In Proc. of ACM MM, pp. 461-470.
- 85.Xiaofei Zhu, Wolfgang Nejdl, and Mihai Georgescu (2014), *An Adaptive Teleportation Random Walk Model for Learning Social Tag Relevance*, In Proc. of ACM SIGIR, pp. 223-232.
- 86.Guangyu Zhu, Shuicheng Yan, and Yi Ma (2010), *Image Tag Refinement Towards Low-Rank, Content-Tag Prior and Error Sparsity*, In Proc. of ACM MM, pp. 461-470.
- 87.Shiai Zhu, Chong-Wah Ngo, and Yu-Gang Jiang (2012), *Sampling and Ontologically Pooling Web Images for Visual Concept Learning*, IEEE Transactions on Multimedia 14, 4 (2012), pp. 1068-1078.

88. Jinfeng Zhuang and Steven C.H. Hoi (2011), *A Two-View Learning Approach for Image Tag Ranking*, In Proc. of ACM WSDM, pp. 625-634.

PHỤ LỤC

Danh sách 1,000 chú giải được lấy từ thông tin phụ trợ của tập dữ liệu NUS-WIDE:

1	abandoned	335	flora	669	picture
2	abstract	336	florida	670	pictures
3	action	337	flower	671	pier
4	actor	338	flowers	672	pilot
5	adorable	339	fly	673	pink
6	adult	340	flying	674	pipes
7	adults	341	foals	675	plane
8	adventure	342	fog	676	plant
9	aerial	343	food	677	plants
10	africa	344	football	678	plastic
11	african	345	forest	679	play
12	afternoon	346	formula	680	playing
13	agent	347	fountain	681	plaza
14	agua	348	fox	682	poland
15	air	349	frame	683	polar
16	aircraft	350	france	684	polaroid
17	airforce	351	free	685	pole
18	airplane	352	freedom	686	police
19	airplanes	353	french	687	politicians
20	airport	354	friend	688	politics
21	alaska	355	friends	689	pond
22	alberta	356	frost	690	pool
23	alley	357	frozen	691	pope
24	amazing	358	fruit	692	port
25	america	359	fun	693	portfolio
26	american	360	funeral	694	portland
27	amsterdam	361	funny	695	portrait
28	analog	362	furniture	696	portraits
29	ancient	363	game	697	portugal
30	angel	364	garden	698	post
31	angle	365	gardens	699	pottery

32	animal	366	gate	700	poverty
33	animals	367	geometry	701	power
34	antenna	368	georgia	702	prayer
35	antique	369	germany	703	president
36	antlers	370	giant	704	pretty
37	apartments	371	giraffe	705	prison
38	apple	372	girl	706	protest
39	april	373	girls	707	protesters
40	aquarium	374	glacier	708	public
41	arch	375	glass	709	puppy
42	architecture	376	glasses	710	pups
43	arctic	377	glow	711	purple
44	argentina	378	goat	712	pyramid
45	arizona	379	god	713	quality
46	army	380	gold	714	quebec
47	art	381	golden	715	rabbit
48	artist	382	golf	716	race
49	asia	383	gorgeous	717	racing
50	asian	384	gothic	718	radar
51	athlete	385	graffiti	719	raft
52	atlanta	386	grain	720	rail
53	atmosphere	387	graphic	721	railing
54	august	388	grass	722	railroad
55	austin	389	graveyard	723	railway
56	australia	390	gray	724	rain
57	austria	391	great	725	rainbow
58	auto	392	greece	726	rainy
59	automobile	393	green	727	rally
60	autumn	394	grey	728	range
61	aviation	395	groom	729	raw
62	awesome	396	ground	730	rays
63	baby	397	group	731	rebel
64	backpack	398	guard	732	red
65	balcony	399	guest	733	reef
66	ball	400	gun	734	reefs

67	balloons	401	guns	735	reflection
68	band	402	hair	736	reflections
69	bank	403	halloween	737	refugees
70	bar	404	hand	738	relief
71	barcelona	405	handmade	739	religion
72	barge	406	hands	740	reptile
73	barn	407	hangar	741	restaurant
74	baseball	408	happiness	742	retro
75	basket	409	happy	743	rice
76	basketball	410	harbor	744	rifles
77	bathroom	411	harbour	745	riot
78	bay	412	harvest	746	river
79	bazaar	413	hat	747	road
80	beach	414	hats	748	roadblock
81	bear	415	hawaii	749	rock
82	beards	416	hawk	750	rocks
83	bears	417	head	751	rodent
84	beautiful	418	health	752	roma
85	beauty	419	heart	753	rome
86	belgium	420	helicopter	754	roof
87	bench	421	helicopters	755	room
88	berlin	422	heritage	756	rose
89	best	423	high	757	ruin
90	bicycle	424	highway	758	ruins
91	big	425	hiking	759	run
92	bike	426	hill	760	running
93	bird	427	hills	761	runway
94	birds	428	historic	762	rural
95	black	429	historical	763	rust
96	bleu	430	history	764	rusty
97	blooms	431	holiday	765	sad
98	blossoms	432	holidays	766	safari
99	blue	433	holland	767	sail
100	blur	434	home	768	sailboat
101	boat	435	hope	769	sailing

102	boats	436	horizon	770	sails
103	booby	437	horns	771	sand
104	book	438	horse	772	save
105	boston	439	horses	773	scenery
106	boy	440	hospital	774	scenic
107	boys	441	host	775	school
108	branch	442	hot	776	science
109	brasil	443	hotel	777	scientists
110	bravo	444	house	778	scotland
111	brazil	445	houseboat	779	scuba
112	breathtaking	446	houses	780	sculpture
113	brick	447	human	781	sea
114	bride	448	hut	782	seascape
115	bridge	449	ice	783	seaside
116	bridges	450	iceland	784	seattle
117	bright	451	illinois	785	self
118	brighton	452	illustration	786	sepia
119	britain	453	image	787	sewing
120	british	454	images	788	shade
121	brooklyn	455	india	789	shadow
122	brown	456	indian	790	shadows
123	buddha	457	individual	791	sheep
124	buddhism	458	indonesia	792	shell
125	buddhist	459	industrial	793	ship
126	bug	460	industry	794	ships
127	building	461	insect	795	shirt
128	buildings	462	interesting	796	shoes
129	bunny	463	interestingness	797	shooting
130	burma	464	interior	798	shop
131	bus	465	international	799	shopping
132	bush	466	iran	800	shops
133	business	467	iraq	801	shore
134	butterfly	468	ireland	802	shot
135	cabin	469	islam	803	shrine
136	cables	470	island	804	sidewalk

137	cactus	471	islands	805	sigma
138	cafe	472	israel	806	sign
139	cake	473	istanbul	807	signs
140	california	474	italia	808	silhouette
141	camera	475	italians	809	silhouettes
142	canada	476	italy	810	silver
143	canal	477	jail	811	singapore
144	candid	478	japan	812	singing
145	canoe	479	japanese	813	sitting
146	canyon	480	jeep	814	skiing
147	capital	481	jet	815	sky
148	car	482	joy	816	skyline
149	caribou	483	judge	817	skyscraper
150	carpet	484	july	818	small
151	cars	485	jump	819	smile
152	castle	486	june	820	smoke
153	cat	487	kauai	821	snake
154	cathedral	488	kid	822	snow
155	cats	489	kids	823	soccer
156	cave	490	kiss	824	social
157	ceiling	491	kit	825	soft
158	cellphones	492	kitchen	826	soldier
159	cemetery	493	kitten	827	soldiers
160	center	494	kitty	828	south
161	ceremony	495	kuwait	829	space
162	chair	496	ladder	830	spain
163	chairs	497	lady	831	speed
164	chapel	498	lake	832	spider
165	charts	499	lakes	833	sport
166	cheering	500	lamp	834	sports
167	cheerleader	501	land	835	spring
168	cheese	502	landing	836	square
169	chicago	503	landscape	837	squirrel
170	child	504	landscapes	838	stadium
171	chile	505	lawn	839	stairs

172	china	506	leaf	840	standing
173	chinese	507	leaves	841	star
174	christmas	508	legs	842	stars
175	church	509	lens	843	state
176	circle	510	leopard	844	station
177	city	511	library	845	statue
178	cityscape	512	life	846	steam
179	classic	513	light	847	steel
180	classroom	514	lighthouse	848	steps
181	cliff	515	lighting	849	stone
182	close	516	lights	850	stones
183	closeup	517	lily	851	store
184	clothes	518	line	852	storm
185	clothing	519	lines	853	storms
186	cloud	520	lion	854	stream
187	clouds	521	little	855	street
188	cloudy	522	live	856	streets
189	club	523	lizard	857	stripes
190	coast	524	locomotive	858	structure
191	coffee	525	london	859	studio
192	cold	526	lonely	860	stunning
193	collage	527	long	861	style
194	college	528	look	862	submarine
195	color	529	love	863	suburban
196	colorado	530	lovely	864	subway
197	colorful	531	mac	865	suits
198	colors	532	macro	866	summer
199	colour	533	madrid	867	sun
200	colourful	534	magic	868	sunglasses
201	colours	535	magical	869	sunlight
202	commercials	536	malaysia	870	sunny
203	composition	537	maldives	871	sunrise
204	concert	538	male	872	sunset
205	concrete	539	males	873	sunshine
206	construction	540	mall	874	surf

207	contrast	541	mammal	875	surreal
208	cool	542	man	876	sweden
209	copyright	543	manhattan	877	sweet
210	coral	544	maps	878	swim
211	costume	545	mar	879	swimming
212	cottage	546	march	880	switzerland
213	country	547	mare	881	sydney
214	countryside	548	marine	882	taiwan
215	county	549	market	883	talking
216	couple	550	marsh	884	tanker
217	court	551	maryland	885	tanks
218	cow	552	massachusetts	886	telephones
219	cows	553	maui	887	temple
220	crab	554	meadow	888	tennis
221	craft	555	meeting	889	tent
222	crafts	556	melbourne	890	texas
223	creative	557	memorial	891	texture
224	creek	558	men	892	textures
225	cross	559	metal	893	thailand
226	crowd	560	metro	894	theater
227	cruise	561	mexico	895	ties
228	crystals	562	michigan	896	tiger
229	cubism	563	microphones	897	time
230	cubs	564	military	898	tokyo
231	culture	565	minnesota	899	tollbooth
232	cute	566	mirror	900	tomb
233	dance	567	missouri	901	tornado
234	dancing	568	mist	902	toronto
235	dark	569	model	903	tortoise
236	darkness	570	modern	904	tour
237	dawn	571	monastery	905	tourism
238	day	572	monks	906	tourist
239	dead	573	monochrome	907	tower
240	death	574	montreal	908	town
241	decay	575	monument	909	toy

242	december	576	moon	910	toys
243	decoration	577	moose	911	track
244	deer	578	morning	912	tracks
245	delete	579	mosque	913	traditional
246	democrat	580	mosques	914	traffic
247	demonstration	581	moss	915	trail
248	desert	582	mother	916	train
249	design	583	motion	917	transportation
250	designs	584	motorcycle	918	travel
251	deutschland	585	mountain	919	tree
252	digital	586	mountains	920	trees
253	display	587	movement	921	trip
254	dive	588	movie	922	tropical
255	diving	589	mug	923	truck
256	dock	590	mural	924	trunk
257	documentary	591	museum	925	tulip
258	dog	592	music	926	tundra
259	dogs	593	muslim	927	tunnel
260	doll	594	national	928	turkey
261	door	595	natural	929	tusks
262	doorway	596	nature	930	twilight
263	downtown	597	navy	931	umbrella
264	dramatic	598	nederland	932	underground
265	drawing	599	needles	933	underwater
266	dream	600	neon	934	uniform
267	dress	601	netherlands	935	united
268	driver	602	nets	936	university
269	drop	603	nevada	937	urban
270	drops	604	new	938	utah
271	drum	605	news	939	vacation
272	dunes	606	nice	940	valley
273	dusk	607	night	941	valleys
274	dust	608	nighttime	942	vancouver
275	dynamic	609	north	943	vegetables
276	eagle	610	norway	944	vehicle

277	earth	611	nose	945	victoria
278	earthquake	612	nsw	946	vietnam
279	east	613	oahu	947	view
280	egypt	614	ocean	948	village
281	election	615	october	949	vintage
282	elephant	616	office	950	virginia
283	elephants	617	officers	951	vivid
284	elk	618	ohio	952	volcano
285	engine	619	oil	953	waiting
286	england	620	oilfield	954	wales
287	entertainment	621	old	955	walk
288	entrance	622	olympus	956	walking
289	environment	623	ontario	957	wall
290	eos	624	orange	958	wallpaper
291	europa	625	orchid	959	war
292	europe	626	oregon	960	warehouse
293	evening	627	outdoor	961	washington
294	excellence	628	outdoors	962	water
295	explore	629	outside	963	waterfall
296	exposure	630	pacific	964	waterfalls
297	eye	631	paint	965	waterways
298	eyes	632	painting	966	wave
299	fab	633	paintings	967	waves
300	fabulous	634	pair	968	weapons
301	facade	635	pakistan	969	weather
302	face	636	palace	970	wedding
303	faces	637	palm	971	west
304	factory	638	panorama	972	wet
305	fall	639	paper	973	white
306	family	640	parade	974	wide
307	fan	641	paradise	975	wild
308	fantastic	642	paris	976	wildlife
309	fantasy	643	park	977	wind
310	farm	644	party	978	window
311	fashion	645	path	979	windows

312	fauna	646	pattern	980	windy
313	favorite	647	peace	981	wing
314	fawn	648	peaceful	982	wings
315	feline	649	pebbles	983	winter
316	female	650	penguin	984	wisconsin
317	fence	651	pennsylvania	985	wolf
318	festival	652	people	986	woman
319	field	653	perfect	987	wonder
320	fields	654	person	988	wonderful
321	fighter	655	perspective	989	wood
322	figures	656	pet	990	woods
323	film	657	petals	991	work
324	finland	658	pets	992	world
325	firefighter	659	philadelphia	993	wow
326	fish	660	philippines	994	writing
327	fisheye	661	phone	995	yellow
328	fishing	662	photo	996	york
329	flag	663	photograph	997	young
330	flags	664	photographer	998	youth
331	flash	665	photographers	999	zebra
332	flight	666	photography	1000	zoo
333	flood	667	photojournalism		
334	floor	668	photos		

