

**ĐẠI HỌC QUỐC GIA TP.HỒ CHÍ MINH**  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**  
**NGÀNH KHOA HỌC MÁY TÍNH**



**BÁO CÁO ĐỒ ÁN**  
**NHẬN DẠNG THỊ GIÁC VÀ ỨNG DỤNG**

**Nội dung: Bài toán tìm kiếm hình ảnh**

**Giảng viên HD:** TS. Lê Đình Duy  
TS. Nguyễn Tân Trần Minh Khang

**Học viên thực hiện:** Phạm Xuân Y

**MSHV:** CH1601044

**Lớp:** KHMT Khóa 11 đợt 2

*Tp HCM, 12/2017*

## **NHẬN XÉT CỦA GIẢNG VIÊN**

---

---

---

---

---

---

---

---

---

---

---

---

## MỤC LỤC

NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN.....	2
LỜI MỞ ĐẦU .....	1
Chương I: Các phương pháp truy vấn thông tin thị giác .....	2
1.    Tìm kiếm theo meta-data.....	2
2.    Tìm kiếm dựa trên nội dung (Content-Based Image Retrieval) .....	2
3.    Các bước xây dựng hệ thống tìm kiếm dựa trên nội dung.....	3
Chương II: Phương pháp, kỹ thuật .....	4
2.1.    Bộ dữ liệu Oxford Building .....	4
2.2.    Xác định đặc trưng .....	4
2.3.    Bag of Visual Words.....	4
2.4.    Trích xuất đặc trưng từ kho dữ liệu hình ảnh Oxford Building.....	6
2.5.    Xây dựng chương trình tìm kiếm .....	7
Chương III: Kết quả .....	9
Chương IV: Kết luận .....	13
Tài liệu tham khảo.....	15

## **LỜI MỞ ĐẦU**

Với sự bùng nổ thông tin toàn cầu hiện nay thì nhu cầu xây dựng những hệ thống tìm kiếm thông tin luôn cấp thiết để đáp ứng nhu cầu lưu trữ, quản lý và tìm kiếm. Trong thời gian vài năm trở lại đây dữ liệu hình ảnh, video tăng một cách bùng nổ, đặc tính của loại dữ liệu này là thường có kích thước lớn, lượng thông tin nhiều, đa dạng cả về số lượng lẫn nội dung nên cần những phương pháp, cách thức xây dựng mới đối với hệ thống tìm kiếm nhằm tăng tính chính xác và hiệu quả trong xử lý và thời gian đáp ứng truy vấn.

Hiện đã có rất nhiều công trình nghiên cứu đưa ra nhiều giải pháp cho việc tìm kiếm dạng thông tin thị giác (gồm hình ảnh và video) như phương pháp meta-data gắn thông tin định danh, phân loại , mô tả nội dung của hình ảnh hoặc video nhằm đưa bài toán về dạng tìm kiếm văn bản, Hay phương pháp tìm kiếm dựa trên nội dung của hình ảnh hoặc video,... Với phạm vi đồ án môn học em sẽ áp dụng các phương pháp tìm kiếm ảnh đã có để xây dựng một hệ thống truy vấn hình ảnh thử nghiệm từ đó kiểm tra, phân tích đánh giá tính hiệu quả cũng như ứng dụng của giải pháp đã chọn.

## **Chương I: Các phương pháp truy vấn thông tin thị giác**

### **1. Tìm kiếm theo meta-data**

Phương pháp này thực hiện bằng cách ghi chú thích bằng từ ngữ cho các nội dung có trong bức ảnh. Kỹ thuật tìm kiếm dựa trên kỹ thuật tìm kiếm văn bản đã có, nhập truy vấn bằng từ ngữ mô tả, tìm kiếm các hình ảnh kết quả có gắn nhãn, mô tả giống với câu query. Ưu điểm của phương pháp là đơn giản dễ thực hiện, tìm kiếm nhanh, nhưng nhược điểm là tốn nhiều thời gian và công sức cho quá trình tiền xử lý, nhập mô tả cho hình ảnh, còn phụ thuộc vào con người.

### **2. Tìm kiếm dựa trên nội dung (Content-Based Image Retrieval)**

Khác với phương pháp tìm kiếm theo meta-data, phương pháp tìm kiếm dựa trên nội dung sử dụng hình ảnh làm câu truy vấn, và trả về các hình ảnh có nội dung giống hoặc gần giống với hình ảnh truy vấn nhất, bằng cách sử dụng một số thuật toán để trích xuất đặc trưng của hình ảnh, lưu trữ lại, khi tìm kiếm, ảnh query của người dùng cũng sẽ được trích xuất đặc trưng, sau đó thuật toán sẽ tiến hành tính toán mức độ tương đồng giữa ảnh query và đặc trưng của ảnh trong kho đã được trích xuất, danh sách các ảnh có mức độ tương đồng từ cao đến thấp sẽ được trả về cho người dùng. Ưu điểm của phương pháp này là tìm kiếm tự động dựa trên nội dung do đó đáp ứng được số lượng hình ảnh lớn (đối với kho dữ liệu lớn, không tồn công mô tả, gắn nhãn cho nội dung bức ảnh), độ chính xác ngày một cao cùng với sự ra đời của nhiều phương pháp rút trích đặc trưng. Nhược điểm về thời gian xây dựng và xử lý tìm kiếm.

Ngoài ra còn có các hệ thống tìm kiếm kết hợp cả 2 phương pháp trên.

### **3. Các bước xây dựng hệ thống tìm kiếm dựa trên nội dung**

Có 4 bước chính trong quá trình xây dựng một hệ thống tìm kiếm hình ảnh dựa trên nội dung

**Bước 1:** Xác định đặc trưng của bức ảnh.

Quyết định những đặc điểm nào của ảnh chúng ta muốn mô tả, trích xuất, như màu sắc, cạnh, biên ảnh,...

**Bước 2:** Trích xuất đặc trưng cho toàn bộ kho dữ liệu ảnh

Với đặc trưng đã xác định ở bước 1, tiến hành trích xuất đặc trưng cho toàn bộ dữ liệu hình ảnh trong kho, lưu trữ các đặc trưng của mỗi bức ảnh trong một file index, để sử dụng cho bước tiếp theo.

**Bước 3:** Xác định độ đo tương đồng

Để xác định 2 bức ảnh là giống nhau cần có một độ đo, có thể dùng khoảng cách Euclidean, Cosine,... Việc lựa chọn độ đo nào tuỳ thuộc vào kho dữ liệu, loại đặc trưng đã xác định.

**Bước 4:** Tìm kiếm hình ảnh

Tại bước này thực hiện tìm kiếm hình ảnh trong kho với ảnh truy vấn. Tiến hành rút trích đặc trưng của ảnh truy vấn, xác định mức độ tương đồng giữa ảnh truy vấn với các ảnh có trong kho dữ liệu và trả về các ảnh kết quả theo thứ tự giảm dần của độ tương đồng.

## **Chương II: Phương pháp, kỹ thuật**

### **2.1. Bộ dữ liệu Oxford Building**

Bộ dữ liệu hình ảnh này chứa các hình ảnh về trường đại học Oxford Building, hình ảnh đa dạng về các tòa nhà, khuôn viên, cũng như cuộc sống của sinh viên trong trường... Đồ án sử dụng bộ dữ liệu này để thực hiện thử nghiệm đánh giá hệ thống tìm kiếm. Do giới hạn về phần cứng máy tính nên đồ án chỉ sử dụng 327 trong tổng số 5063 tấm ảnh trong bộ dữ liệu.

### **2.2. Xác định đặc trưng**

Giới thiệu về SURF (Speeded Up Robust Features): Một phiên bản tăng tốc độ của SIFT (Scale Invariant Feature Transform) – Phương pháp phát hiện và mô tả các điểm đặc trưng(keypoint) của một tấm ảnh nhưng được cải thiện về tốc độ thực thi so với SIFT mà vẫn đảm bảo độ chính xác trong việc phát hiện các điểm đặc trưng. SURF được giới thiệu vào năm 2006 bởi Bay, H., Tuytelaars, T. and Van Gool, L

Với SIFT việc tìm Scale-space dựa trên việc tính gần đúng LoG (Laplace of Gaussian) dùng DoG (Difference of Gaussian), trong khi đó SURF sử dụng Box Filter, tốc độ xử lý sẽ được cải thiện đáng kể với việc dùng ảnh tích phân (integral image)

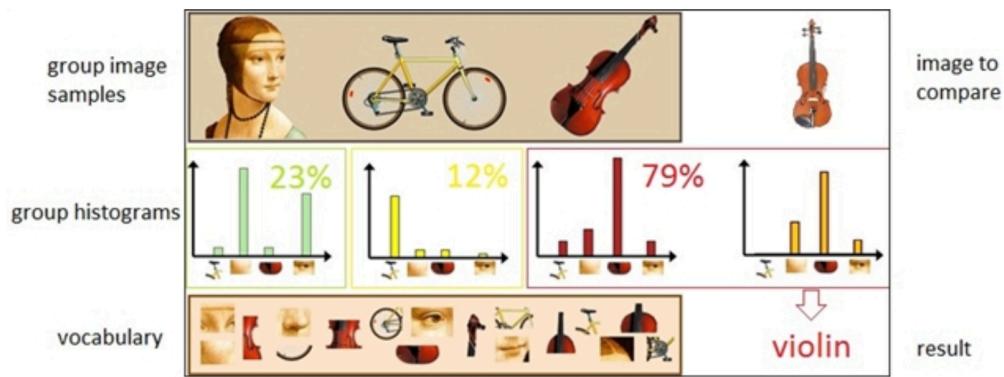
Ở bước Orientation Assignment, SURF sử dụng wavelet response theo 2 chiều dọc và ngang, sau đó tính hướng chính bằng cách tính tổng các response đó, có một điều đáng chú ý là wavelet response cũng dễ dàng tính được với ảnh tích phân (integral image)

### **2.3. Bag of Visual Words**

Bag of visual words (BOVW) là một phương pháp để biểu diễn ảnh nhằm giải quyết vấn đề không gian lưu trữ và tốc độ truy vấn bằng phương pháp phân cụm. Trình tự các bước thực hiện của phương pháp này, như sau:

- Phân cụm bằng giải thuật k-means.
- Gán từng descriptor của ảnh vào cluster gần nhất.
- Đếm số descriptor của ảnh được gán vào từng cluster.

- Cuối cùng biểu diễn ảnh bằng 1 histogram các visual words.



### Tổng quan phương pháp BOVW

Trong thực tế chúng ta không so khớp ảnh trực tiếp với các đặc trưng của 2 ảnh với nhau vì không gian và thời gian tìm kiếm lớn, mà sử dụng một phương pháp phân lớp rồi sau đó mới thực hiện so khớp ảnh. BOVW sử dụng giải thuật phân lớp là k-Means với số lần lặp xác định để tạo ra k cluster bằng cách tính trung bình khoảng cách giữa các vector đặc trưng gần nhất. Tập hợp các cluster gọi là visual word vocabulary, kích thước của vocabulary phụ thuộc vào số lượng các cluster.

### Tính khoảng cách 2 vector visual words

Để so khớp hai vector các visual words, sử dụng phương pháp tf\*idf (term frequency, inverse document frequency). Tf\*idf là một phương pháp thường được sử dụng trong truy vấn thông tin và khai thác văn bản. Phương pháp này sử dụng để đánh giá tầm quan trọng của một từ trong một tài liệu và trong toàn bộ cơ sở dữ liệu tài liệu. Tầm quan trọng tăng lên tương ứng với số lần một từ xuất hiện trong tài liệu. Biến thể của tf\*idf thường được sử dụng như một công cụ tìm kiếm và xếp hạng mức độ phù hợp của một tài liệu ứng với một truy vấn của người dùng.

Áp dụng phương pháp này cho bài toán tìm kiếm ảnh, với mỗi ảnh được biểu diễn bằng 1 vector k chiều (k-vector)  $V_d = (t_1, \dots, t_i, \dots, t_k)$ , ta được công thức sau:

$$t_i = \frac{n_{id}}{n_d} \log \frac{N}{n_i}$$

Với

- $n_{id}$  là tổng số descriptor của ảnh d gán vào cluster i.
- $n_d$  tổng số descriptor của ảnh d.
- N tổng số ảnh có trong CSDL.
- $n_i$  tổng số ảnh có descriptor được gán vào cluster i.
- Để so khớp sự giống nhau 2 ảnh ta tính khoảng cách giữa chúng bằng độ đo khoảng cách Euclidean hay Cosine

#### 2.4. Trích xuất đặc trưng từ kho dữ liệu hình ảnh Oxford Building

Tiến hành trích xuất các điểm đặc trưng cho tất cả các ảnh trong kho dữ liệu. Lưu trữ các vector đặc trưng trích xuất được vào file index.mat để lưu trữ.

Code chương trình thực hiện index kho dữ liệu

```
function IndexDataset()
    strFileName = 'index.mat';
    rootFolder = fullfile('oxi_test');
    imds = imageDatastore(rootFolder);
    imageIndex = indexImages(imds);
    save(strFileName, 'imageIndex');
end
```

Dòng đầu tiên khởi tạo tên file index.mat để lưu trữ dữ liệu index kho hình ảnh.

Dòng tiếp theo lấy đường dẫn đầy đủ đến thư mục chứa các ảnh cần index. Sau đó khởi tạo một imageDataStore để lưu trữ toàn bộ hình ảnh.

Thực hiện index các hình ảnh trong **imds** sử dụng hàm **indexImages** do matlab cung cấp. Kết quả trả về là một ma trận chứa id của ảnh và các vector đặc trưng trong inverted Index.

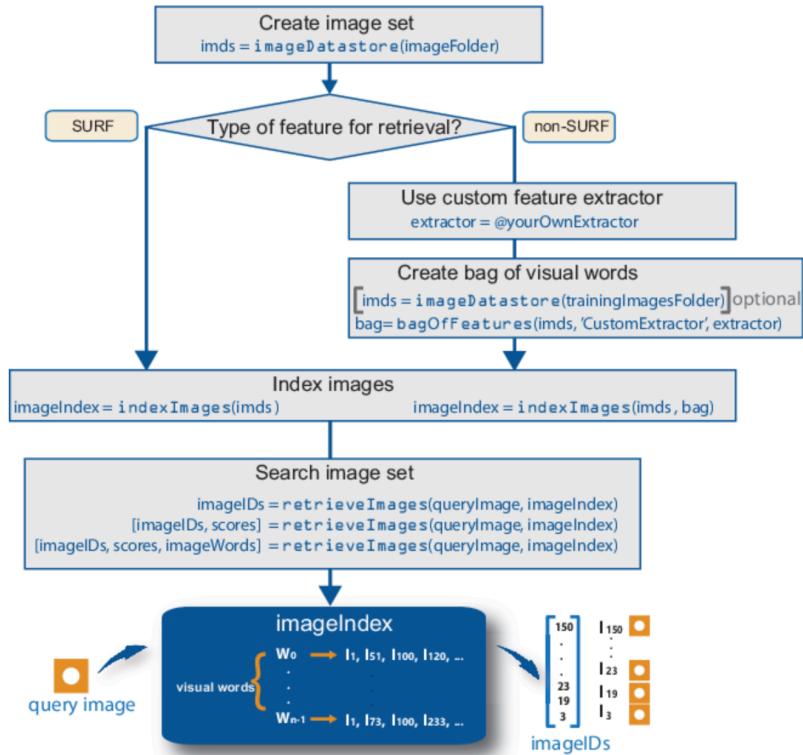
Lưu ma trận này vào file index.mat để sử dụng cho việc tìm kiếm một hình ảnh trong kho dữ liệu.

## **2.5. Xây dựng chương trình tìm kiếm**

Hệ thống khởi tạo bộ từ điển từ thị giác (“visual words”), tập các từ này được xây dựng từ kho dữ liệu ban đầu, được gọi là túi từ (hay bộ từ điển dùng trong kho dữ liệu). Các hình ảnh được đánh chỉ mục (index) tương ứng, một ánh xạ từ ảnh gốc đến một vector đặc trưng đại diện cho hình ảnh đó mà dựa vào đó máy tính có thể xác định độ tương đồng giữa ảnh truy vấn và ảnh hiện có trong kho dữ liệu. Giá trị của các vector đặc trưng này được xác định bằng số lần xuất hiện visual word đó trong bức ảnh cần index. Mỗi ảnh sẽ được ánh xạ thành một vector có số chiều bằng số lượng từ có trong túi (từ điển), trọng số mỗi chiều của vector chính là số lần xuất hiện của từ đó trong ảnh.

Khi người dùng cung cấp một ảnh query, ảnh này sẽ được ánh xạ về một vector đặc trưng. Hệ thống xác định mức độ tương đồng giữa ảnh truy vấn và các ảnh có trong kho dữ liệu, bằng cách tính toán khoảng cách (Cosine, Euclidian, ...) tập các kết quả có độ tương cao nhất sẽ được trả về.

Hệ thống mô tả bằng sơ đồ bên dưới, quy trình làm việc của hệ thống CBIR trên Matlab. Sử dụng phương pháp Bag of visual words, với trích xuất đặc trưng SURF để tăng tốc độ tính toán. Ngoài ra matlab còn hỗ trợ áp dụng các phương pháp trích xuất đặc trưng khác để tạo ra túi từ cũng như xác định độ đo tương đồng.



Mô hình triển khai chương trình tìm kiếm hình ảnh trên MatLab theo phương pháp Bag of words [4]

Chương trình tìm kiếm hình ảnh:

```

function SearchImage(imagePath)
%Load and show image query
queryImage = imread(imagePath);
figure(1)
imshow(queryImage);

load('index.mat');
imageResults = retrieveImages(queryImage, imageIndex);
fprintf('\n So ket qua: %d\n', size(imageResults,1));
index = 0;
imgResults = [];
for(i = 1:size(imageResults,1))
    index = index + 1;
    imgResult = imread(imageIndex.ImageLocation{imageResults(i, 1)});
    thumbnail = imresize(imgResult,[300 300]);
    %Show multi image results in one figure
    imgResults = cat(4,imgResults,thumbnail);
end
figure(2)
montage(imgResults);
end

```

## Chương III: Kết quả

Thực hiện index 327 tấm ảnh trong kho dữ liệu đặt tên là oxi\_test

```
Command Window
>> IndexDataset
Creating an inverted image index using Bag-Of-Features.
-----
Creating Bag-Of-Features.
-----
* Selecting feature point locations using the Detector method.
* Extracting SURF features from the selected feature point locations.
** detectSURFFeatures is used to detect key points for feature extraction.

* Extracting features from 327 images...done. Extracted 510298 features.

* Keeping 80 percent of the strongest features from each category.

* Balancing the number of features across all image categories to improve clustering.
** Image category 1 has the least number of strongest features: 408238.
** Using the strongest 408238 features from each of the other image categories.

* Using K-Means clustering to create a 20000 word visual vocabulary.
* Number of features      : 408238
* Number of clusters (K)   : 20000

* Initializing cluster centers...100.00%.
* Clustering...completed 39/100 iterations (~8.39 seconds/iteration)...converged in 39 iterations.

* Finished creating Bag-Of-Features

Encoding images using Bag-Of-Features.
-----
* Encoding 327 images...done.
fx Finished creating the image index.
```

Chạy chương trình **IndexDataset** từ cửa sổ Command Window của Matlab

Chương trình tiến hành tạo một inverted image index (tránh lãng phí tài nguyên và thời gian tính toán đối với trường hợp sparse vector) sử dụng Bag of Features.

Khi gọi hàm imageDataStore với tham số **location** (đường dẫn đến thư mục chứa các hình ảnh trong kho dữ liệu cần được index). Mặc định SURF features được sử dụng để trích xuất các điểm đặc trưng cho mỗi hình ảnh trong kho dữ liệu.

Kết quả được 510298 điểm được trung được phát hiện đối với 327 tấm ảnh trong folder oxi\_test.

Chương trình tiếp tục lọc lại chỉ lấy 80% điểm đặc trưng mạnh từ mỗi mục => còn lại 408238 điểm đặc trưng để đưa vào gom cụm xác định visual words.

Thuật toán k-Means được sử dụng với k là 20000 => túi từ (bộ từ điển) có 20000 từ.

Quá trình cluster kết thúc kết quả ta thu được bộ từ điển lưu trong biến **imds**.

Từ bộ từ điển thu được chương trình tiếp tục index tất cả các hình ảnh có trong kho dữ liệu. Mỗi ảnh sẽ được trích xuất đặc trưng thành các từ, chuyển thành một vector với 20000 chiều tương ứng với số lượng từ có trong từ điển, trọng số của mỗi từ là số lần xuất hiện của từ đó trong bức ảnh. Vector này được xử lý inverted index để tiết kiệm không gian và thời gian tính toán.

Kết quả index ta thu được ma trận index, ghi ma trận này xuống file để lưu trữ sử dụng cho bước truy vấn bằng hình ảnh trong kho dữ liệu.

Kết quả thu được của chương trình với các ảnh query:

*SearchImage('oxi\_test/all\_souls\_000001.jpg')*

ST T	Ảnh Query	Kết quả
1.		 <p>Hệ thống trả về ảnh đầu tiên khớp với ảnh query.</p> <p><b>Kết quả đúng: 12/20</b></p>
2		 <p>Hệ thống trả về ảnh đầu tiên khớp với ảnh query.</p> <p><b>Kết quả đúng 14/20</b></p>

3



Hệ thống trả về ảnh đầu tiên khớp với ảnh query.

**Kết quả đúng 6/20**



4



Hệ thống trả về ảnh đầu tiên khớp với ảnh query.

**Kết quả đúng 11/20**



5

(This row is empty)

	 <p>Hệ thống trả về ảnh đầu tiên khớp với ảnh query.</p> <p><b>Kết quả đúng 4/20</b></p>	
6	 <p>Hệ thống trả về ảnh đầu tiên khớp với ảnh query.</p> <p><b>Kết quả đúng 2/20</b></p>	

## Chương IV: Kết luận

Đồ án thực hiện cài đặt hệ thống tìm kiếm hình ảnh sử dụng phương pháp Bag of visual words. Với phương pháp SURF để trích xuất đặc trưng, ưu điểm của phương pháp này là thời gian index hình ảnh trong kho ảnh nhanh, đặc trưng tốt, chính xác, và thời gian đáp ứng truy vấn nhanh, kết quả trả về đạt độ chính xác cao.

Hệ thống tương lai sẽ xây dựng dạng module có thể thay đổi các phương pháp trích xuất đặc trưng khác cho hiệu quả cao và đổi với các kho dữ liệu ảnh có tính đặc thù (Cần chạy lại index trích xuất đặc trưng cho hình ảnh trong kho dữ liệu mới). Cài đặt Matlabs Runtime để có thể triển khai hệ thống dạng web, có giao diện người dùng để upload hình ảnh truy vấn và nhận kết quả thuận tiện.

## Tài liệu tham khảo

[1] <https://www.mathworks.com/help/vision/ug/image-retrieval-with-bag-of-visual-words.html>

[2] Dữ liệu: Bộ dữ liệu Oxford Building (5K),  
<http://www.robots.ox.ac.uk/~vgg/data/oxbuildings/>

[3] Demo tham khảo: <http://www.robots.ox.ac.uk/~vgg/research/oxbuildings/index.html>

[4] [https://docs.opencv.org/3.0-beta/doc/py\\_tutorials/py\\_feature2d/py\\_surf\\_intro/py\\_surf\\_intro.html](https://docs.opencv.org/3.0-beta/doc/py_tutorials/py_feature2d/py_surf_intro/py_surf_intro.html)

[5] Brian C. Becker (2009), SIFT Lecture, Computer Vision 16