

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN



TRUY VẤN THÔNG TIN THỊ GIÁC

Truy Vấn Ảnh Dựa Vào Nội Dung Thị Giác

Giảng viên hướng dẫn

Thầy Võ Hoài Việt

Thầy Nguyễn Trọng Việt

Thầy Phạm Minh Hoàng

Sinh viên thực hiện

Võ Nguyễn Hoàng Kim

21127090

Phụ lục

A.	BẢNG ĐÁNH GIÁ.....	3
I.	Đánh giá hệ thống truy vấn.....	3
II.	Đánh giá đồ án.....	3
B.	NỘI DUNG BÁO CÁO.....	3
I.	Tổng quan hệ thống truy vấn ảnh	3
II.	Tập dữ liệu.....	4
1.	Tập dữ liệu CD.....	4
2.	Tập dữ liệu TMBuD	4
III.	Xử lý tập dữ liệu.....	4
1.	Đánh nhãn dữ liệu	4
2.	Lưu trữ dữ liệu	6
IV.	Chuẩn bị cơ sở dữ liệu	7
1.	Rút trích đặc trưng.....	7
2.	Mô hình Bag of Features.....	7
3.	Lưu trữ dữ liệu	8
V.	Thực hiện truy vấn.....	8
1.	Tải cơ sở dữ liệu.....	8
2.	Tải dữ liệu truy vấn.....	8
3.	Truy vấn.....	9
4.	Độ đo.....	9
VI.	Kết quả.....	9
1.	Kết quả hiển thị	9
2.	Thống kê kết quả	10
3.	Nhận xét.....	11
VII.	Một số vấn đề xảy ra.....	12
1.	Cải thiện độ chính xác.....	12
2.	Cải thiện tốc độ truy vấn.....	12
VIII.	Hướng dẫn sử dụng	12
C.	TÀI LIỆU THAM KHẢO	13

A. BẢNG ĐÁNH GIÁ

I. Đánh giá hệ thống truy vấn

Tác vụ	Chức năng	Mức độ hoàn thành	Ghi chú
Chuẩn bị cơ sở dữ liệu	Xử lý dữ liệu ban đầu	100%	Hoàn thành
	Trích xuất đặc trưng SIFT, ORB, Color Histogram và Color Correlogram	100%	Hoàn thành
	Tìm hiểu và xây dựng mô hình phù hợp với các đặc trưng (Cục bộ và Toàn cục)	100%	Hoàn thành
	Tìm hiểu và cài đặt độ đo mAP	100%	Hoàn thành
Thực hiện truy vấn	Truy vấn bằng các đặc trưng (SIFT, ORB, Color Histogram, Color Correlogram)	100%	Hoàn thành
	Xây dựng giao diện trình bày ảnh truy vấn và kết quả truy vấn được	100%	Hoàn thành
	Tính toán độ đo mAP cho tập kết quả truy vấn và thời gian truy vấn trung bình cho một ảnh	100%	Hoàn thành

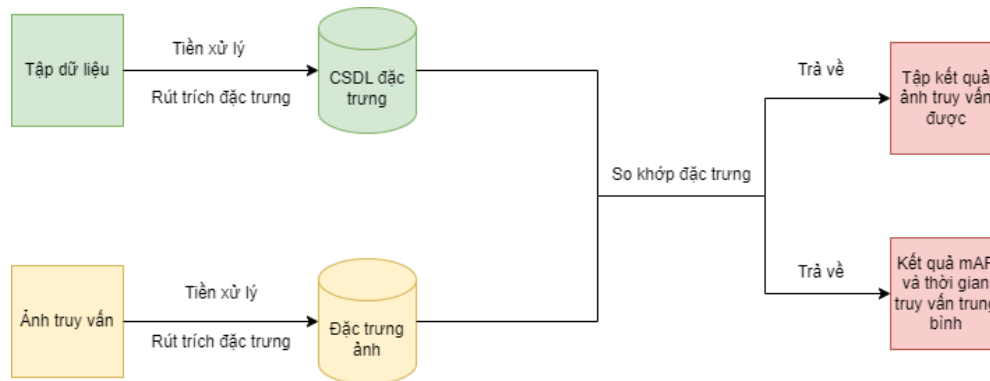
II. Đánh giá đồ án

STT	Yêu cầu	Mức độ hoàn thành	Ghi chú
1	Cho phép người dùng truy vấn ảnh sử dụng Color Histogram, Color Correlogram và đặc trưng cục bộ SIFT, ORB. Đánh giá kết quả thực hiện của hệ thống truy vấn.	100%	Hoàn thành
2	Tìm hiểu các đặc trưng để tăng hiệu quả để nâng cao độ chính xác của hệ thống. So sánh kết quả truy vấn của các phương pháp.	50%	Vẫn chưa tìm được cách để nâng cao độ chính xác
3	Tìm hiểu và cài đặt cải tiến tốc độ tìm kiếm ảnh trong CSDL.	100%	Hoàn thành

B. NỘI DUNG BÁO CÁO

I. Tổng quan hệ thống truy vấn ảnh

- Hệ thống truy vấn ảnh được thực hiện dựa trên hai tác vụ chính là Chuẩn bị cơ sở dữ liệu và Thực hiện truy vấn. Sơ đồ dưới đây thể hiện quy trình tổng quát mà hệ thống này thực hiện:



II. Tập dữ liệu

1. Tập dữ liệu CD

- Link tải tập dữ liệu: <https://drive.google.com/drive/folders/1EjOew0oXgnz5aeqPpNbIB8IrpUTnPdT1>
- Tập dữ liệu CD được chia thành 2 folders chính là **training_set** và **TestImages**, trong đó:
 - o **training_set**: là folder chứa các ảnh dành cho việc huấn luyện, với tổng số ảnh là 99. Theo quan sát và phân tích (thủ công) về tính tương đồng giữa các ảnh, em thấy rằng, các ảnh được đặt tên theo cú pháp <Nhãn ảnh>_<Thứ tự các ảnh trong cùng một nhãn>. Đây chính “mẫu chốt” để thực hiện việc tạo thông tin về ảnh và nhãn tương ứng cho tập dữ liệu.
 - o **TestImages**: là folder chứa các ảnh dành cho việc kiểm thử, với tổng số ảnh là 10. Giống với tập huấn luyện, tập kiểm thử này cũng không tồn tại bất cứ thông tin về ảnh và nhãn của chúng. Do đó, việc nhận biết và đánh nhãn cho chúng sẽ được thực hiện thủ công.

2. Tập dữ liệu TMBuD

- Link tải tập dữ liệu: <https://github.com/CipiOrhei/TMBuD>
- Tập dữ liệu TMBuD cung cấp khá nhiều thông tin về tập dữ liệu, nhưng trong phạm vi hệ thống này, chúng ta chỉ sử dụng 2 tập tin chính là **DATASET SPLIT.csv** và **images**, trong đó:
 - o **DATASET SPLIT.csv**: là file chứa thông tin của ảnh, trong đó có 2 thông tin quan trọng mà ta cần lưu trữ lại là **Picture Name** (xem như id ảnh) và **Building Name** (xem như là nhãn).
 - o **images**: là thư mục chứa tất cả các ảnh của tập dữ liệu. Do tập này cơ bản không chia đầy đủ các ảnh train và test (có thể thấy điều này trong các cột còn lại của file csv), do đó, ta sử dụng toàn bộ ảnh trong thư mục này để làm dữ liệu cho tập huấn luyện.

III. Xử lý tập dữ liệu

1. Đánh nhãn dữ liệu

a. Tập huấn luyện

- Tập dữ liệu CD: Như đã trình bày ở mục II, mẫu chốt để thực hiện đánh nhãn cho tập dữ liệu này chính là dựa vào tên của nó với cấu trúc <Nhãn ảnh>_<Thứ tự các ảnh trong cùng một nhãn>. Khi đó, ta chỉ cần thiết lập hàm là có thể có được nhãn ảnh tương ứng cho mỗi ảnh.
 - o Hàm được sử dụng trong tác vụ này là **createLabel()** thuộc lớp **Dataset**. Hàm sẽ duyệt qua toàn bộ ảnh có trong folder training, nó thực hiện lấy ra hai thông tin quan trọng của ảnh là **FileName** và **Label** (Tức tên ảnh và nhãn ảnh).

- Ví dụ: Các ảnh **01_1.jpg**, **01_2.jpg** và **01_03.jpg** thể hiện rằng chúng cùng một nhóm. Vì vậy, nhãn của chúng sẽ là **01**. Tuy nhiên, tên ảnh không chỉ dừng lại là **01_1.jpg** mà nó sẽ bao gồm cả đường dẫn tương đối (đứng tại vị trí chạy file exe) đến nơi lưu trữ ảnh.
- Tập dữ liệu TMBuD: Như đã trình bày ở mục II, ta sẽ sử dụng **Building Name** để thực hiện tác vụ đánh nhãn cho ảnh. Tuy nhiên, các thông tin này chứa trong file csv (có nhiều cột không cần thiết cho hệ thống), điều này có thể ảnh hưởng đến hiệu suất của hệ thống, do đó, ta sẽ thực hiện lấy các thông tin này từ file csv và chuyển về kiểu file khác để đồng nhất với tập CD.
 - Hàm được sử dụng trong tác vụ này là **readCSV()** thuộc lớp **Dataset**. Hàm sẽ mở file csv được truyền vào, tiến hành đọc từng dòng của file. Trong file csv có dòng đầu tiên chứa các thông tin của cột tương ứng, do đó, ta sẽ bỏ qua dòng này (bằng cách thiết lập cờ cho hàm để nhận biết điều này).
 - Sau đó, tiến hành đọc và xử lý dữ liệu của từng dòng (các thông tin trên cùng một dòng được phân tách bởi ký tự “,”).
 - Đối với tên ảnh, do thông tin lưu trên file csv và tên được lưu trong folder **images** không đồng nhất, do đó, ta bắt buộc phải xử lý dữ liệu ở file csv để có thể lấy đúng thông tin về tên file (điều này ảnh hưởng đến việc mở ảnh khi xử lý). Đối với các ảnh trong folder **images**, tên ảnh là một chuỗi thứ tự có độ dài là 5, do đó, khi thực hiện đọc thông tin từ file csv, ta sẽ thêm các ký tự 0 ở phía trước sao cho độ dài tên ảnh vừa đúng 5 thì dừng. Sau đó, kết hợp (nối) với đường dẫn đến thư mục chứa ảnh để cho ra tên ảnh hoàn chỉnh.
 - Đối với nhãn, chúng ta chỉ việc lấy giá trị mà không cần chỉnh sửa bất cứ thông tin gì.
 - Khi này, kết quả đạt được sau khi gọi hàm sẽ là thông tin về tên ảnh và nhãn của toàn bộ ảnh trong tập dữ liệu TMBuD.

b. Tập kiểm thử

- Tập dữ liệu CD:
 - Tập dữ liệu này dù đã có sẵn tập kiểm thử, tuy nhiên lại thiếu thông tin về nhãn tương ứng của từng ảnh, việc này sẽ gây khó khăn cho việc cài đặt và sử dụng độ đo mAP ở kết quả truy vấn. Do đó, ta sẽ tạo một file TXT để chứa các thông tin về tập kiểm thử này.
- Tập dữ liệu TMBuD:
 - Như đã trình bày khó khăn trong việc chia tập huấn luyện và kiểm thử của tập dữ liệu này ở mục II, do đó, ta sẽ tự tạo dữ liệu kiểm thử cho tập dữ liệu này. Tập dữ liệu kiểm thử được lấy ngẫu nhiên 10 ảnh trong thư mục **images**, thông tin về tập này sẽ được lưu trữ trong một file TXT.
- Việc tạo file TXT chứa thông tin về tập kiểm thử được khởi tạo thủ công, trong đó:
 - Dòng đầu, sẽ là đường dẫn đến nơi lưu trữ các ảnh của tập kiểm thử (đây là đường dẫn tương đối).
 - Các dòng tiếp theo sẽ là thông tin của các ảnh có trong tập kiểm thử, với mỗi dòng sẽ chứa tên ảnh và nhãn tương ứng của nó.

```

|../Data/TMBuD-main/TMBuD-main/images/
00004.png Orthodox Metropolitan Cathedral
00513.png Bruck House
03404.png Serbian Community House
04202.png Military Casino of Timisoara
05604.png Fire Brigade 2
06103.png The National Opera House
09001.png Building 7 Str. E. Ungureanu
10602.png Franz Anheuer Palace
11211.png Marczel Kaldori Palace
12506.png Banatul Philharmonic of Timisoara

```

2. Lưu trữ dữ liệu

- Do yêu cầu của đề án là thực hiện truy vấn với hai tập dữ liệu là CD và TMBuD, nhưng cách tổ chức dữ liệu của chúng ban đầu không giống nhau. Do đó, cần thực hiện đồng nhất cách lưu trữ dữ liệu của chúng để việc chuẩn bị cơ sở dữ liệu được thực hiện dễ hơn.
- Ở đây, ta sử dụng file XML để lưu trữ dữ liệu của chúng dưới dạng các thẻ:
 - o Mỗi ảnh trong một tập dữ liệu sẽ là một phần tử và có các thẻ để chứa thông tin cần thiết của ảnh gồm:
 - filename: đường dẫn đến file ảnh.
 - label: nhãn tương ứng của ảnh.
 - sift, orb, hist, correlogram: là các dữ liệu về đặc trưng tương ứng của ảnh.
 - o Dưới đây là ảnh minh họa cho một phần tử trong tập tin **CD_LocalFeature.xml** với các thông tin về đặc trưng SIFT và ORB

```

<?xml version="1.0"?>
<opencv_storage>
<CD>
  <item>
    <filename>../../Data/CD_data/training_set/training_images\01_1.jpg</filename>
    <label>"01"</label>
    <sift type_id="opencv-matrix">
      <rows>1</rows>
      <cols>100</cols>
      <dt>f</dt>
      <data>
        4. 33. 2. 1. 5. 4. 7. 11. 5. 9. 16. 0. 4. 4. 8. 10. 19. 4. 3. 2.
        2. 12. 9. 3. 4. 8. 7. 3. 6. 10. 4. 2. 0. 6. 7. 7. 8. 2. 9. 8.
        17. 2. 10. 3. 2. 6. 14. 54. 7. 7. 7. 14. 7. 4. 10. 3. 8. 3. 4.
        7. 9. 6. 6. 14. 8. 4. 11. 7. 4. 8. 4. 3. 2. 12. 6. 5. 4. 4. 8.
        4. 6. 6. 5. 10. 5. 4. 6. 8. 2. 11. 8. 7. 2. 7. 7. 2. 4. 6. 7. 6.</data></sift>
    <orb type_id="opencv-matrix">
      <rows>1</rows>
      <cols>100</cols>
      <dt>f</dt>
      <data>
        7. 3. 2. 5. 6. 3. 3. 2. 3. 8. 4. 6. 1. 1. 12. 4. 6. 13. 8. 9.
        10. 5. 2. 3. 4. 1. 2. 9. 7. 2. 4. 6. 2. 1. 6. 4. 9. 3. 5. 0. 3.
        0. 0. 2. 4. 1. 0. 9. 4. 0. 8. 3. 3. 2. 6. 2. 3. 4. 4. 2. 5. 9.
        7. 12. 7. 6. 2. 7. 6. 7. 13. 2. 7. 8. 2. 7. 5. 3. 18. 4. 1. 4.
        13. 5. 3. 2. 2. 4. 2. 19. 12. 2. 3. 0. 4. 8. 9. 7. 6. 6.</data></orb>
    <hist type_id="opencv-matrix">
      <rows>0</rows>
      <cols>0</cols>
      <dt>u</dt>
      <data></data></hist>
    <correlogram type_id="opencv-matrix">
      <rows>0</rows>
      <cols>0</cols>
      <dt>u</dt>
      <data></data></correlogram>
    <combine type_id="opencv-matrix">
      <rows>0</rows>
      <cols>0</cols>
      <dt>u</dt>
      <data></data></combine></item>
</item>

```

IV. Chuẩn bị cơ sở dữ liệu

1. Rút trích đặc trưng

- Do tập dữ liệu khá lớn, nếu giữ nguyên ảnh ban đầu để thực hiện rút trích đặc trưng sẽ gây ảnh hưởng đến tài nguyên và tốc độ xử lý. Do đó, việc tiền xử lý là cần thiết, là giảm kích thước của ảnh xuống.
 - o Bước này, ta sử dụng hàm **cv::resize()** để giảm đi 50% kích thước của ảnh (cả chiều dài và rộng).
- Các đặc trưng được sử dụng trong chương trình này gồm SIFT, ORB, Color Histogram và Color Correlogram:
 - o Với các đặc trưng SIFT và ORB:
 - Trước hết, cần chuyển đổi ảnh về dạng ảnh xám. Sau đó dựa vào hàm **cv::detectAndCompute()** để rút trích đặc trưng tương ứng. Kết quả của descriptor sẽ được lưu tạm vào biến tương ứng của ảnh.
 - o Đối với Color Histogram:
 - Vì các ảnh được sử dụng trong tập dữ liệu đều là ảnh màu, do đó, ta cũng sẽ thực hiện tính toán Color Histogram đối với ảnh màu.
 - Ta thực hiện tính toán giá trị histogram cho từng kênh màu riêng biệt (các kênh màu được tính toán với 256 bin, giới hạn từ 0 đến 255). Việc này được hỗ trợ bởi hàm **cv::calcHist()**.
 - Sau đó, thực hiện chuẩn hoá từng kênh màu và thêm chúng vào cùng một ma trận duy nhất (sử dụng hàm **reshape()** để chuyển đổi từng kênh màu thành một hàng duy nhất). Trước khi trả về kết quả, thực hiện **reshape** một lần nữa với ma trận trên để chuyển đổi kết quả histogram của một ảnh về thành một hàng duy nhất (mảng 1 chiều) giúp đơn giản hóa cho việc so khớp ở sau.
 - o Đối với Correlogram
 - Đối với đặc trưng này, ta sử dụng không gian màu HSV để tính toán và thực hiện. Do đó, cần chuyển đổi ảnh ban đầu sang không gian màu HSV.
 - Ta giả định rằng, màu sắc trong ảnh sẽ được lượng tử hoá thành 16 mức và xác định khoảng cách tối đa để xem xét các cặp màu là 5 pixel. Thực hiện lượng tử hoá để giảm số lượng màu sắc cũng như giúp đơn giản hoá việc tính toán.
 - Khởi tạo một ma trận Correlogram để lưu trữ thông tin về các cặp màu và tần suất xuất hiện của chúng trong ảnh. Thực hiện duyệt qua từng pixel, xác định chỉ số màu của chúng và kiểm tra các pixel lân cận (trong phạm vi đã thiết lập ở trên) để tính toán chỉ số màu của chúng. Cuối cùng, cập nhật lại ma trận Correlogram dựa trên các cặp màu đã tìm thấy.
 - Lưu trữ tạm kết quả ma trận Correlogram của ảnh vào biến tương ứng.

2. Mô hình Bag of Features

- Mô hình Bag of Features (nguồn gốc là Bag of Word) là một phương pháp tốt để biểu diễn hình ảnh bằng các đặc trưng cục bộ (như SIFT và ORB). Việc xây dựng mô hình trong mã nguồn này được thực hiện chính dựa vào hàm **cv::kmeans()**, với mục đích tìm ra các đặc trưng đại diện (là các centers tương ứng trong thuật toán kmeans) trong dữ liệu.

- Dữ liệu được sử dụng làm dữ liệu đầu vào cho thuật toán **kmeans** này chính là một descriptor lớn (được gom lại từ các descriptor của ảnh trong tập dữ liệu) theo một đặc trưng được sử dụng để truy vấn.
- Khi thực hiện gọi mô hình này, ta sẽ thu được các centers tương ứng (là các đặc trưng nổi bật trong tập dữ liệu).
 - o Các centers này được lưu lại (trong thư mục **Database/ <Data tương ứng (CD/TMBuD)/Centroids**) dưới dạng file nhị phân **.bin**. Tên file của chúng thể hiện rằng đó là file lưu trữ centers của đặc trưng nào, với số lượng từ là bao nhiêu.
 - Ví dụ, với file **Centroid_SIFT_100.bin** sẽ cho biết rằng file này lưu trữ center của đặc trưng SIFT với số lượng từ là 100.
- Sau khi thực hiện việc lượng tử hoá các đặc trưng bằng mô hình Bag of Features, ta sẽ tiến hành tính toán sự xuất hiện của chúng trong ảnh. Kết quả này được biểu diễn dưới dạng một histogram.
 - o Ta sẽ thực hiện riêng biệt đối với đặc trưng SIFT, ORB.
 - o Histogram (kết quả trên) sẽ được lưu tạm vào biến tương ứng của ảnh.

3. Lưu trữ dữ liệu

- Giá trị của thẻ sẽ tương ứng với giá trị mà nó lưu trữ.
- Ở đây, các thẻ đại diện cho đặc trưng “sift” hay “orb” chỉ lưu trữ histogram tương ứng mà nó tính toán được (không phải là ma trận các đặc trưng descriptor mà nó đã rút trích).
- Do các dữ liệu về đặc trưng toàn cục chiếm dung lượng khá lớn, nên nếu lưu trữ toàn bộ đặc trưng trong cùng một file XML có thể ảnh hưởng đến việc tải dữ liệu khi thực hiện truy vấn, do đó, ta sẽ lưu trữ chúng riêng biệt với nhau:
 - o Đặc trưng cục bộ (SIFT và ORB) có kích thước khá nhỏ, do đó, ta có thể lưu trữ chúng chung với nhau trong cùng một file là **<Data tương ứng>_LocalFeature.xml**. Các đặc trưng khác như Histogram và Correlogram sẽ lưu dưới dạng ma trận riêng.
 - o Đặc trưng toàn cục (Histogram và Correlogram) có kích thước lớn, nên mỗi đặc trưng sẽ lưu một file XML riêng.
 - Đối với Histogram, tên file sẽ là **<Data tương ứng>_Histogram.xml**.
 - Đối với Correlogram, tên file sẽ là **<Data tương ứng>_Correlogram.xml**.

V. Thực hiện truy vấn

1. Tải cơ sở dữ liệu

- Khi người dùng muốn thực hiện truy vấn dựa trên một đặc trưng cụ thể, hệ thống sẽ tiến hành tải cơ sở dữ liệu của đặc trưng tương ứng lên hệ thống.
 - o Ví dụ, khi người dùng muốn truy vấn ảnh bằng đặc trưng SIFT, hệ thống chỉ tải mỗi đặc trưng SIFT trong cơ sở dữ liệu lên hệ thống để phục vụ tác vụ truy vấn.
- Việc tải duy nhất một đặc trưng tương ứng trong cơ sở dữ liệu sẽ giúp cho hệ thống tiết kiệm thời gian thực hiện truy vấn và giảm bớt tải nguyên sử dụng để thực hiện hoặc tính toán.

2. Tải dữ liệu truy vấn

- Để có thể thực hiện truy vấn ảnh, người dùng cần cung cấp đủ thông tin về tên ảnh (đường dẫn đến nơi lưu trữ ảnh) và nhãn tương ứng (để hệ thống có thể đánh giá mAP). Tuy nhiên, hệ thống hướng đến việc người dùng muốn truy vấn một tập ảnh (sử dụng một tập ảnh để kiểm thử hệ thống), nên đã thiết lập việc truy vấn thông qua bước trung gian là đọc dữ liệu từ file txt (được trình bày trong mục III).

- Với dữ liệu có được từ việc đọc file, ta tiến hành đọc các ảnh và nhãn tương ứng của nó. Khi đó, ta sẽ có được tập dữ liệu cần truy vấn.

3. Truy vấn

- Để đồng nhất tính chất với các cơ sở dữ liệu, ta cũng tiến hành tiền xử lý bằng cách giảm kích thước của ảnh đi 50%.
- Chương trình cho phép truy vấn dữ liệu dựa trên các đặc trưng SIFT, ORB, Color Histogram, Color Correlogram. Điều này được quyết định dựa trên tham số mà người dùng truyền vào.
 - Ví dụ: người dùng mong muốn truy vấn ảnh bằng đặc trưng **SIFT**, chương trình sẽ thực hiện như sau:
 - Rút trích đặc trưng của ảnh truy vấn bằng SIFT, lưu lại kết quả của descriptor.
 - Tải dữ liệu (từ file) center của đặc trưng SIFT và tính toán histogram cho ảnh truy vấn bằng descriptor ở trên và center. Ta sẽ gọi kết quả của việc này là **query_histogram**.
 - So sánh **query_histogram** với các histogram (SIFT) có trong cơ sở dữ liệu để tìm ảnh tương tự với ảnh truy vấn. Điều này được thực hiện thông qua việc tính toán khoảng cách giữa **query_histogram** với từng histogram (SIFT) có trong cơ sở dữ liệu (ghi nhận lại kết quả của khoảng cách này).
 - Sau khi tính toán khoảng cách (độ dị biệt) giữa **query_histogram** và toàn bộ cơ sở dữ liệu, ta thực hiện sắp xếp theo thứ tự giảm dần.
 - Cuối cùng, lấy ra **K** ảnh có khoảng cách bé nhất để trả về kết quả truy vấn.
- Một số lưu ý trong hệ thống truy vấn:
 - Đối với tập TMBuD, do ta sử dụng các ảnh trong tập huấn luyện để làm kiểm thử, do đó, chắc chắn nó sẽ luôn trả về chính ảnh đó (ở vị trí đầu tiên) khi thực hiện truy vấn. Vì thế, ta sẽ loại bỏ kết quả đó khỏi tập kết quả truy vấn.
 - Đối với tập CD, ta không chắc rằng tất cả các ảnh được sử dụng trong tập kiểm thử có được lấy từ tập huấn luyện hay không, do đó, ta vẫn sử dụng kết quả đầu tiên của tập kết quả truy vấn được.

4. Độ đo

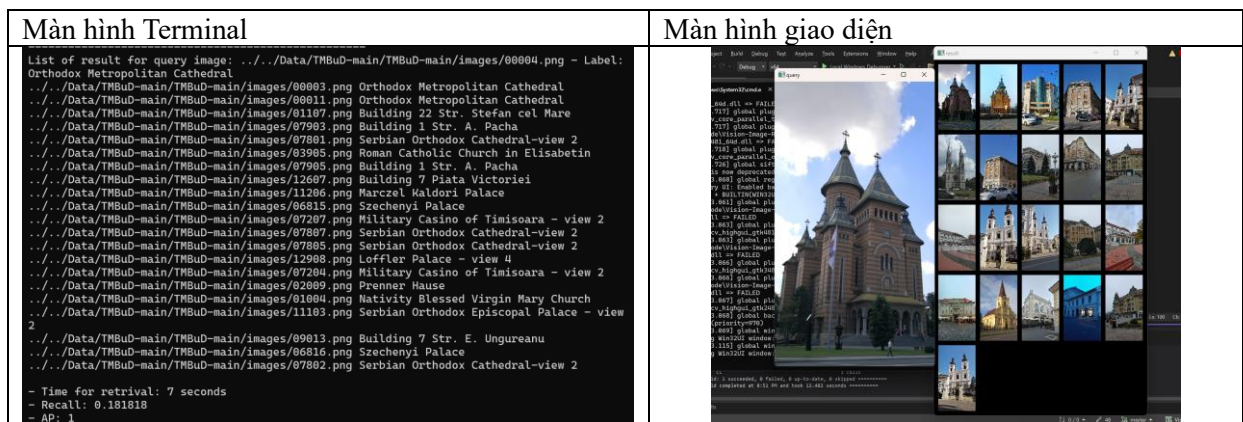
- Trong bài này, ta sử dụng độ đo chính là **mean Average Precision (mAP)** để đánh giá kết quả truy vấn của chương trình, trong đó:
 - **AP**: là giá trị trung bình của các giá trị Precision tại mỗi điểm mà một ảnh đúng được tìm thấy trong danh sách kết quả. Việc tính **AP** được thực hiện như sau:
 - Với kết quả truy vấn trả về theo thứ tự giảm dần của độ tương đồng, ta tính precision tại mỗi vị trí mà một ảnh đúng xuất hiện.
 - Sau đó, tính trung bình các giá trị precision này.
 - Giá trị của **mAP** được tính bằng trung bình của các **AP** khi thực hiện truy vấn một tập các ảnh đưa vào (tập test).
- Ngoài ra, chương trình còn cung cấp độ bao phủ **Recall** cho từng kết quả trong tập test, nó được tính dựa vào số ảnh truy vấn đúng chia cho số lượng ảnh có nhãn tương ứng trong tập dữ liệu.

VI. Kết quả

1. Kết quả hiển thị

- Khi người dùng truyền lệnh để chương trình hiển thị các ảnh truy vấn được, chương trình sẽ hiển thị hai cửa sổ cùng lúc, trong đó:
 - Query: là cửa sổ hiển thị ảnh mà người dùng sử dụng để truy vấn.

- Result: là kết quả các ảnh truy vấn được, được sắp xếp theo độ tương đồng giảm dần từ trái sang phải và từ trên xuống dưới.
- Chương trình sẽ đợi lệnh để đóng hai cửa sổ này (người dùng nhấn bất kỳ phím nào cũng có thể đóng cửa sổ). Sau khi cửa sổ giao diện về kết quả được đóng lại, chương trình sẽ hiển thị thông tin của ảnh vừa được thực hiện truy vấn gồm:
 - Dòng đầu tiên cung cấp thông tin về ảnh truy vấn bao gồm tên ảnh (đường dẫn đến nơi lưu trữ ảnh) và nhãn của nó
 - K dòng tiếp theo là kết quả mà hệ thống thực hiện truy vấn được, được sắp xếp theo độ tương đồng giảm dần
 - 3 dòng cuối hiển thị thông tin về kết quả của một lần truy vấn, lần lượt là thời gian truy vấn ảnh (tính từ lúc đọc thông tin của ảnh truy vấn và kết thúc khi có được kết quả truy vấn, không tính thời gian để thực hiện hiển thị kết quả trên giao diện), Độ bao phủ Recall và AP.
- Dưới đây là ảnh minh họa cho kết quả mà người dùng nhận được khi hiển thị ảnh truy vấn:



2. Thống kê kết quả

- Số liệu dưới đây được thống kê từ kết quả truy vấn trên tập ảnh TMBuD và CD với các đặc trưng SIFT, ORB, Histogram và Correlogram với các số lượng ảnh trả về lần lượt là 3, 5, 11, 21.
- Việc thực hiện truy vấn và thống kê lại kết quả được thực hiện trên tập ảnh kiểm thử (gồm 10 ảnh) và ghi nhận lại giá trị trung bình về độ chính xác (mAP) và thời gian truy vấn trung bình cho một ảnh (đơn vị là giây).

TMBuD				
K	SIFT	ORB	Histogram	Correlogram
3	0.783333	0.408333	0.416667	0.4
	12 secs / image	10.1 secs / image	8.6 secs / image	14.9 secs / image
5	0.753333	0.403333	0.416667	0.391667
	11.5 secs / image	7.2 secs / image	8 secs / image	14.1 secs / image
11	0.735202	0.412623	0.376667	0.388952
	5.7 secs / image	10.7 secs / image	9.5 secs / image	5.7 secs / image
21	0.718833	0.398385	0.376667	0.331031
	5.2 secs / image	4.6 secs / image	9.8 secs / image	13 secs / image

CD				
K	SIFT	ORB	Histogram	Correlogram

3	0.791667	0.633333	0	0
	1.2 secs / image	0.7 secs / image	0 secs / image	4.3 secs / image
5	0.791667	0.607222	0	0.04
	1.2 secs / image	0.9 secs / image	0.1 secs / image	4.6 secs / image
11	0.79246	0.60373	0.0358766	0.0785714
	1.9 secs / image	0.8 secs / image	0 secs / image	4.6 secs / image
21	0.700148	0.54342	0.0358766	0.0921269
	1.6 secs / image	0.7 secs / image	0.1 secs / image	5.5 secs / image

3. Nhận xét

- Về tập dữ liệu:
 - o Nhìn chung, các đặc trưng sử dụng truy trên hai tập dữ liệu cho ra tính chất kết quả tương đồng nhau (về thời gian lẫn độ chính xác). Tuy nhiên giữa tập dữ liệu TMBuD và CD có sự chênh lệch nhau về độ đa dạng (giữa các ảnh trong cùng một nhãn) và số lượng ảnh trong tập huấn luyện, do đó, không thể so sánh kết quả (độ chính xác) mà hệ thống trả ra giữa 2 tập dữ liệu.
- Về tốc độ:
 - o Tốc độ truy vấn ảnh được quyết định dựa trên nhiều yếu tố như kích thước của ảnh truy vấn, độ lớn của tập cơ sở dữ liệu đặc trưng, đặc trưng sử dụng để truy vấn, ...
 - o So về kích thước tập dữ liệu: với tập dữ liệu CD, kích thước của tập dữ liệu không quá lớn, do đó, thời gian truy vấn sẽ rất nhanh. Ngược lại, với tập dữ liệu TMBuD, kích thước lại rất lớn, dẫn đến thời gian truy vấn sẽ bị kéo theo, do đó, nó chỉ nằm ở mức ổn.
 - o So về các đặc trưng: có thể thấy, đặc trưng Color Histogram và ORB có phần nhanh hơn so với hai đặc trưng còn lại, đứng ở vị trí cuối cùng chính là đặc trưng Color Correlogram.
 - Điều này có thể giải thích do đặc trưng mà Color Correlogram trích xuất được là quá lớn (dù đã giảm số lượng các bins khi thực hiện tính toán). Dẫn đến việc so khớp tốn khá nhiều thời gian so với các đặc trưng còn lại.
- Về độ chính xác:
 - o Dựa vào bảng thống kê trên, độ chính xác khi thực hiện truy vấn với số lượng ảnh kết quả trả về càng nhỏ, thì giá trị của **mAP** càng cao (điều này đúng đối với các đặc trưng cục bộ, và ngược lại đối với các đặc trưng toàn cục).
 - Các giá trị K càng lớn, thì kết quả về độ chính xác cho các đặc trưng toàn cục càng được cải thiện.
 - o Tổng quan, đặc trưng SIFT cho ra kết quả tốt nhất so với các đặc trưng được sử dụng để truy vấn.
- Kết luận:
 - o Bảng sau đây sẽ thể hiện tổng quan về đánh giá đối các đặc trưng:

Đặc trưng	SIFT	ORB	Histogram	Correlogram
Đánh giá	Cho ra kết quả truy vấn với độ chính xác cao, tuy nhiên đánh đổi lại bằng thời gian truy vấn chỉ nằm ở mức tạm ổn, chưa thật sự nhanh.	Cho ra kết quả với độ chính xác tạm ổn, tuy nhiên thời gian truy vấn nhanh.	Kết quả cho ra thấp, chỉ đưa ra các ảnh đúng khi đặt số lượng ảnh truy vấn lớn, thời gian thực hiện truy vấn nhanh	Kết quả cho ra thấp, chỉ đưa ra các ảnh đúng khi đặt số lượng ảnh truy vấn lớn, thời gian thực hiện truy vấn nhanh

VII. Một số vấn đề xảy ra

- Trong mục này, em xin được trình bày vấn đề cũng như những nỗ lực mình đã thực hiện trong đồ án này.

1. Cải thiện độ chính xác

- Em nhận thấy ra mô hình Bag Of Feature (BOF) cho ra một kết quả khá tốt đối với đặc trưng SIFT, do đó, em quyết định sử dụng chúng (đặc trưng SIFT với mô hình BOFe) để có thể cải tiến mô hình.
- Việc sử dụng kết hợp một số thuật toán tiền xử lý như làm mịn ảnh bằng Gaussian và phát hiện biên cạnh đã được sử dụng, sau đó trích xuất đặc trưng SIFT trên chúng với mô hình BOF. Tuy nhiên, kết quả cho ra lại không khả quan như em mong đợi (độ chính xác cho ra thậm chí còn thấp hơn so với kết quả ban đầu), dẫn đến thất bại.
- Một cách khác mà em nghĩ đến là kết hợp SIFT và ORB, với ý tưởng rằng có thể rằng những đặc trưng mà ORB trích xuất được sau khi đưa qua SIFT sẽ có thể tìm thấy những đặc trưng tốt và quan trọng mà đã bị bỏ lỡ bởi ORB. Ý tưởng này được thực hiện bằng cách kết hợp descriptor giữa 2 đặc trưng này với nhau (điều này đã được thể hiện ở yêu cầu 2, tuy nhiên gặp một số trục trặc dẫn đến kết quả của sáng kiến này và kết quả của đặc trưng SIFT trùng nhau). Sau đó, em đã sửa và thử lại mã nguồn, thì kết quả lại không được như mong đợi, dẫn đến thất bại tiếp tục.
- Và nỗ lực cuối cùng là thử kết hợp giữa 2 kết quả truy vấn độc lập với nhau (là kết quả tốt nhất của đặc trưng cục bộ và toàn cục), là SIFT và Correlogram. Ý tưởng được đưa ra và thực hiện bằng cách lấy kết quả của SIFT và Correlogram nhân với trọng số tương ứng (trọng số này được thiết lập là 0.8 và 0.2, lần lượt cho đặc trưng SIFT và Correlogram) với mong muốn rằng với kết quả tốt của SIFT kết hợp với đặc trưng toàn cục của Correlogram sẽ tầm soát được các ảnh liên quan. Tuy nhiên điều này cũng không khả thi, do kết quả về độ chính xác mà Correlogram cho ra là rất tệ, kéo theo kết quả của SIFT cũng sẽ ảnh hưởng, do đó, làm giảm độ chính xác so với ban đầu.

2. Cải thiện tốc độ truy vấn

- Xét về cách cài đặt và xây dựng cơ sở dữ liệu, việc xây dựng cơ sở dữ liệu dưới mô hình BOF cho ra tốc độ truy vấn nhanh chóng so với việc sử dụng so khớp các đặc trưng (điều này có thể thấy rõ khi thực hiện truy vấn với các đặc trưng cục bộ và toàn cục).
- Bên cạnh đó, đối với các tập dữ liệu như TMBuD, số lượng ảnh trong tập dữ liệu rất nhiều, kết hợp với kích thước một ảnh khá lớn, do đó, nếu giữ nguyên kích thước ban đầu để thực hiện rút trích hoặc truy vấn dữ liệu, sẽ tiêu tốn thời gian khá lâu. Do đó, việc tiền xử lý bằng cách **resize()** là một điều cần thiết và quan trọng được sử dụng để tối ưu thời gian truy vấn ảnh.
- Đồng thời, việc thực hiện lưu trữ ảnh bằng file XML cũng là một cách tối ưu trong việc xây dựng hệ thống truy vấn. Trước khi thực hiện lưu trữ dữ liệu dưới dạng XML, em đã thử lưu trữ chúng bằng file TXT hoặc BIN và riêng biệt (giữa đặc trưng và thông tin ảnh). Khi thực hiện tải thông tin về các ảnh (tên, nhãn) và các đặc trưng tương ứng rất lâu, và thời gian thực hiện truy vấn cũng bị kéo theo.

VIII. Hướng dẫn sử dụng

- Chương trình được xây dựng theo cấu trúc thư mục (bắt buộc) như sau:
 - o <Folder> 21127090

- <Folder> Data: Đây là thư mục lưu trữ tập dữ liệu ảnh (gồm **CD_data** và **TMBuD-main**). Do dữ liệu quá lớn, nên trong bài nộp sẽ không đính kèm vào, thay vào đó sẽ là tập tin TXT chứa link để tải thư mục.
- <Folder> Database: Đây là thư mục lưu trữ cơ sở dữ liệu của hai tập dữ liệu **CD** và **TMBuD**. Tuy nhiên, do kích thước thư mục quá lớn, do đó, trong thư mục này sẽ có một file TXT chứa link để tải dữ liệu. (Lưu ý, riêng đặc trưng Histogram và Correlogram quá lớn, nó sẽ được zip lại để giảm kích thước, nên cần thực hiện unzip trước khi thực hiện truy vấn).
- <Folder> Source: Đây là thư mục chứa mã nguồn chính của chương trình (chứa các file **.cpp** và **.h**).
- <Folder> Query: Đây là thư mục chứa thông tin về ảnh muốn truy vấn.
- <Folder> Exe: Chứa file **.exe** của chương trình.
- <Folder> Report: Chứa báo cáo trình bày nội dung của chương trình (docs và pdf) và ppt.
- <Folder> Demo: Chứa link video demo các chức năng của đồ án.
- Chương trình được chạy trên CMD bằng các lệnh command line. Đầu vào chương trình nhận 5 tham số, gồm:
 - Tên chương trình thực thi (21127090.exe).
 - Đường dẫn đến file chứa thông tin truy vấn (file **.txt**).
 - Tập dữ liệu sử dụng để truy vấn (CD hoặc TMBuD).
 - Loại đặc trưng sử dụng để truy vấn (SIFT, ORB, Histogram, Correlogram).
 - Số lượng ảnh của kết quả truy vấn trả về (3, 5, 11, 21).
 - Cờ đánh dấu rằng sẽ hiển thị kết quả ảnh truy vấn hay không (true hoặc false).
- Ví dụ: *Ta muốn truy vấn các ảnh trong tập tin **test_CD.txt** với bộ dữ liệu **CD**, truy vấn bằng đặc trưng **SIFT** với số lượng ảnh truy vấn là **3** và hiển thị ảnh kết quả truy vấn, ta thực hiện lệnh sau:*

21127090.exe ../../Query/test_CD.txt CD SIFT 11 true

C. TÀI LIỆU THAM KHẢO

[1] Tài liệu môn học – Slide bài giảng

[2] <https://nlp.stanford.edu/IR-book/html/htmledition/evaluation-of-ranked-retrieval-results-1.html>

[3] <https://docs.opencv.org/4.x/index.html>