

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

Báo cáo đồ án cá nhân
Truy vấn ảnh dựa vào nội dung thị giác



Truy vấn thông tin thị giác

Sinh viên thực hiện:

MSSV: 21127730

Họ tên: Hoàng Lê Cát Thanh

Email: hlcthanh21@clc.fitus.edu.vn

SĐT: 0842472242

Giảng viên hướng dẫn:

Võ Hoài Việt

Thành phố Hồ Chí Minh, Ngày 23 tháng 7 năm 2024

Mục lục

| | | |
|----------|--|-----------|
| 1 | Tự đánh giá kết quả | 2 |
| 2 | Các giai đoạn thực hiện | 2 |
| 2.1 | Giai đoạn chuẩn bị cơ sở dữ liệu | 2 |
| 2.1.1 | Trích xuất đặc trưng toàn cục | 2 |
| 2.1.2 | Trích xuất đặc trưng cục bộ | 4 |
| 2.1.3 | Tạo codebook cho các đặc trưng cục bộ | 6 |
| 2.1.4 | Khởi tạo và lưu các đặc trưng | 6 |
| 2.2 | Giai đoạn truy vấn | 7 |
| 3 | Kết quả ban đầu và so sánh | 9 |
| 3.1 | Kết quả ban đầu | 9 |
| 3.1.1 | Color histogram | 9 |
| 3.1.2 | Color correlogram | 11 |
| 3.1.3 | SIFT Features | 12 |
| 3.1.4 | ORB Features | 13 |
| 3.2 | So sánh kết quả | 15 |
| 4 | Cải thiện độ chính xác của hệ thống | 16 |
| 4.1 | Sử dụng kết hợp đặc trưng cục bộ và toàn cục | 16 |
| 4.2 | Kết quả thực nghiệm | 17 |
| 4.3 | Đánh giá kết quả | 19 |
| 5 | Tài liệu tham khảo | 21 |

1 Tự đánh giá kết quả

| STT | Yêu cầu | Ghi chú | Tỷ lệ hoàn thành |
|-----|---|---|------------------|
| 1 | <ul style="list-style-type: none"> 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. | <ul style="list-style-type: none"> Các tấm ảnh được hiển thị trên màn hình giảm dần theo mức độ giống và giá trị so khớp tương ứng. So sánh kết quả khi sử dụng các đặc trưng khác nhau. Sử dụng độ đo MAP để đánh giá độ chính xác kết quả truy vấn lần lượt theo số lượng kết quả trả về là 3, 5, 11 và 21. | 100% |
| 2 | <ul style="list-style-type: none"> 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. | <ul style="list-style-type: none"> Sử dụng các đặc trưng liên quan đến hình dáng, biên cạnh, vân (texture). Có thể kết hợp các đặc trưng để tăng độ chính xác của hệ thống. Hoặc sử dụng các phương pháp để nâng cao chất lượng ảnh trước khi rút trích đặc trưng. Sử dụng độ đo MAP để đánh giá độ chính xác kết quả truy vấn lần lượt theo số lượng kết quả trả về là 3, 5, 11 và 21. | 50% |
| 3 | Tìm hiểu và cài đặt cải tiến tốc độ tìm kiếm ảnh trong CSDL | Ghi rõ thời gian truy vấn trên giao diện của chương trình | 0% |

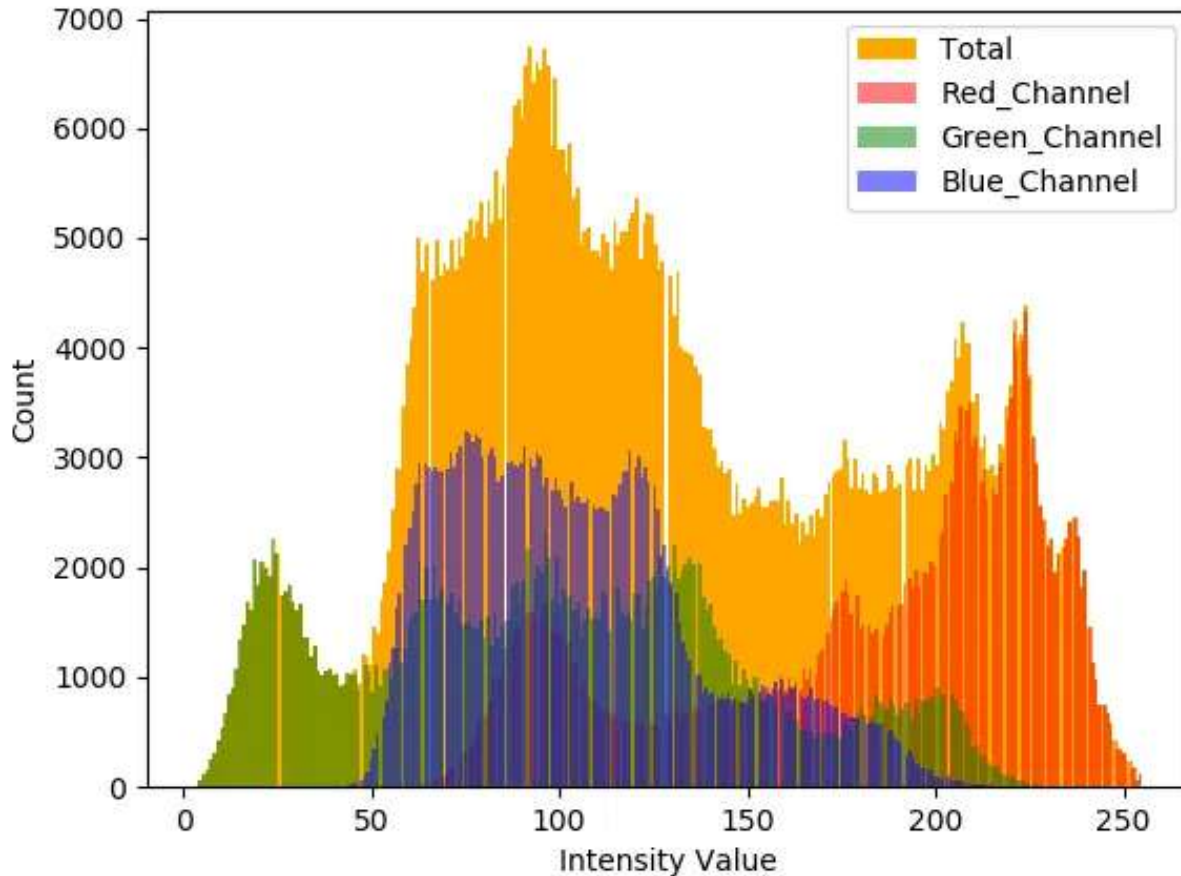
2 Các giai đoạn thực hiện

2.1 Giai đoạn chuẩn bị cơ sở dữ liệu

2.1.1 Trích xuất đặc trưng toàn cục

Đặc trưng toàn cục của ảnh mô tả các thuộc tính chung của toàn bộ ảnh, thay vì tập trung vào các chi tiết cục bộ như điểm ảnh riêng lẻ hoặc vùng nhỏ. Những đặc trưng này cung cấp thông tin tóm tắt về nội dung ảnh, chẳng hạn như phân bố màu sắc, độ sáng tổng thể và độ tương phản. Hai đặc trưng toàn cục được sử dụng để làm cơ sở cho quá trình truy vấn là *Color Histogram* và *Color*

Correlogram. *Color Histogram* giúp nắm được phân bố màu sắc tổng thể một cách dễ dàng và hiệu quả, trong khi *Color Correlogram* cung cấp thông tin chi tiết hơn về mối quan hệ không gian giữa các màu sắc, dẫn đến việc khả năng truy xuất chính xác hơn cho các ảnh có sắp xếp màu sắc tương tự.



Hình 1: Biểu diễn của Color Histogram

Color Histogram Color Histogram cho thấy được một biểu đồ thống kê thể hiện sự phân bố màu sắc trong một hình ảnh.

Các bước thực hiện trích xuất đặc trưng Color Histogram:

- **Phân tách kênh màu:** ảnh gốc được phân tách thành các kênh màu riêng biệt (đỏ, lục, lam), giúp phân tích độc lập phân bố màu sắc trong từng kênh.
- **Tính toán histogram:** đối với mỗi kênh màu, thực hiện tính toán histogram dựa vào số lần xuất hiện của mỗi giá trị cường độ màu trong một phạm vi nhất định.
- **Nối các kênh màu lại:** sau khi tính toán histogram cho từng kênh, các histogram này được nối lại với nhau thành một histogram duy nhất và được dàn phẳng về dưới dạng vector đặc trưng.

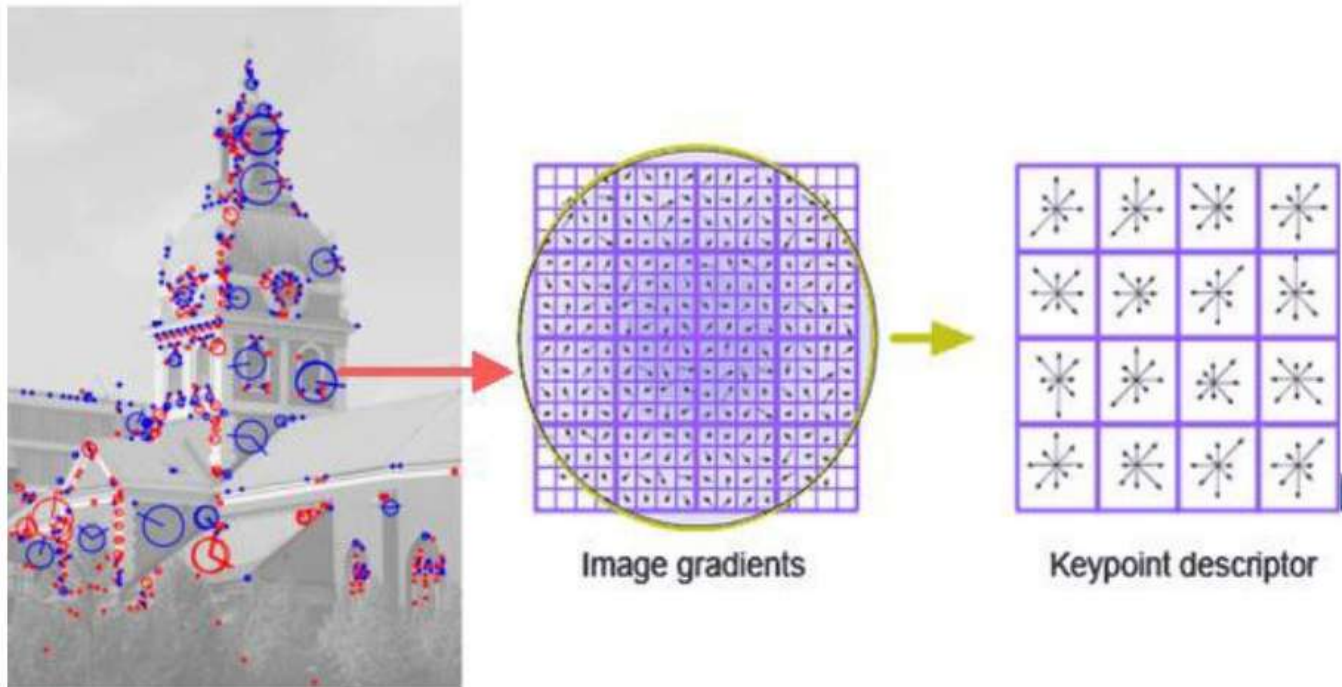
Color Correlogram Tương tự như Color Histogram, Color Correlogram cho thấy biểu đồ thống kê về mối quan hệ không gian giữa các cặp pixel có màu sắc khác nhau trong một hình ảnh. Color Correlogram đo lường mức độ phụ thuộc lẫn nhau giữa các kênh màu, cho biết mức độ đồng biến hoặc ngẫu nhiên của màu sắc trong ảnh.

Các bước thực hiện trích xuất đặc trưng Color Correlogram cho ảnh:

- **Color Quantization:** Đây là bước quan trọng trong quá trình trích xuất đặc trưng, làm giảm số lượng màu khác biệt trong ảnh bằng cách ánh xạ các màu tương tự về cùng một giá trị đại diện.
- **Tính toán Correlogram:**
 - **Phân chia các Bin Allocation:** Đối với mỗi pixel, giá trị màu đã được giảm sẽ được sử dụng làm chỉ mục để gán vào một bin cụ thể trong không gian 3D được xác định trước.
 - **Neighbor Analysis:** Lặp qua một vùng lân cận được chỉ định xung quanh mỗi pixel. Đối với mỗi neighbor, thực hiện kiểm tra xem giá trị màu đã giảm có nằm trong cùng một bin với pixel trung tâm hay không.
 - **Đếm Co-Occurrences:** Nếu màu của neighbor thuộc cùng một bin, biến đếm cho mức khoảng cách đó được tăng lên, hỗ trợ nắm bắt tần suất xuất hiện của các màu tương tự trong một khoảng cách không gian nhất định.
 - **Chuẩn hóa:** Cuối cùng, correlogram được chuẩn hóa bằng cách chia mỗi giá trị bin cho tổng số pixel trong ảnh, đảm bảo correlogram được trích xuất độc lập với kích thước ảnh.

2.1.2 Trích xuất đặc trưng cục bộ

Đặc trưng cục bộ là những mô tả về các phần nhỏ, riêng biệt trong một hình ảnh, được sử dụng rộng rãi trong các ứng dụng thị giác máy tính như nhận dạng đối tượng, ghép ảnh, và theo dõi chuyển động. Trong đó, hai đặc trưng cục bộ phổ biến là SIFT (Scale-Invariant Feature Transform) và ORB (Oriented FAST and Rotated BRIEF).



Hình 2: Minh họa cho bộ mô tả của đặc trưng cục bộ

Đặc trưng SIFT Các đặc trưng SIFT là các đặc trưng bất biến khi ảnh thay đổi tỷ lệ, gồm bộ mô tả (descriptors) 128 chiều cho mỗi điểm đặc trưng trong ảnh (keypoint), mô tả các đặc điểm dựa trên hướng, độ lớn và vị trí, được tính toán bằng cách lấy gradient của ảnh xung quanh điểm đặc trưng và tạo ra một histogram hướng. Histogram này có khả năng bất biến đối với các thay đổi về tỷ lệ, độ xoay và độ sáng của ảnh.

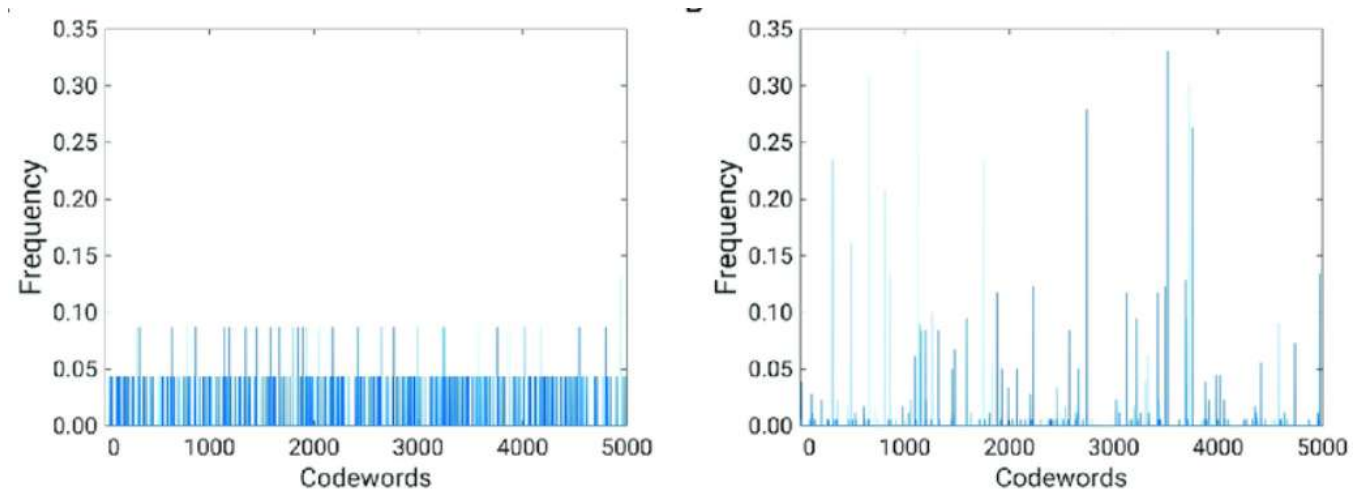
Đặc trưng ORB Đặc trưng này sử dụng các điểm FAST (Features from Accelerated Segment Test) để phát hiện điểm đặc trưng và mô tả BRIEF (Binary Robust Independent Elementary Features) để xây dựng bộ mô tả 31 chiều. ORB nhanh hơn SIFT và có khả năng chống lại nhiễu và thay đổi ánh sáng.

Để thực hiện trích xuất các đặc trưng cục bộ, thư viện OpenCV được sử dụng để gọi đến các hàm xử lý dùng để trích xuất các đặc trưng này từ ảnh. Các bước thực hiện trích xuất đặc trưng bao gồm:

- **Chuyển đổi sang ảnh xám:** Nếu ảnh đầu vào là ảnh màu, hệ thống sẽ chuyển đổi sang ảnh xám vì SIFT thường hoạt động tốt hơn với ảnh xám. Bên cạnh đó, ảnh xám nắm bắt sự thay đổi cường độ mà không có thông tin màu sắc, giúp dễ dàng phát hiện các điểm đặc trưng.
- **Tạo đối tượng và phát hiện điểm đặc trưng:** hệ thống thực hiện khởi tạo đối tượng theo đặc trưng đang được trích xuất và sử dụng nó để xác định các điểm đặc trưng có trong ảnh. Từ các điểm đặc trưng này, một bộ mô tả cho mỗi điểm sẽ được tính toán qua các hàm có sẵn trong đối tượng như *detectAndCompute*.

2.1.3 Tạo codebook cho các đặc trưng cục bộ

Từ các bộ mô tả điểm đặc trưng được trích xuất từ ảnh trong bộ dữ liệu, một bộ codebook được tạo ra bằng cách thực hiện gom nhóm các đặc trưng này. Các codewords sẽ được khởi tạo dựa trên tâm điểm của mỗi nhóm sau khi hoàn tất gom nhóm. Các codewords này là cơ sở để biểu diễn lại các vector trong ảnh dưới dạng một histogram các codewords.



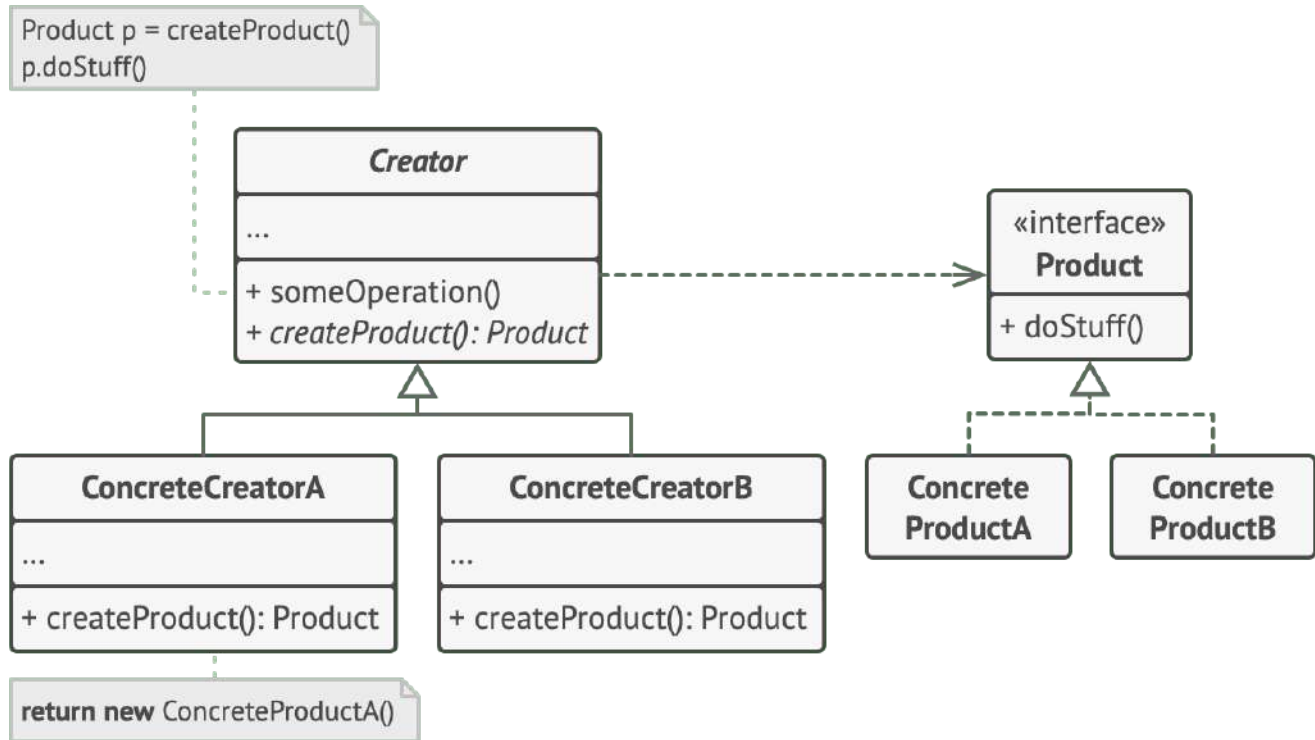
Hình 3: Minh họa cho Histogram của codewords

Để thực hiện, hệ thống sẽ tiến hành xử lý các bước sau:

- **Trích xuất các đặc trưng cần thiết:** đầu tiên, thực hiện trích xuất đặc trưng cục bộ cho tất cả ảnh trong bộ dữ liệu. Các vector mô tả này sau đó được kết hợp lại với nhau và gộp thành một ma trận để thực hiện gom cụm dữ liệu bằng *K-means*.
- **Thực hiện phân cụm:** thuật toán được chọn để thực hiện gom cụm của hệ thống là thuật toán gom cụm K-means, giúp nhóm các điểm dữ liệu tương tự nhau vào một nhóm, tạo thành k từ vựng (codewords) trực quan dựa trên các tâm cụm.
- **Tạo codewords histogram:**
 - **Khởi tạo histogram:** Hệ thống thực hiện khởi tạo một histogram dưới dạng một vector hàng với độ dài bằng số lượng cụm (k) trong bộ từ vựng. Mỗi phần tử trong histogram sẽ đại diện cho số lần xuất hiện của một codeword.
 - **Gán codewords:** Đối với mỗi vector đặc trưng, hệ thống sẽ tìm tâm cụm gần nhất (visual codewords) trong bộ từ vựng bằng cách sử dụng khoảng cách L2 (khoảng cách Euclid) như một thước đo sự tương đồng với chỉ số của tâm gần nhất trong bộ từ vựng tương ứng với một từ vựng trực quan cụ thể. Histogram tại chỉ số đó được tăng lên 1, về cơ bản ghi lại số lượng đặc trưng trong ảnh thuộc về từ vựng trực quan đó.

2.1.4 Khởi tạo và lưu các đặc trưng

- Các đặc trưng được khởi tạo dựa trên lớp Factory, giúp việc khởi tạo các đặc trưng trở nên dễ dàng và linh hoạt hơn. Quá trình trích xuất đặc trưng được thực hiện một cách tự động



Hình 4: Design Pattern được sử dụng: Factory

và hiệu quả, giảm thiểu sai sót và tăng tốc độ xử lý.

- Các đặc trưng sau khi trích xuất sẽ được lưu trữ trong các tệp XML. Mỗi tệp XML chứa thông tin về đường dẫn đến hình ảnh và đặc trưng tương ứng đã được trích xuất. Cụ thể, tên tệp ảnh và các mô tả đặc trưng được ghi vào tệp dưới dạng các cặp khóa-giá trị trong một cấu trúc map. Khi cần đọc lại tệp, hệ thống sẽ sử dụng cùng cấu trúc cặp khóa-giá trị này để lấy lại tên tệp ảnh và mô tả đặc trưng một cách chính xác và nhanh chóng.
- Dựa vào loại đặc trưng được sử dụng, hệ thống sẽ xác định xem đặc trưng đó là cục bộ hay toàn cục. Quá trình này dựa trên các cài đặt đã được xác định trước trong tệp cấu hình (config file). Tệp cấu hình này không chỉ chỉ định rõ ràng đặc trưng nào là cục bộ và đặc trưng nào là toàn cục, mà còn bao gồm số lượng cụm được sử dụng để phân cụm các đặc trưng cục bộ, cùng với số lượng hình ảnh sẽ được trả về sau khi truy xuất. Nhờ vào việc sử dụng tệp cấu hình, hệ thống có thể hoạt động một cách linh hoạt và hiệu quả, đáp ứng đúng yêu cầu của người sử dụng và đảm bảo chất lượng kết quả truy xuất.

2.2 Giai đoạn truy vấn

- Tùy theo đặc trưng mà người dùng chọn khi chạy chương trình, đặc trưng tương ứng sẽ được trích xuất ra từ ảnh truy vấn và được đưa vào làm đầu vào của hệ thống truy vấn ảnh. Nếu đặc trưng đang được chọn là đặc trưng cục bộ như SIFT hoặc ORB, hệ thống sẽ thực hiện tạo codewords histogram cho ảnh truy vấn với các tâm cụm đại diện cho từng codeword được lấy từ dữ liệu đã được lưu về sau khi phân cụm.

- Sau khi có được vector đặc trưng cho ảnh truy vấn, vector đặc trưng tương ứng cho các ảnh trong bộ dữ liệu cũng được đọc từ file được lưu sau khi trích xuất.
- Để xác định độ giống và độ so khớp của ảnh truy vấn với các ảnh trong bộ dữ liệu, độ đo cosine được sử dụng với giá trị của độ đo càng lớn thì hai ảnh càng gần giống nhau.
- Sau khi đã trích xuất toàn bộ các đặc trưng cần thiết, hệ thống sẽ bắt đầu thực hiện quá trình truy vấn. Với mỗi ảnh trong bộ dữ liệu, độ đo tương đồng cosine được tính trên vector đặc trưng của ảnh truy vấn và vector đặc trưng của ảnh trong bộ dữ liệu đang được xét đến. Khi đã duyệt qua hết các ảnh trong bộ dữ liệu hệ thống thực hiện sắp xếp các giá trị của độ đo theo thứ tự giảm dần với các ảnh ở đầu được xem là các ảnh gần nhất với ảnh truy vấn.
- Để đánh giá kết quả truy vấn trả về, độ đo MAP được sử dụng như một thang đo đánh giá kết quả.

$$\text{MAP}(Q) = \frac{1}{|Q|} \sum_{j=1}^{|Q|} \frac{1}{m_j} \sum_{k=1}^{m_j} \text{Precision}(R_{jk})$$

Hình 5: Độ đo MAP

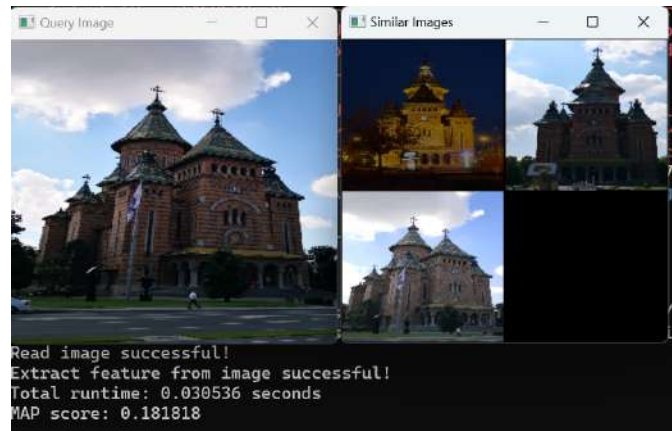
- Độ đo **MAP** (Mean Average Precision), là một thước đo hiệu suất phổ biến để đánh giá chất lượng kết quả trả về của hệ thống truy vấn, độ đo này đo lường mức độ chính xác và tính toàn diện của hệ thống trong việc xếp hạng các kết quả theo mức độ liên quan đến ảnh truy vấn của người dùng. Độ đo này được tính toán qua các bước:
 - **Tính toán độ chính xác (Precision) tại mỗi vị trí k**: là tỷ lệ số lượng ảnh có liên quan trong top k kết quả trả về chia cho tổng số ảnh trong top k.
 - **Tính toán đường cong Precision-Recall**: với Precision được tính trên trục y và Recall là tỷ lệ số lượng ảnh có liên quan được tìm thấy chia cho tổng số ảnh có liên quan trong tập dữ liệu. Các điểm (Recall, Precision) được tính toán tại mỗi ngưỡng IoU (Intersection over Union) khác nhau sẽ được nối lại tạo thành đường cong Precision-Recall.
 - **Tính toán diện tích dưới đường cong Precision-Recall (AUC)**: giá trị này thể hiện mức độ tổng thể mà hệ thống xếp hạng các ảnh có liên quan cao hơn các ảnh không liên quan.
 - **Tính giá trị MAP**: là giá trị trung bình của AUC trên tất cả các lớp truy vấn hoặc ngưỡng IoU.
- Để có thể tính toán được độ đo MAP, dữ liệu cần được đánh nhãn trước (ground truth) để có thể kiểm tra đâu là ảnh có liên quan đến ảnh truy vấn đưa vào. Đối với tập dữ liệu TMBuD, nhãn của ảnh chính là tên ảnh được lưu trong file csv đi kèm với bộ dữ liệu. Tuy nhiên bộ dữ liệu CD lại không có sẵn nhãn, do đó, nhãn cho các ảnh được đánh thủ công và lưu vào một file csv để sử dụng khi đánh giá kết quả trả về.

- Quá trình tính toán giá trị độ đo MAP để đánh giá kết quả trả về được thực hiện qua các bước:
 - **Đọc ground truth:** tùy theo bộ dữ liệu đang được sử dụng mà file csv tương ứng sẽ được load lên để lấy ra tên ảnh và nhãn của ảnh.
 - **Tính toán độ chính xác (Precision) tại mỗi vị trí k:** hàm tính toán độ chính xác lấy danh sách tên ảnh trả về sau truy xuất và một tập tên ảnh có liên quan, kiểm tra từng ảnh trả về xem có nằm trong tập các ảnh có liên quan không. Nếu có, thực hiện tăng biến đếm liên quan cho ảnh đó, độ chính xác tại thứ hạng hiện tại (k) được tính bằng cách chia số lượng ảnh truy xuất chính xác cho tổng số ảnh có liên quan.
 - **Tính toán AP:** Độ chính xác trung bình (AP) được tính bằng cách chia tổng độ chính xác cho tổng số giá trị độ chính xác trong kết quả trả về.
 - **Tính toán MAP:** MAP khi này được xác định bằng giá trị trung bình của tất cả các AP được tính.

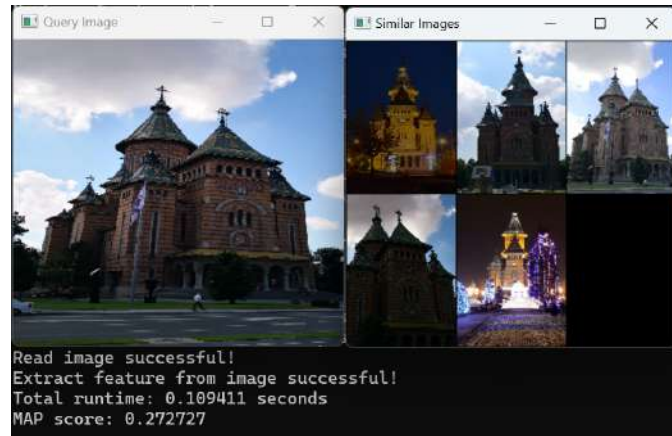
3 Kết quả ban đầu và so sánh

3.1 Kết quả ban đầu

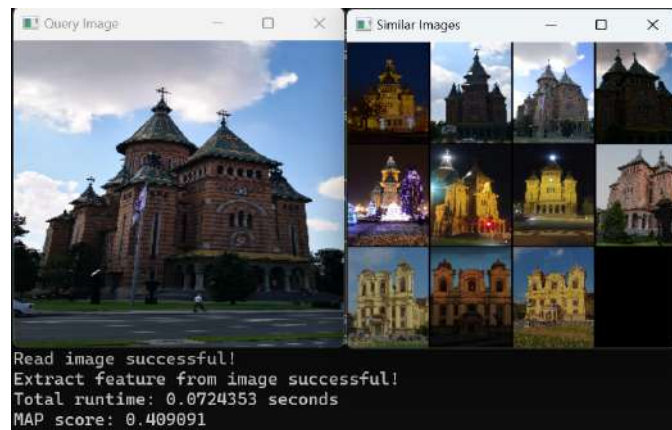
3.1.1 Color histogram



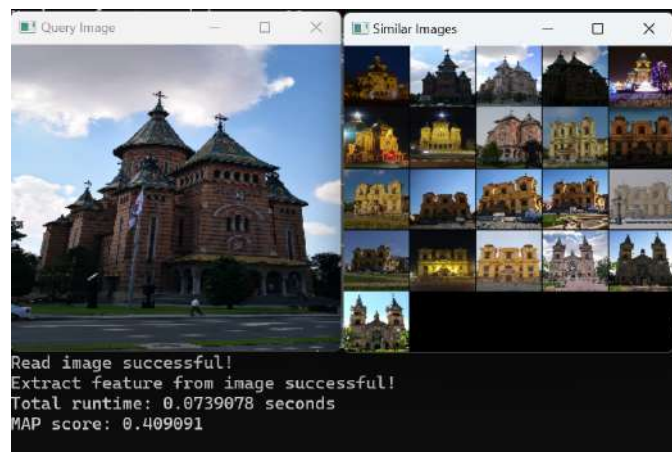
Hình 6: Trả về 3 ảnh khi sử dụng đặc trưng color histogram



Hình 7: Trả về 5 ảnh khi sử dụng đặc trưng color histogram

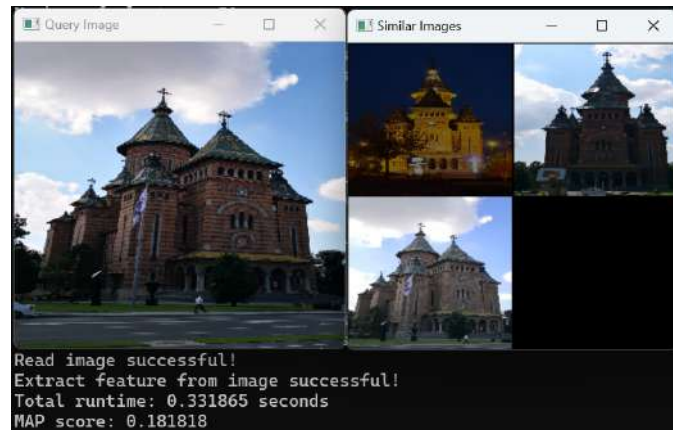


Hình 8: Trả về 11 ảnh khi sử dụng đặc trưng color histogram

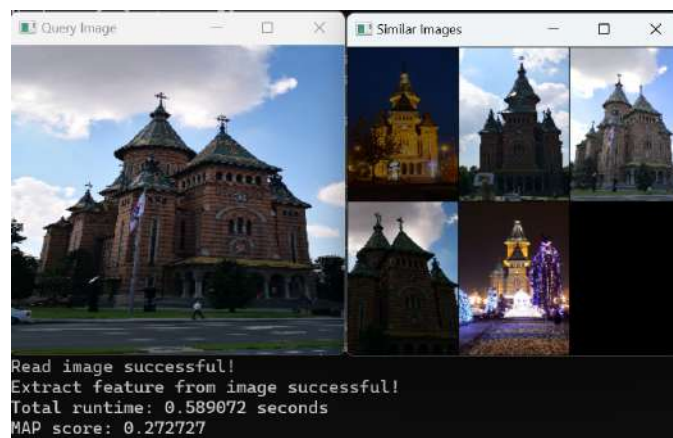


Hình 9: Trả về 21 ảnh khi sử dụng đặc trưng color histogram

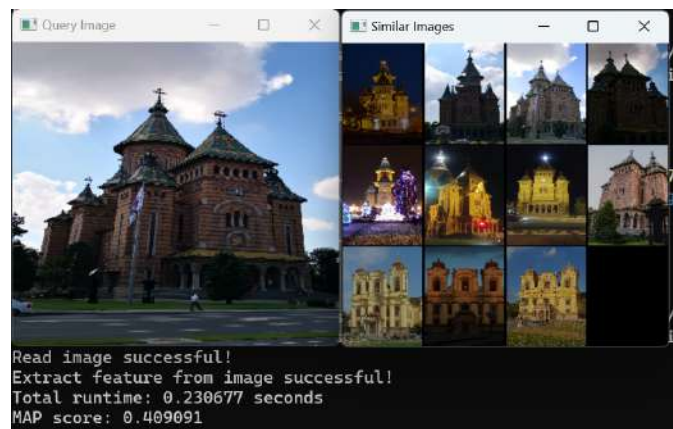
3.1.2 Color correlogram



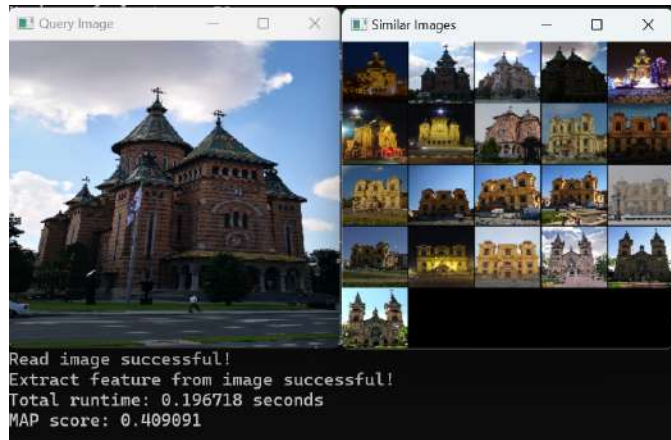
Hình 10: Trả về 3 ảnh khi sử dụng đặc trưng color correlogram



Hình 11: Trả về 5 ảnh khi sử dụng đặc trưng color correlogram

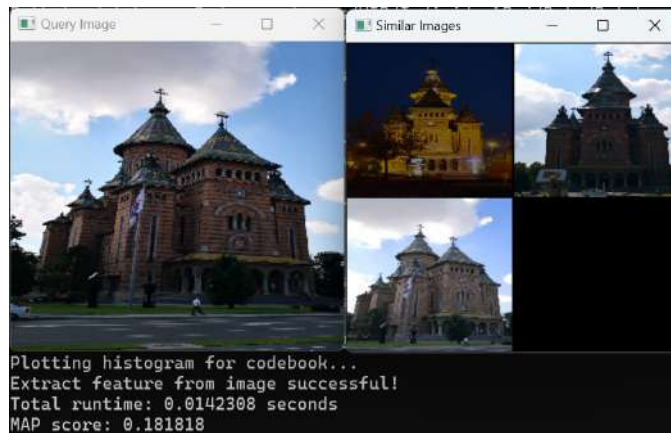


Hình 12: Trả về 11 ảnh khi sử dụng đặc trưng color correlogram

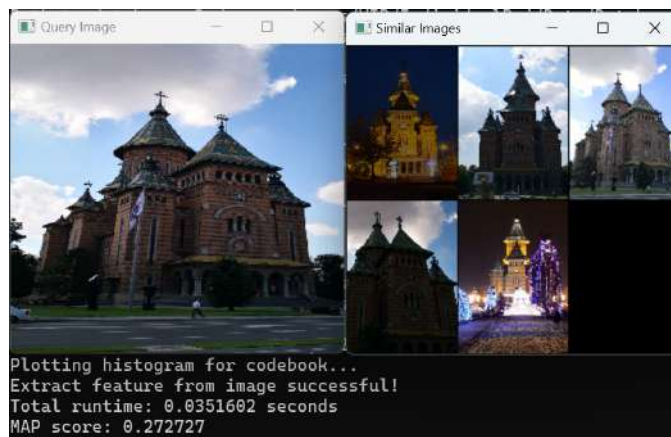


Hình 13: Trả về 21 ảnh khi sử dụng đặc trưng color correlogram

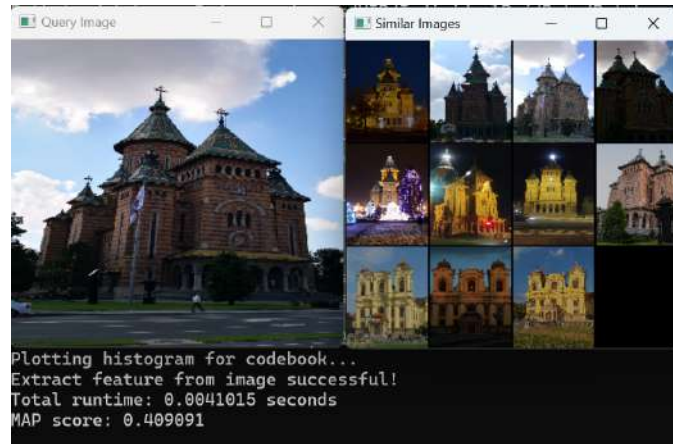
3.1.3 SIFT Features



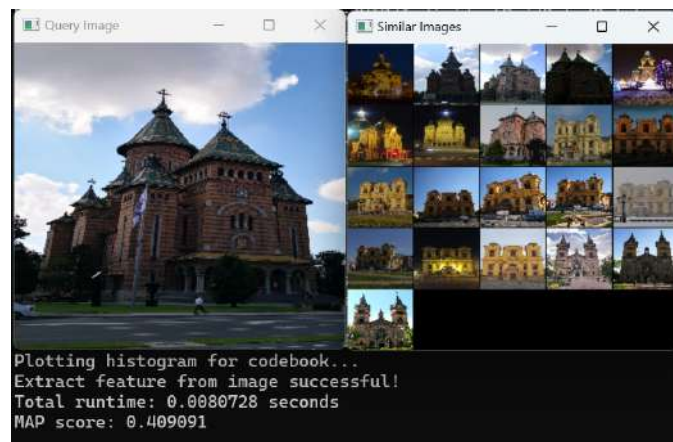
Hình 14: Trả về 3 ảnh khi sử dụng đặc trưng SIFT



Hình 15: Trả về 5 ảnh khi sử dụng đặc trưng SIFT

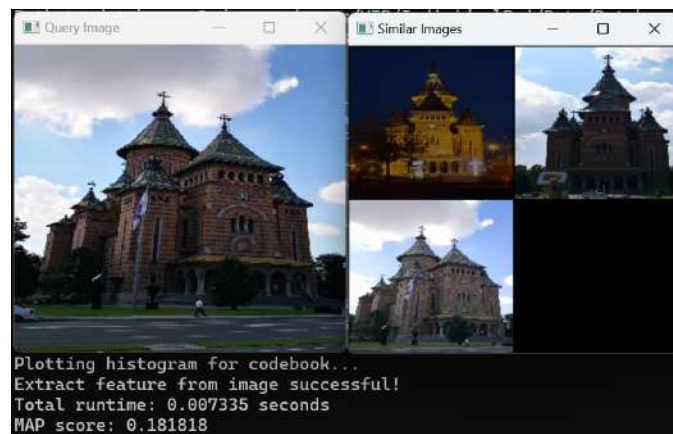


Hình 16: Trả về 11 ảnh khi sử dụng đặc trưng SIFT

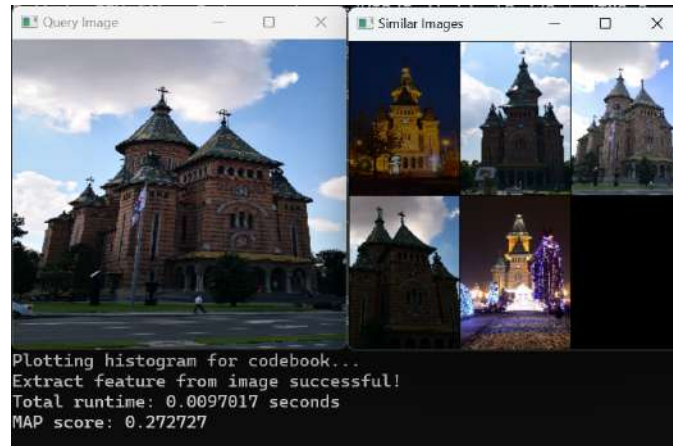


Hình 17: Trả về 21 ảnh khi sử dụng đặc trưng SIFT

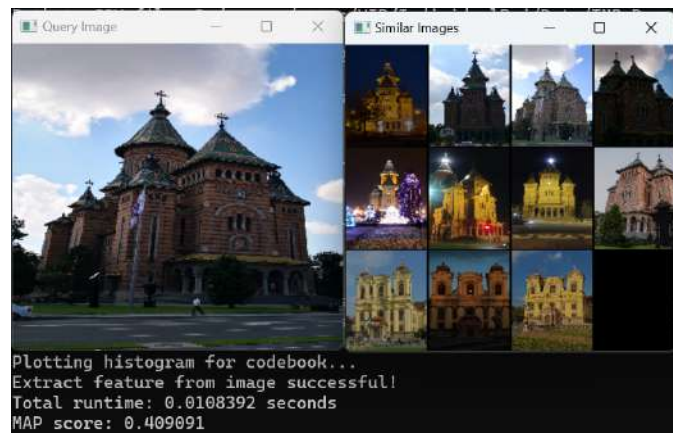
3.1.4 ORB Features



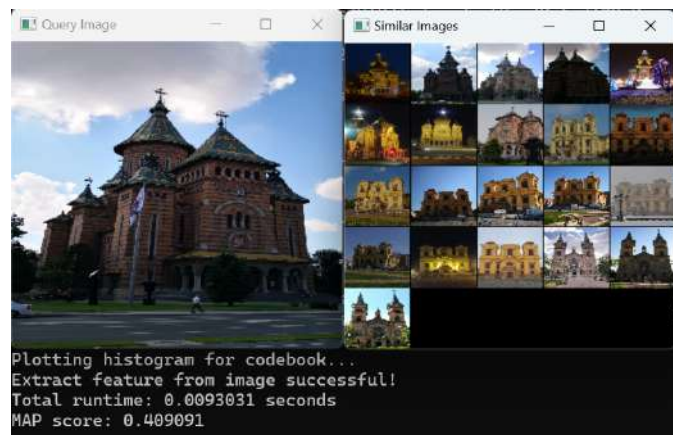
Hình 18: Trả về 3 ảnh khi sử dụng đặc trưng ORB



Hình 19: Trả về 5 ảnh khi sử dụng đặc trưng ORB



Hình 20: Trả về 11 ảnh khi sử dụng đặc trưng ORB



Hình 21: Trả về 21 ảnh khi sử dụng đặc trưng ORB

3.2 So sánh kết quả

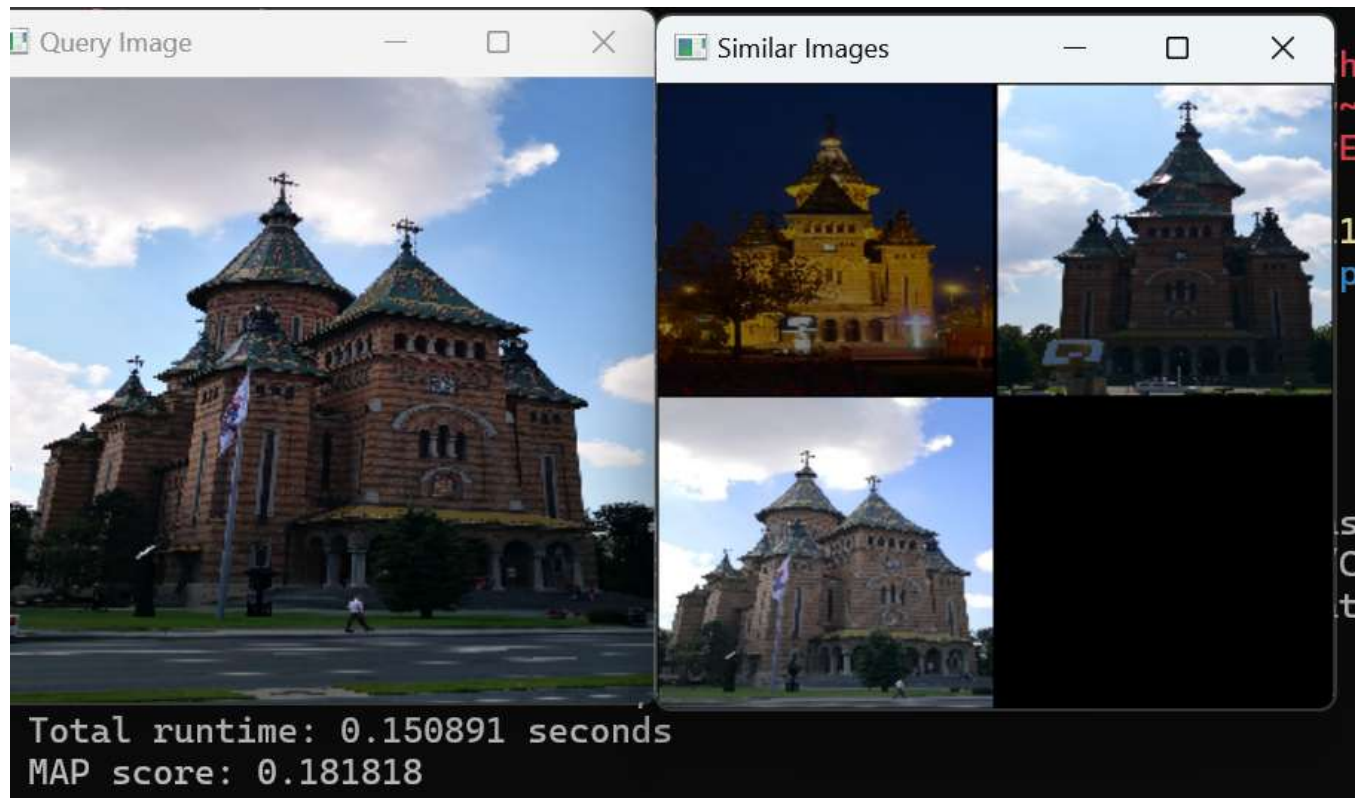
- Từ các kết quả thu được khi thực nghiệm trên các đặc trưng khác nhau với nhiều kết quả trả về khác nhau, có thể thấy rằng kết quả truy vấn đối với các đặc trưng đều tương đồng. Các ảnh tương đồng nhất luôn được xếp đầu tiên và giảm dần về sau.
- Qua thử nghiệm và phân tích kết quả truy vấn trên nhiều trường hợp khác nhau đối với bốn đặc trưng: Color Histogram, Color Correlogram, SIFT và ORB. Có thể thấy rằng việc chỉ sử dụng một đặc trưng toàn cục hoặc cục bộ duy nhất cho tác vụ truy vấn ảnh không thể mang lại kết quả tối ưu nhất. Một phần nguyên nhân là do trong bộ dữ liệu có nhiều ảnh tương tự nhau, cùng với việc các ảnh được thu thập trong nhiều điều kiện ánh sáng và góc chụp khác nhau, dẫn đến các vector đặc trưng khi rút trích có tính tương đồng cao. Do đó, kết quả truy vấn trả về cho các đặc trưng khác nhau đều tương tự nhau, khiến cho giá trị độ đo MAP (Mean Average Precision) khi trả về đối với tất cả các đặc trưng là như nhau.
- Đặc biệt, đối với các đặc trưng SIFT và ORB, những đặc trưng này được trích xuất từ ảnh xám thay vì ảnh màu, làm mất đi một số thông tin cần thiết về màu sắc. Điều này gây ra tình trạng các ảnh khác nhau nhưng lại được nhận định là giống nhau do các đặc trưng cục bộ này có tính tương tự cao.
- Ngược lại, các đặc trưng toàn cục như Color Histogram và Color Correlogram lại sử dụng thông tin màu sắc của ảnh để tạo ra vector đặc trưng. Color Histogram biểu diễn sự phân bố màu sắc trong toàn bộ ảnh bằng cách đếm số lượng pixel tương ứng với từng giá trị màu sắc. Điều này giúp giữ lại thông tin về màu sắc, nhưng lại bỏ qua các thông tin về vị trí không gian của các màu sắc trong ảnh. Mặc dù Color Histogram có thể nhận diện được các ảnh có màu sắc tương tự nhau, nhưng nó không phân biệt được các ảnh có bố cục màu sắc khác nhau.
- Color Correlogram cải tiến hơn bằng cách không chỉ xem xét tần suất xuất hiện của các cặp màu ở các khoảng cách khác nhau trong ảnh, mà còn xem xét mối quan hệ không gian giữa các pixel màu sắc. Điều này giúp Color Correlogram giữ lại nhiều thông tin không gian hơn so với Color Histogram, nhưng việc tính toán lại phức tạp và tốn kém hơn về mặt thời gian và tài nguyên.
- Tuy nhiên, khi xét về thời gian truy vấn, các đặc trưng cục bộ có thời gian truy vấn nhanh hơn so với các đặc trưng toàn cục. Trong quá trình rút trích đặc trưng, việc trích xuất các đặc trưng cục bộ đòi hỏi nhiều thời gian hơn so với việc tính toán các histogram. Đặc trưng cục bộ cần phải trải qua giai đoạn gom nhóm để tạo codebook và sau đó tạo histogram cho các codewords trên ảnh. Điều này làm cho quá trình rút trích đặc trưng trở nên phức tạp và tốn nhiều thời gian hơn.
- Ngoài ra, đối với tập dữ liệu lớn như TMBuD, việc tăng số lượng ảnh trả về có thể làm tăng giá trị của độ đo MAP bởi hệ thống có thể trả về được nhiều ảnh tương đồng tìm được hơn. Ngược lại, với tập dữ liệu nhỏ như tập CD, việc tăng số lượng kết quả trả về có thể sẽ làm giảm đi giá trị của MAP bởi không có nhiều ảnh tương đồng để trả về khi thực hiện truy vấn.

4 Cải thiện độ chính xác của hệ thống

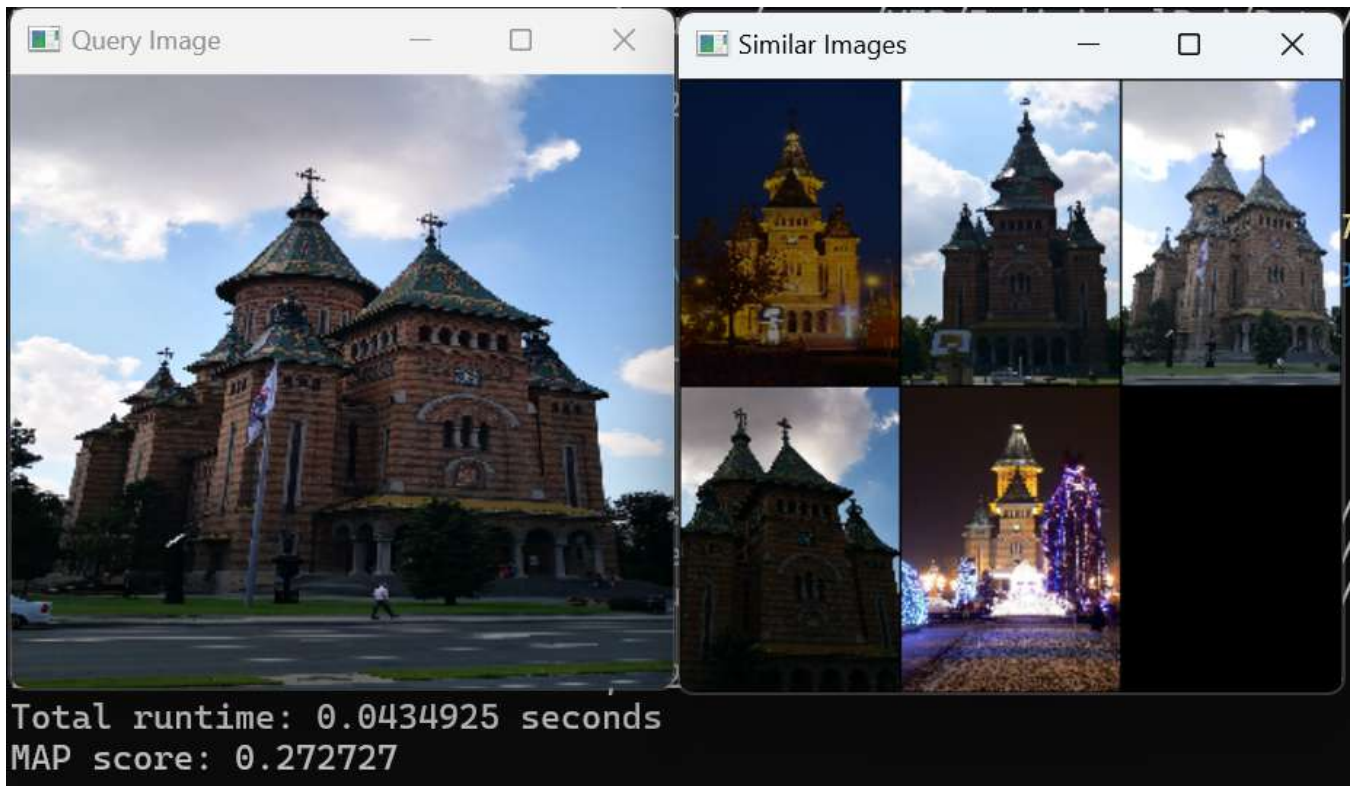
4.1 Sử dụng kết hợp đặc trưng cục bộ và toàn cục

- Thông qua phân tích đánh giá kết quả thực nghiệm đối với từng đặc trưng khác nhau, có thể thấy rằng việc chỉ sử dụng một đặc trưng cục bộ hoặc toàn cục để thực hiện truy vấn không đem lại kết quả với độ chính xác như mong đợi. Do đó, một phương pháp để tăng độ chính xác của hệ thống được nghiên cứu thử nghiệm là sử dụng kết hợp đặc trưng cục bộ *SIFT* và đặc trưng toàn cục *Color Histogram*.
- Đối với đặc trưng cục bộ *SIFT* (*Scale-Invariant Feature Transform*), các đặc trưng này được trích xuất từ ảnh xám thay vì ảnh màu, làm mất đi một số thông tin cần thiết về màu sắc. *SIFT* tập trung vào việc nhận diện các điểm đặc biệt và mô tả chúng một cách không phụ thuộc vào sự thay đổi về ánh sáng, góc chụp hay phóng to/thu nhỏ của ảnh. Điều này giúp *SIFT* rất mạnh mẽ trong việc phát hiện các đặc điểm chính của ảnh, nhưng lại bỏ qua các thông tin về màu sắc, khiến cho các ảnh khác nhau nhưng có cấu trúc hình học tương tự có thể bị nhận diện là giống nhau.
- Ngược lại, đặc trưng toàn cục như *Color Histogram* lại sử dụng thông tin màu sắc của ảnh để tạo ra vector đặc trưng. *Color Histogram* biểu diễn sự phân bố màu sắc trong toàn bộ ảnh bằng cách đếm số lượng pixel tương ứng với từng giá trị màu sắc. Điều này giúp giữ lại thông tin về màu sắc, nhưng lại bỏ qua các thông tin về vị trí không gian của các màu sắc trong ảnh. Mặc dù *Color Histogram* có thể nhận diện được các ảnh có màu sắc tương tự nhau, nhưng nó không phân biệt được các ảnh có bố cục màu sắc khác nhau.
- Việc kết hợp đặc trưng cục bộ *SIFT* và đặc trưng toàn cục *Color Histogram* được kỳ vọng sẽ mang lại một phương pháp cân bằng giữa thông tin cấu trúc và màu sắc của ảnh. Đặc trưng *SIFT* cung cấp khả năng nhận diện các đặc điểm hình học mạnh mẽ, trong khi *Color Histogram* bổ sung thêm thông tin về màu sắc, giúp phân biệt các ảnh có màu sắc khác nhau nhưng có cấu trúc tương tự.
- Quá trình thực hiện truy vấn ảnh sử dụng kết hợp hai đặc trưng được thực hiện qua các bước:
 - Lần lượt thực hiện trích xuất đặc trưng *Color Histogram* cũng như *Codeword Histogram* cho ảnh truy vấn.
 - Thực hiện tính toán độ đo tương đồng cosine giữa *Color Histogram* của ảnh truy vấn với các ảnh trong cơ sở dữ liệu và tương tự đối với *Codeword Histogram* của ảnh.
 - Giá trị tương đồng cuối cùng cho ảnh truy vấn với mỗi ảnh trong cơ sở dữ liệu là giá trị trung bình cho đặc trưng *Color Histogram* và *SIFT*. Cả hai đặc trưng này đều được đánh trọng số tương đồng nhau.
 - Thực hiện sắp xếp theo thứ tự độ tương đồng giảm dần và trả về một danh sách các tên ảnh tùy theo số lượng ảnh trả về đã được xác định trước.

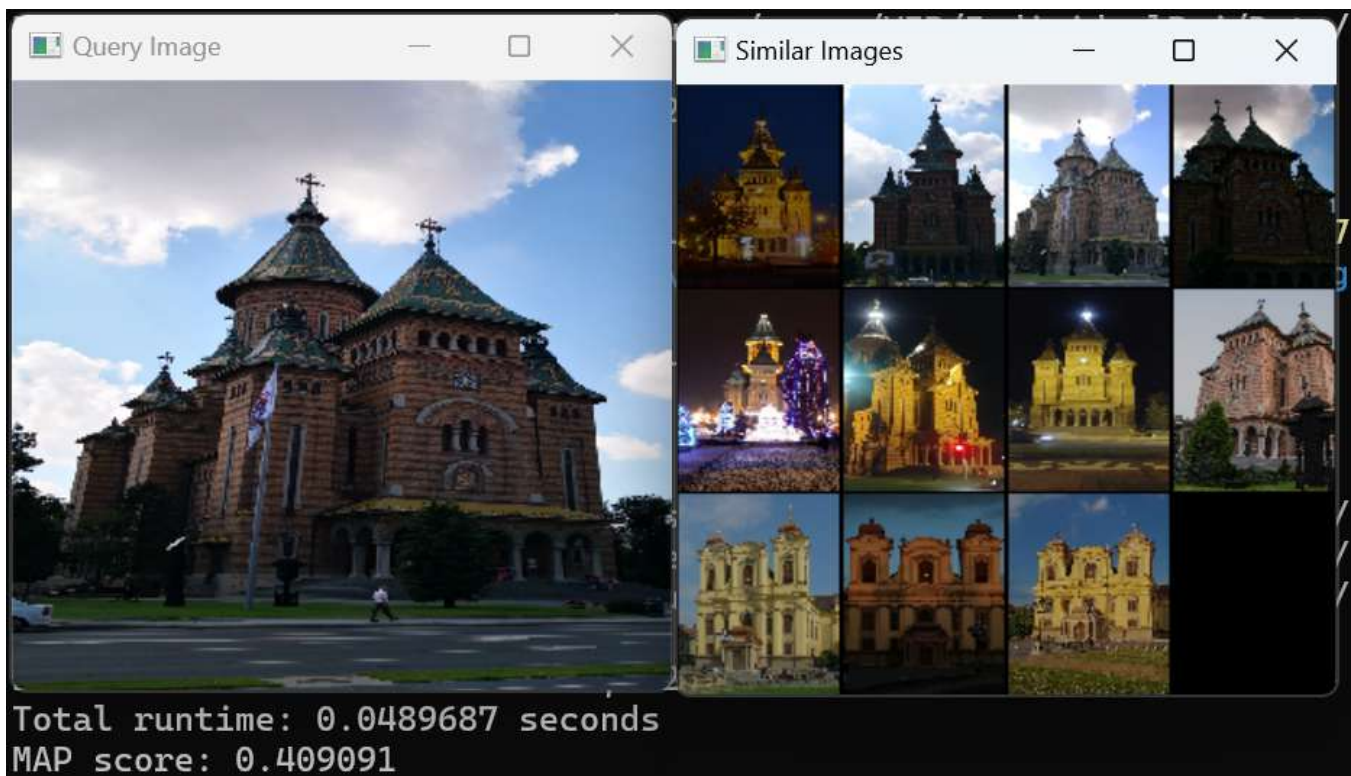
4.2 Kết quả thực nghiệm



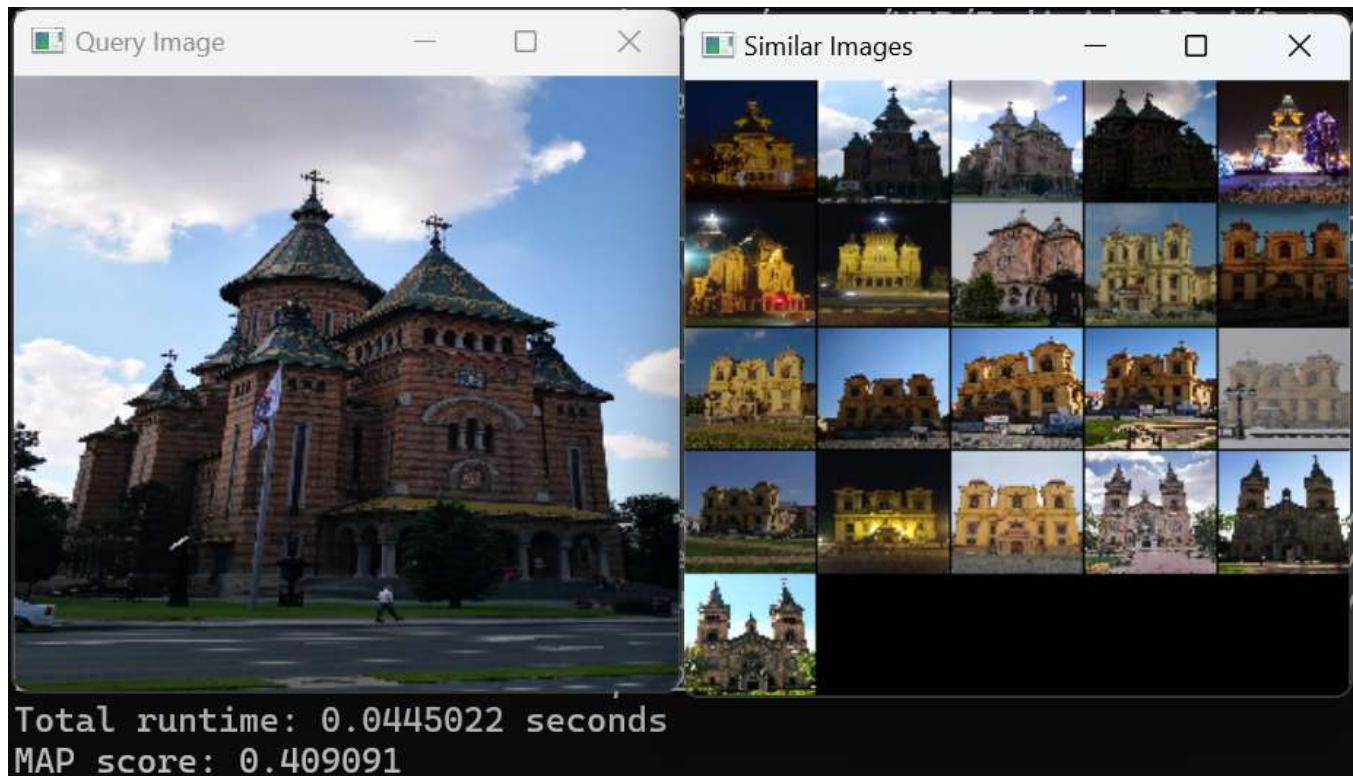
Hình 22: Trả về 3 ảnh



Hình 23: Trả về 5 ảnh



Hình 24: Trả về 11 ảnh



Hình 25: Trả về 21 ảnh

4.3 Đánh giá kết quả

- Dựa vào kết quả truy vấn trả về, có thể thấy được rằng việc kết hợp hai đặc trưng cục bộ (SIFT) và toàn cục (Color Histogram) vẫn chưa có thể đem lại được kết quả như mong muốn. Việc kết hợp hai đặc trưng yêu cầu nhiều bước tính toán hơn, bao gồm trích xuất đặc trưng, tính toán độ đo tương đồng và kết hợp kết quả. Điều này dẫn đến thời gian truy vấn lâu hơn so với việc chỉ sử dụng một đặc trưng duy nhất. Mặc dù thời gian truy vấn tăng lên, độ chính xác của hệ thống vẫn chưa được cải thiện như mong đợi. Việc kết hợp hai đặc trưng không tạo ra sự khác biệt rõ rệt trong việc phân loại và truy vấn ảnh. Điều này có thể do các đặc trưng này chưa đủ mạnh để bù đắp cho nhau trong việc phân biệt các ảnh tương tự và khác nhau.
- Để cải thiện kết quả truy vấn ảnh, một trong những hướng giải quyết tiềm năng là thực hiện các bước tiền xử lý để tăng chất lượng ảnh trong bộ dữ liệu trước khi thực hiện rút trích đặc trưng. Các bước này có thể giúp làm nổi bật các đặc trưng quan trọng, giảm nhiễu và tăng tính đồng nhất của các đặc trưng được trích xuất. Một số phương pháp tiền xử lý có thể áp dụng như:
 - **Histogram Equalization**: giúp cân bằng độ sáng của ảnh, làm nổi bật các chi tiết và tăng độ tương phản.
 - **Sử dụng bộ lọc Gauss để loại bỏ nhiễu**: làm giảm nhiễu trong ảnh, giúp các đặc trưng quan trọng trở nên rõ ràng hơn.
 - **Làm sắc nét (Sharpening)**: giúp làm nổi bật các cạnh và chi tiết nhỏ trong ảnh.

- **Chuẩn hóa (Normalization):** giúp đảm bảo các ảnh có cùng tỷ lệ độ sáng và độ tương phản, giúp các đặc trưng trở nên đồng nhất hơn.
- Việc thực hiện các bước tiền xử lý để tăng chất lượng ảnh trước khi trích xuất đặc trưng có thể giúp cải thiện đáng kể độ chính xác của hệ thống truy vấn ảnh. Bằng cách làm nổi bật các đặc trưng quan trọng, giảm nhiễu và chuẩn hóa các ảnh, các đặc trưng được trích xuất sẽ trở nên đồng nhất và chính xác hơn, từ đó nâng cao khả năng phân biệt và truy vấn ảnh trong cơ sở dữ liệu.
- Tuy nhiên, do giới hạn về thời gian thực hiện đồ án cũng như việc cài đặt các đặc trưng ban đầu đã tốn nhiều thời gian hơn dự kiến để đảm bảo rằng các đặc trưng được trích xuất một cách chính xác và hiệu quả nên tuy đã xác định được tầm quan trọng của các bước tiền xử lý trong việc nâng cao độ chính xác của hệ thống, em vẫn chưa thể cài đặt và chạy thực nghiệm hệ thống khi thực hiện truy vấn trên các ảnh đã qua xử lý trước.

5 Tài liệu tham khảo

- [1] Tài liệu môn học slide, textbook
- [2] [Evaluation of ranked retrieval results](#)
- [3] [OpenCV modules](#)
- [4] [Introduction to SIFT \(Scale-Invariant Feature Transform\)](#)
- [5] [ORB \(Oriented FAST and Rotated BRIEF\)](#)
- [6] [Color Histograms in Image Retrieval](#)
- [7] [What Is Content-Based Image Retrieval?](#)