

**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**



**ĐỒ ÁN MÔN HỌC:**

**NHẬN DẠNG THỊ GIÁC VÀ ỨNG DỤNG**

**Đề tài:**

**HỆ THỐNG TÌM KIẾM BẰNG HÌNH ẢNH  
SỬ DỤNG PHƯƠNG PHÁP BAG OF WORD**

**GVHD: TS. Lê Đình Duy**

**TS. Nguyễn Tấn Trần Minh Khang**

**HỌC VIÊN: Phan Văn Khải – CH1601029**

**Lớp: Cao học khóa 11 – Đợt 2**

**TP. HỒ CHÍ MINH - NĂM 2017**

## MỤC LỤC

<b>MỤC LỤC .....</b>	<b>2</b>
<b>CHƯƠNG 1: TỔNG QUAN ĐỀ TÀI.....</b>	<b>1</b>
1.1. Mục tiêu và phạm vi đề tài.....	1
1.2. Môi trường, công cụ thực hiện.....	1
<b>CHƯƠNG 2: CHI TIẾT CÁC BƯỚC XÂY DỰNG HỆ THỐNG.....</b>	<b>2</b>
2.1. Xây dựng dữ liệu huấn luyện.....	2
2.2. Truy vấn ảnh .....	3
2.3. Xây dựng chức năng hệ thống .....	3
<b>CHƯƠNG 3: KẾT QUẢ VÀ ĐÁNH GIÁ THỰC NGHIỆM .....</b>	<b>5</b>
3.1. Kết quả đạt được .....	5
3.1.1. Tìm kiếm theo toàn bộ nội dung ảnh.....	6
3.1.2. Tìm theo một phần nội dung ảnh.....	8
3.1.3. Load ảnh từ thư mục máy tính.....	9
3.1.4. Các chức năng khác .....	12
3.2. Đánh giá .....	14
3.3. Hướng phát triển .....	14
<b>CHƯƠNG 4: TÀI LIỆU THAM KHẢO .....</b>	<b>15</b>

## **LỜI NÓI ĐẦU**

Sự phát triển của công nghệ thông, các thiết bị điện tử, thiết bị ghi nhận hình ảnh, video, âm thanh ngày càng mạnh mẽ. Ngày càng tạo ra nhiều thông tin, trong đó bao gồm cả thông tin hữu ích và cả thông tin không cần thiết với từng cá nhân. Việc tìm kiếm thông tin liên quan và hữu ích thực sự là một mối quan tâm của cá nhân và xã hội. Tránh lãng phí thời gian, công sức tìm kiếm. Việc tìm kiếm thủ công giữa kho ảnh khổng lồ với lượng lớn thông tin kèm theo, việc tìm được ảnh tương đồng hay nội dung liên quan đến ảnh truy vấn là khó khả thi. Do đó xây dựng một hệ thống tìm kiếm những ảnh tương đồng, liên quan đến ảnh truy vấn của người dùng.

Em chân thành cảm ơn những kiến thức quý báu đã được TS. Lê Đình Duy và TS. Nguyễn Tấn Trần Minh Khang truyền đạt, cùng việc được khơi dậy cảm hứng đam mê nghiên cứu. Đã tạo động lực và giúp em hoàn thành đồ án.

Do thời gian có hạn, kinh nghiệm và năng lực hạn chế và những điều kiện chủ quan cũng như khách quan dẫn đến đồ án còn nhiều sai sót, mong được các Thầy góp ý, sửa chữa để em hoàn thiện, nâng cao năng lực nghiên cứu, trình độ để tiếp tục học tập nghiên cứu thêm.

Một lần nữa em xin chân thành cảm ơn!

*TP Hồ Chí Minh, ngày 30 tháng 12 năm 2017*

Học viên thực hiện

**Phan Văn Khải**

# CHƯƠNG 1: TỔNG QUAN ĐỀ TÀI

## 1.1. Mục tiêu và phạm vi đề tài

Mục tiêu của đề án là xây dựng hệ thống truy vấn ảnh có chức năng tìm các ảnh có nội dung tương tự sử dụng bộ dataset Oxford building. Hệ thống có chức năng như sau:

- Dataset sử dụng Oxford Building (5K) có 5063 ảnh.
- Giao diện người dùng cho phép truy vấn trên các ảnh có sẵn trong thư mục ImageSearch hoặc chọn ảnh bất kỳ trong image oxford.
- Tìm kiếm theo toàn bộ nội dung của ảnh hoặc một phần nội dung ảnh đầu vào
- Xác định số lượng ảnh trả về trong đoạn [1 5063] kết quả.

Phạm vi đề tài chỉ sử dụng bộ dataset Oxford building để training, kết quả cũng sử dụng image trong bộ dataset có sẵn.

## 1.2. Môi trường, công cụ thực hiện

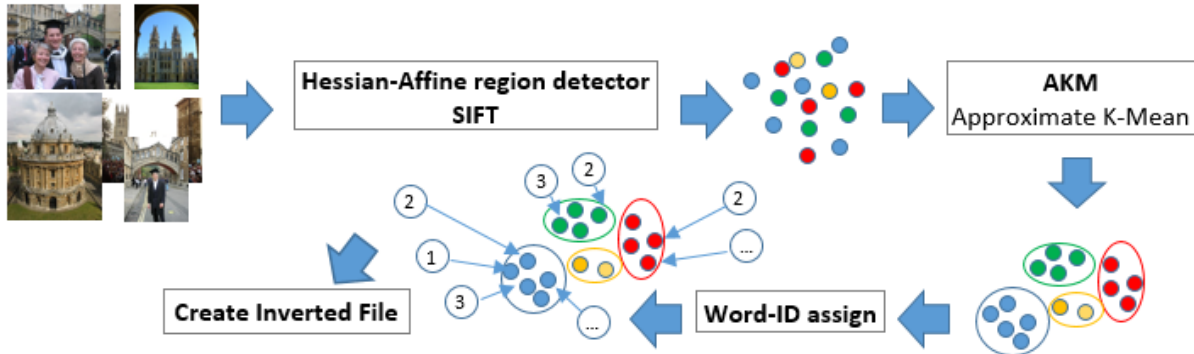
- Matlab 2017b.
- Vlfeat 0.9.20  
<http://www.vlfeat.org/download/vlfeat-0.9.20-bin.tar.gz>
- Dataset dữ liệu tòa nhà Oxford  
[http://www.robots.ox.ac.uk/~vgg/data/oxbuild\\_images.tgz](http://www.robots.ox.ac.uk/~vgg/data/oxbuildings/oxbuild_images.tgz)
- Các hàm trong thư viện AKM

## 1.3. Các thông tin liên quan

- Source code và các báo cáo lưu tại github theo địa chỉ  
<https://github.com/khaiphanvan/VRA>
- Địa chỉ YouTube demo chương trình, kết quả thực hiện:  
<https://www.youtube.com/watch?v=1nGJLImXXpI&feature=youtu.be>

## CHƯƠNG 2: CHI TIẾT CÁC BƯỚC XÂY DỰNG HỆ THỐNG

### 2.1. Xây dựng dữ liệu huấn luyện



#### Bước 1: Rút trích đặc trưng

- Ảnh đầu vào: 5063 ảnh
- Sử dụng Hessian-Affine region detector để rút trích các keypoint, sử dụng hàm `vl_codet` trong thư viện `vlfeat`.
- Tính đặc trưng SIFT trên các keypoint.
- Số chiều đặc trưng: 128.

#### Bước 2: Xây dựng Visual Words (dictionary)

- Sử dụng thuật toán gom cụm Approximate K-Mean (AKM).
- Số lượng cluster: 1.000.000
- **Em đã thử với số lượng cluster lần lượt là: 60.000 600.000 và 1.000.000 và nhận thấy với số lượng cluster 1.000.000 cho kết quả chính xác nhất**
- Số lượng k-d tree: 8.
- Số lần lặp: 5

#### Bước 3: Tính Word-ID cho từng ảnh trong dataset

- Với mỗi word trong từng document, ta tìm Word-ID của word dựa trên dictionary đã xây dựng ở bước 2
- Các khái niệm của xử lý ảnh được chuyển thành bài toán xử lý văn bản hay ngôn ngữ tự nhiên:

- visual words → dictionary
- feature → word
- index của feature → word ID
- ảnh → documents

#### **Bước 4: Xây dựng Inverted file**

- Với mỗi document, ta sẽ biết được trong document này có các word nào theo thứ tự.
- Inverted file: với mỗi word, ta sẽ lưu danh sách những document có chứa nó theo trọng số ‘tf-idf’.

## **2.2. Truy vấn ảnh**

**Bước 1:** Rút trích đặc trưng SIFT sử dụng Hessian-Affine region detector (tương tự bước 1 của quá trình xây dựng dữ liệu training).

**Bước 2:** tính Word-ID cho từng đặc trưng trong ảnh query đã tính ở bước 1.

**Bước 3:** Tính ranked list

- Sử dụng hàm ccvBowGetWords trong thư viện vlfeat tính vào tạo bảng word count
- Sử dụng inverted file để so sánh query document với tất cả các document trong inverted file bằng khoảng cách euclidean.
- Tạo bảng ranked list theo tự giảm dần.

**Bước 4:** Trả kết quả về

- Dựa theo số lượng kết quả người dùng nhập ở màn hình, lấy **n** hình trả về hiển thị lên màn hình kết quả.

## **2.3. Xây dựng chức năng hệ thống**

Xây dựng giao diện trên Matlab có chức năng tìm kiếm theo toàn bộ nội dung ảnh, một phần nội dung ảnh. Xác định số lượng kết quả trả về. Tìm kiếm bằng hình ảnh bất kỳ

**Bước 1:** Load sẵn 5 ảnh lên thumb, cho phép người dùng chọn để load lên ảnh truy vấn

**Bước 2:** Load ảnh bất kỳ lên dùng để truy vấn

**Bước 3:** Chọn một ảnh bất kỳ từ thư mục máy tính dùng truy vấn

**Bước 4:** Set số ảnh kết quả muốn trả về

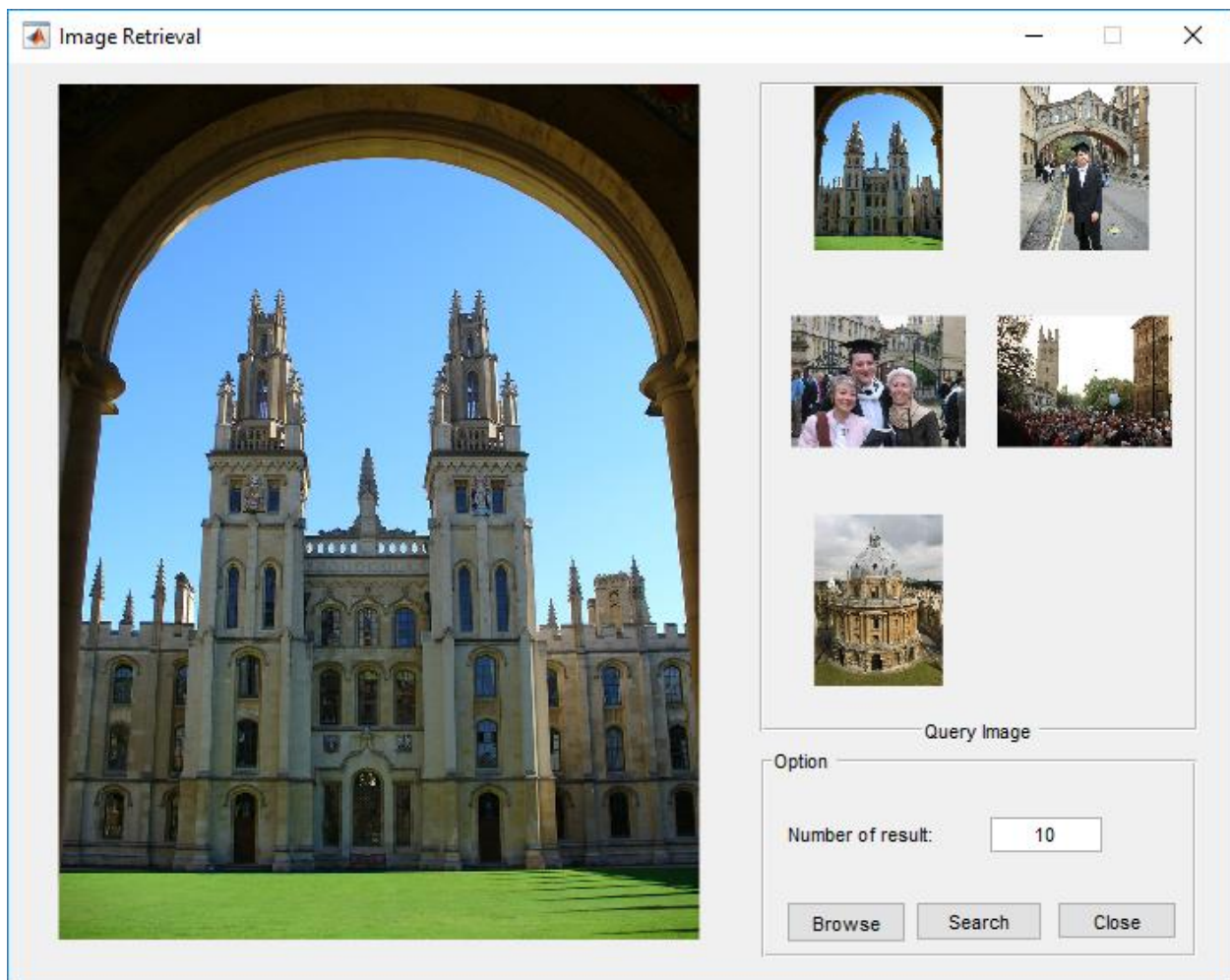
**Bước 5:** Tìm kiếm ảnh trả về kết quả nếu có hiển thị lên màn hình kết quả 10 ảnh/ trang, thông báo nếu không tìm thấy ảnh nào phù hợp.

## CHƯƠNG 3: KẾT QUẢ VÀ ĐÁNH GIÁ THỰC NGHIỆM

### 3.1. Kết quả đạt được

- Xây dựng được hệ thống tìm kiếm hình ảnh trên tập dataset Oxford Building, có giao diện
- Có chức năng tìm kiếm theo nội dung toàn bộ ảnh.
- Có chức năng tìm kiếm theo một phần nội dung ảnh.
- Trả về số kết quả theo thiết lập của người dùng.

#### Giao diện chương trình



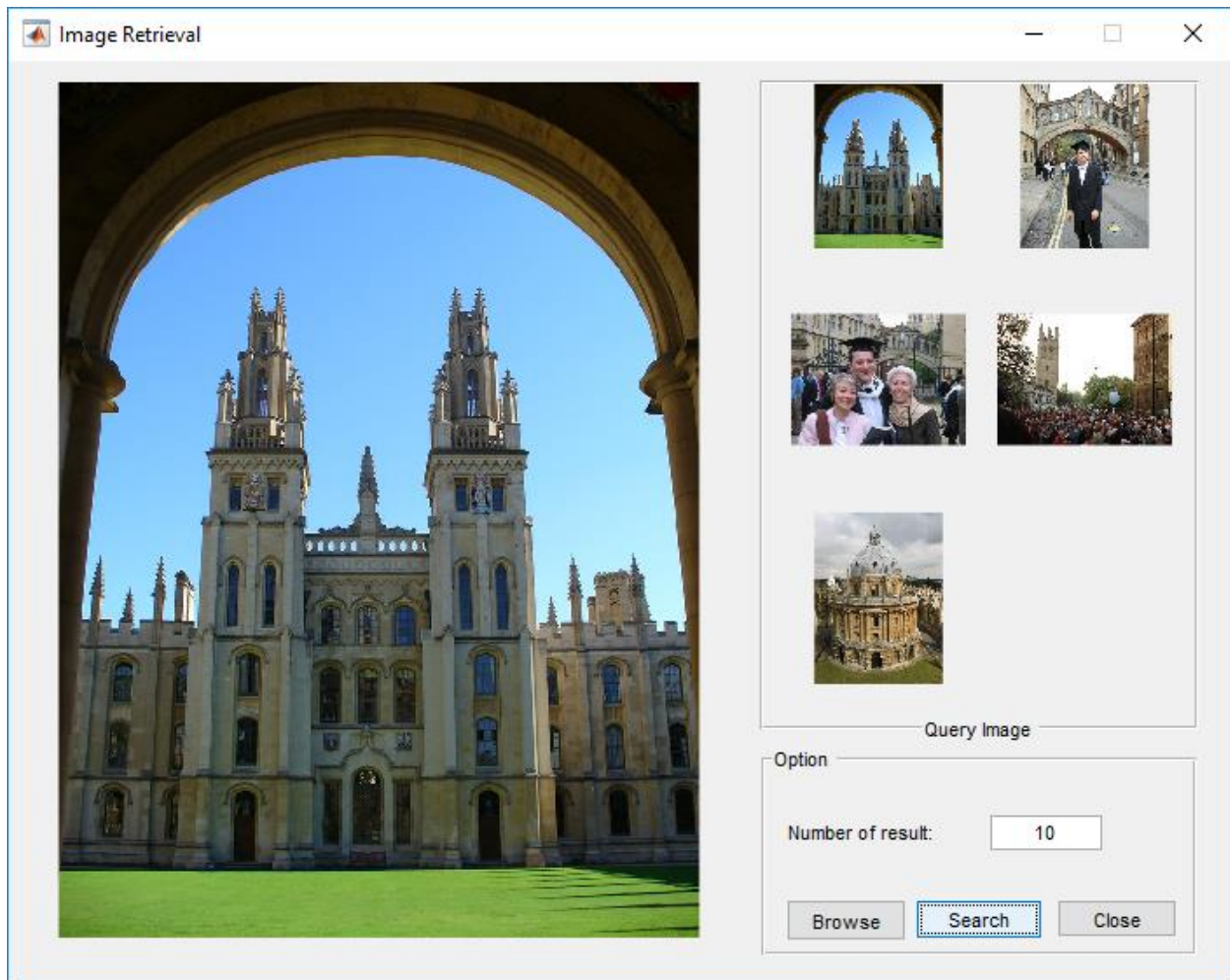
Load sẵn 5 ảnh làm query

Số lượng image result trả về, mặc định 10

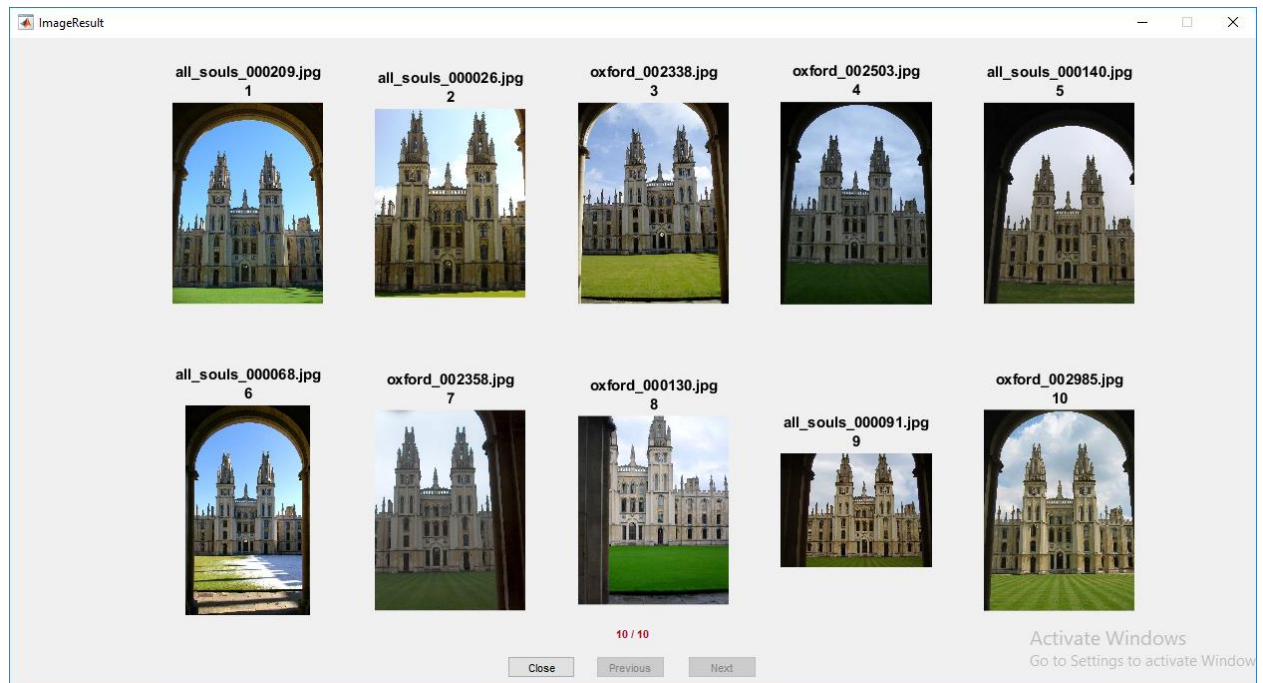
Nút Browse để load ảnh bất kỳ lên truy vấn



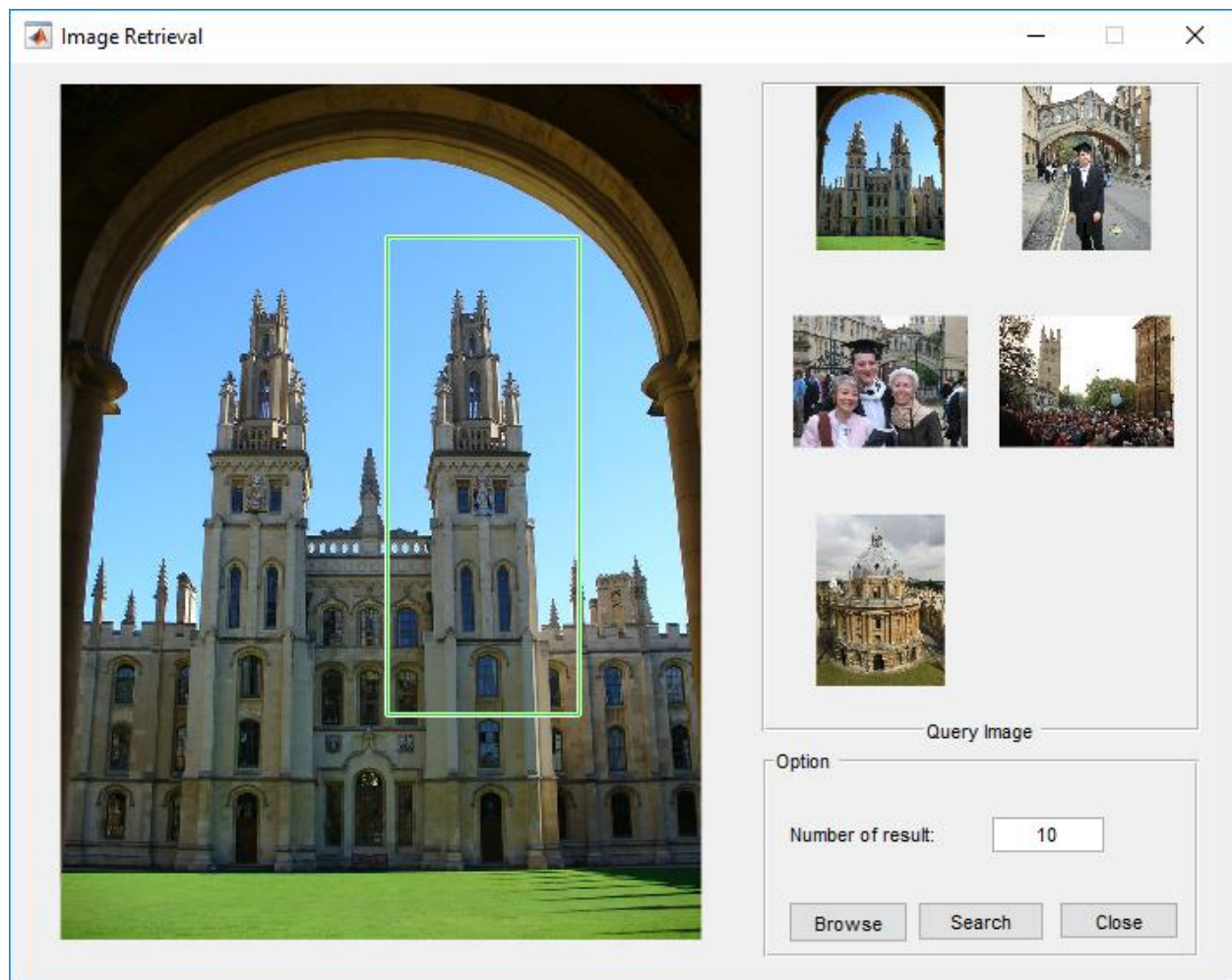
### 3.1.1. Tìm kiếm theo toàn bộ nội dung ảnh



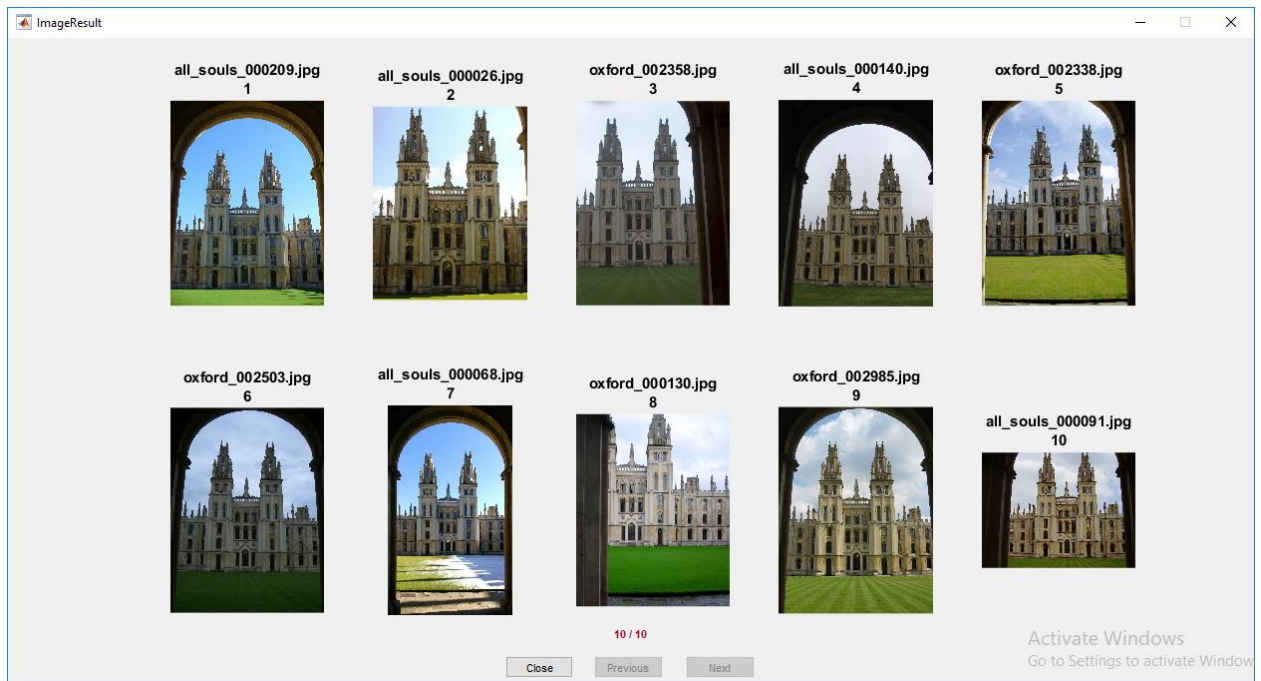
## Kết quả trả về



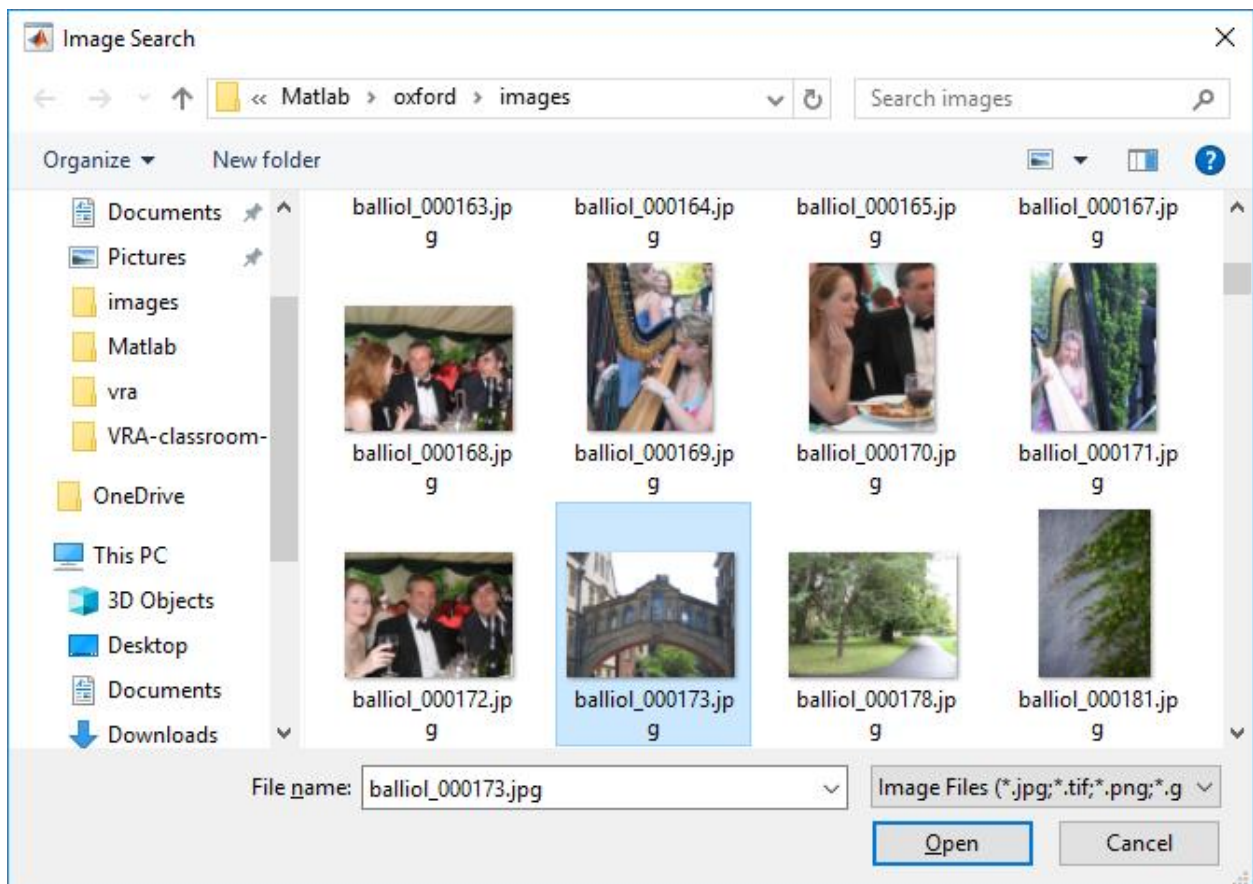
### 3.1.2. Tìm theo một phần nội dung ảnh



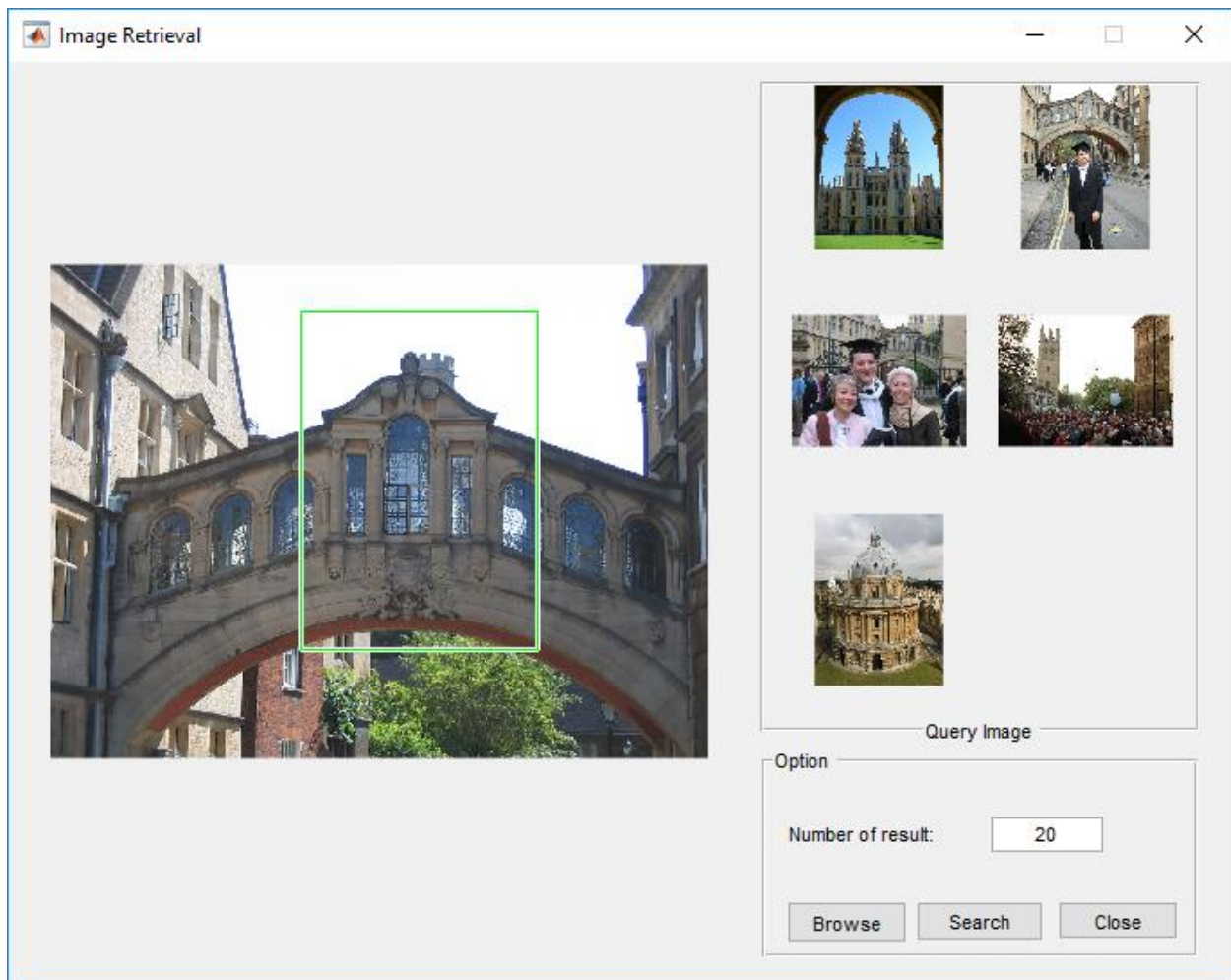
## Kết quả trả về



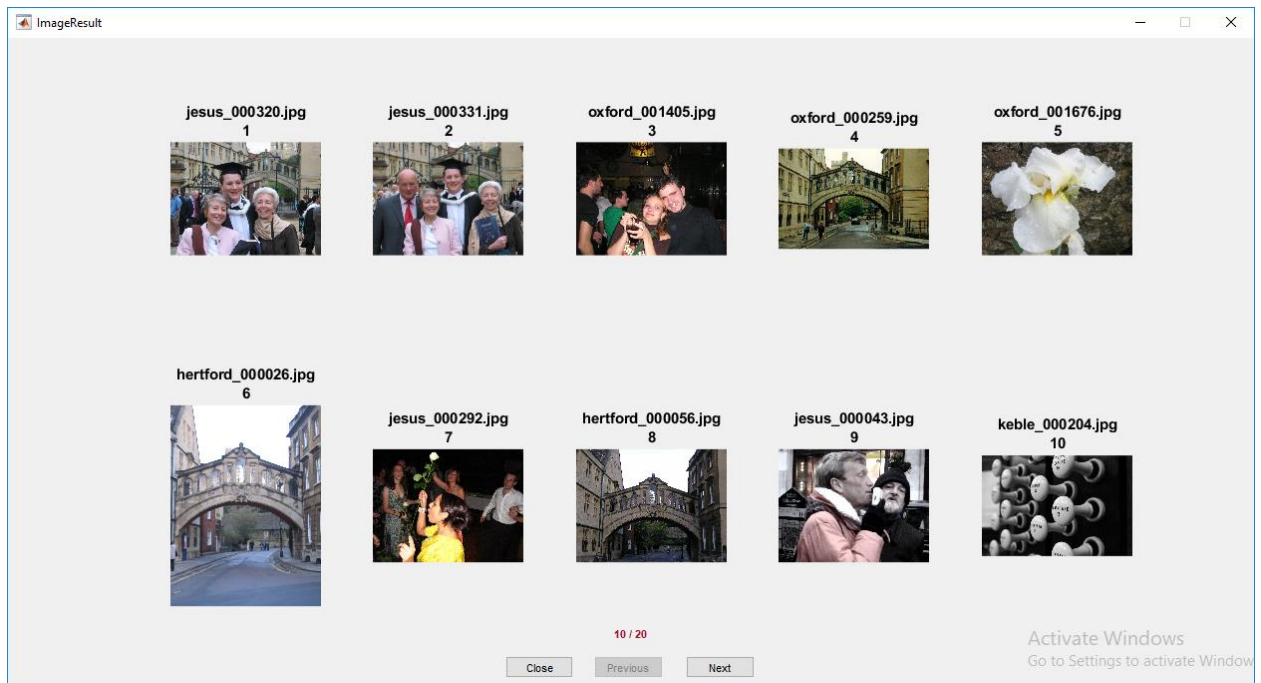
### 3.1.3. Load ảnh từ thư mục máy tính





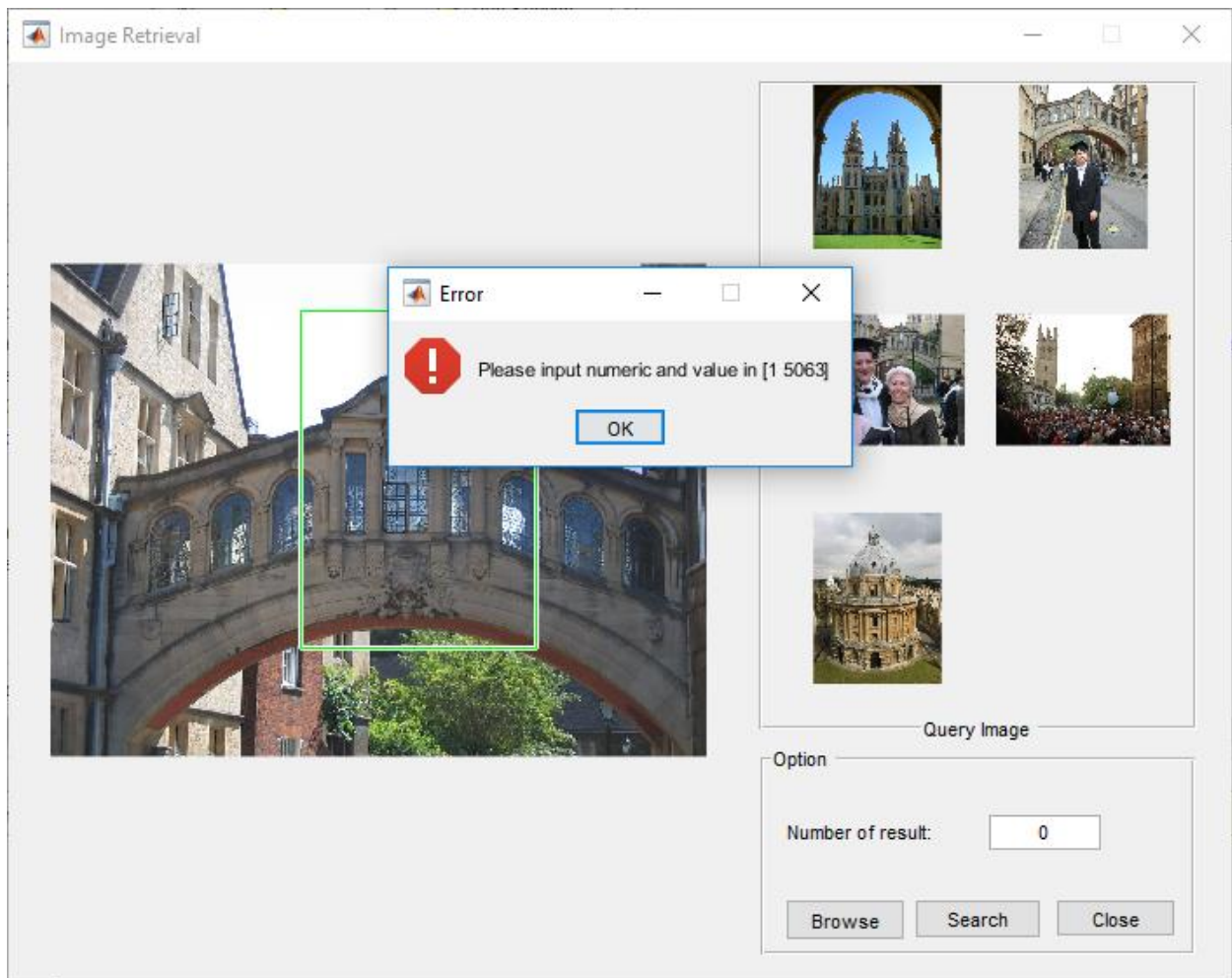


## Kết quả trả về

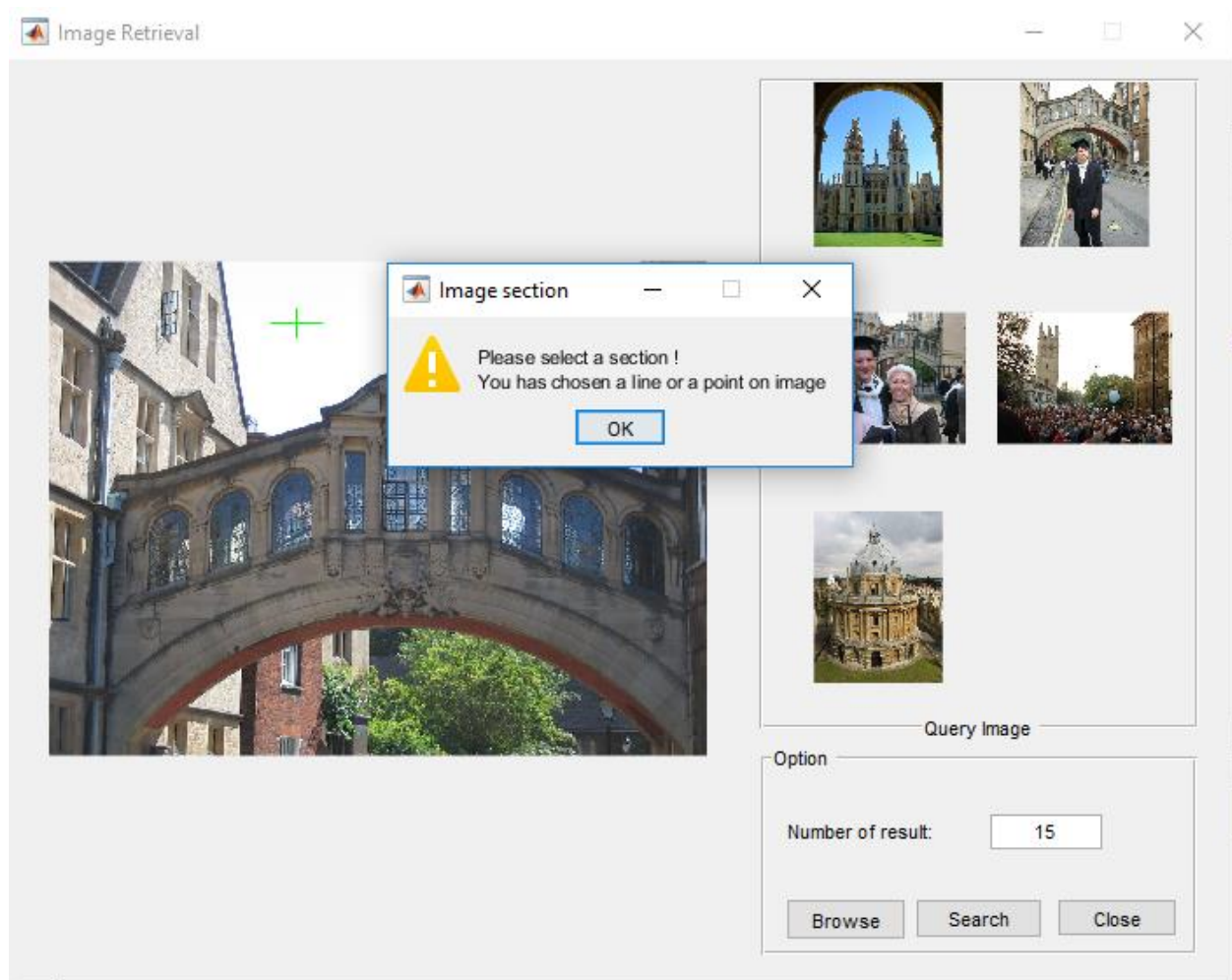


### 3.1.4. Các chức năng khác

Thông báo khi người dùng nhập số kết quả trả về không chính xác



Chọn vùng quá nhỏ





### **3.2. Đánh giá**

- Sau khi tiến hành thực nghiệm, em nhận thấy kết quả trả về tương đối chính xác khi truy vấn.
- Trường hợp truy vấn bằng vùng ảnh, nếu vùng ảnh truy vấn không thuộc vùng rút trích đặc trưng, có thể coi trong file \*query.txt thì kết quả trả về có độ chính xác chưa cao

### **3.3. Hướng phát triển**

- Đưa hệ thống lên website.
- Tăng tốc độ xử lý quá trình tìm kiếm.
- Training với bộ data lớn hơn, đa dạng hơn.

## CHƯƠNG 4: TÀI LIỆU THAM KHẢO

### Tiếng việt

[1] Bài giảng môn: Nhận dạng thị giác và ứng dụng Lê Đình Duy, Nguyễn Tấn Trần Minh Khang.

### Tiếng Anh

[2] <http://www.robots.ox.ac.uk/~vgg/data/oxbuildings/>

### Tài liệu Website

[3] <https://github.com/nvtiep/Instance-Search>

[4] <http://www.vlfeat.org/benchmarks/overview/retrieval.html>

[5] <http://viet.wordnet.vn/wnms>

[6] T. MathWorks, " Image Retrieval with Bag of Visual Words" 2017. [Online]. Available: <https://www.mathworks.com/help/vision/ug/image-retrieval-with-bag-of-visual-words.html>. [Accessed 7 December 2017].