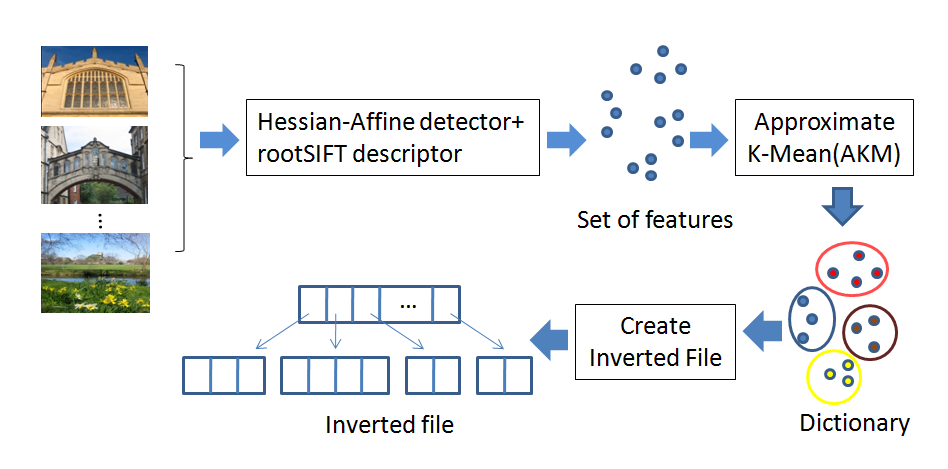
# 1. Mục tiêu của task

Mục tiêu của task này là xây dựng một framework truy vấn thông tin hình ảnh trên dataset Oxford bao gồm 5063 ảnh. Input đầu vào là một ảnh truy vấn, output: ranked list được sắp xếp từ thấp đến cao theo mức độ tương đồng.

# 2. Chi tiết các bước thực hiện

## 2.1. Phase 1: Training Offline

****

Hình 1. Quy trình training offline

Bước 1: Rút trích đặc trưng

* Sử dụng Hessian-Affine region detector để rút trích các keypoint
* Tính đặc trưng SIFT trên các keypoint
* Tính đặc trưng rootSIFT: rootsift = sqrt(sift / sum(sift) )
* 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
* Số lượng k-d tree: 8
* Số lần lặp: 5

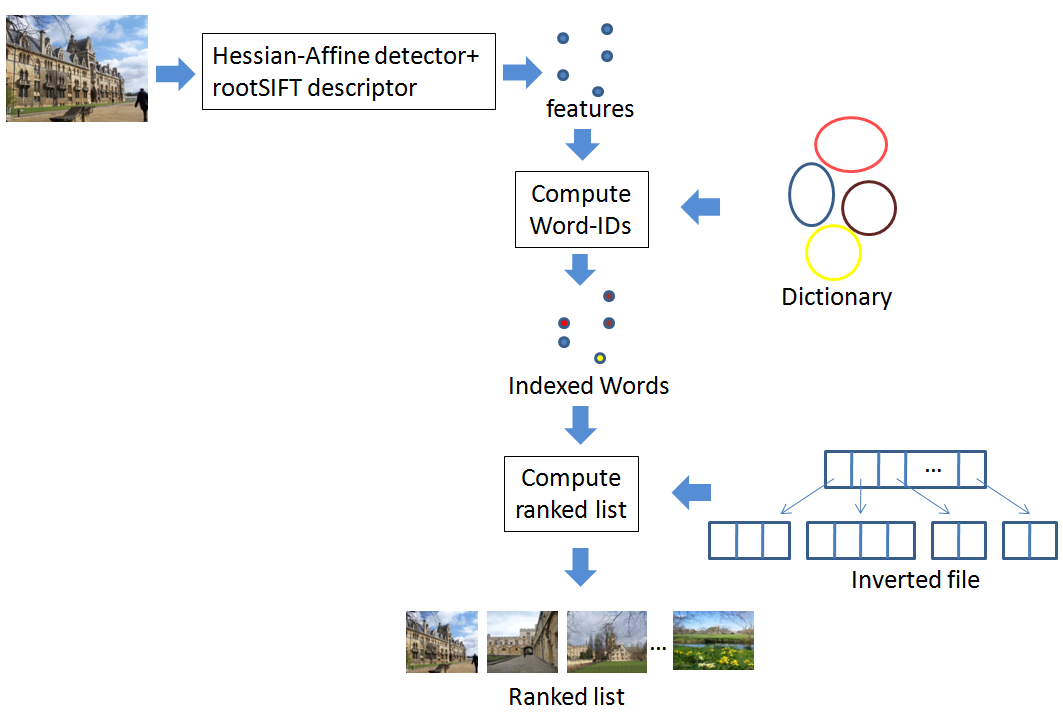
Bước 3: Tính Word-ID cho từng ảnh trong dataset

* Với mỗi word (rootSIFT feature) trong từng document (ảnh), ta tìm Word-ID của word dựa trên dictionary (tập Visual Words) đã xây dựng ở bước 2
* Ở bước này ta sẽ chuyển các khái niệm của xử lý ảnh sang 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

* Theo thứ tự bình thường, với mỗi document, ta sẽ biết được trong document này có các word nào
* Inverted file: với mỗi word, ta sẽ lưu danh sách những document có chứa nó
* ‘tf-idf’ weighting: các visual word xuất hiện ở nhiều class of document thì càng ít có vai trò để phân loại một document nên được đánh trọng số thấp hơn. Các visual word xuất hiện ở càng ít các class of document thì có trọng số cao hơn.

## 2.2. Phase 2: Truy vấn ảnh



Hình 2. Quy trình truy vấn ánh

Bước 5: phát hiện và rút trích đặc trưng rootSIFT sử dụng Hessian-Affine region detector (tương tự bước 1)

Bước 6: tính Word-ID cho từng feature trong ảnh query (tương tự bước 3)

Bước 7: Tính ranked list

* Xây dựng bảng Word Count để thống kê các word và tần số xuất hiện tương ứng
* Sử dụng inverted file để so sánh query document với tất cả các document trong trong inverted file 🡪list score distance
* Sắp xếp list score theo tự giảm dần

Bước 8: Evaluation

* Tập ảnh truy vấn gồm 55 ảnh khoanh vùng các đối tượng truy vấn chính
* Các ảnh thuộc tập ground truth “good” và “ok” đứng ở những vị trí đầu tiên của rank\_list thì độ chính xác càng cao và ngược lại.

# 3. Kết quả thực nghiệm

Bảng 1. Độ chính xác khi truy vấn trên 55 ảnh ground truth

|  |  |  |
| --- | --- | --- |
| **STT** | **Query image** | **Accuracy** |
| 1 | all\_souls\_000013 | 0.450577 |
| 2 | all\_souls\_000026 | 0.300251 |
| 3 | all\_souls\_000051 | 0.760197 |
| 4 | ashmolean\_000000 | 0.828235 |
| 5 | ashmolean\_000007 | 0.737466 |
| 6 | ashmolean\_000058 | 0.503754 |
| 7 | ashmolean\_000269 | 0.745224 |
| 8 | ashmolean\_000305 | 0.621897 |
| 9 | balliol\_000051 | 0.637355 |
| 10 | balliol\_000167 | 0.174537 |
| 11 | balliol\_000187 | 0.304682 |
| 12 | balliol\_000194 | 0.520144 |
| 13 | bodleian\_000107 | 0.526571 |
| 14 | bodleian\_000108 | 0.686305 |
| 15 | bodleian\_000132 | 0.614126 |
| 16 | bodleian\_000163 | 0.513866 |
| 17 | bodleian\_000407 | 0.438443 |
| 18 | christ\_church\_000179 | 0.786827 |
| 19 | christ\_church\_000999 | 0.475605 |
| 20 | christ\_church\_001020 | 0.719164 |
| 21 | cornmarket\_000019 | 0.851731 |
| 22 | cornmarket\_000047 | 0.763913 |
| 23 | cornmarket\_000105 | 0.333794 |
| 24 | cornmarket\_000131 | 0.654789 |
| 25 | hertford\_000015 | 0.845602 |
| 26 | hertford\_000027 | 0.883455 |
| 27 | hertford\_000063 | 0.857492 |
| 28 | keble\_000028 | 0.482374 |
| 29 | keble\_000055 | 0.887559 |
| 30 | keble\_000214 | 1 |
| 31 | keble\_000227 | 1 |
| 32 | keble\_000245 | 1 |
| 33 | magdalen\_000058 | 0.236529 |
| 34 | magdalen\_000078 | 0.0488831 |
| 35 | magdalen\_000560 | 0.197525 |
| 36 | oxford\_000317 | 0.686436 |
| 37 | oxford\_000545 | 0.654268 |
| 38 | oxford\_001115 | 0.0510465 |
| 39 | oxford\_001752 | 0.861477 |
| 40 | oxford\_001753 | 0.441524 |
| 41 | oxford\_002416 | 0.752533 |
| 42 | oxford\_002562 | 0.508901 |
| 43 | oxford\_002734 | 0.647257 |
| 44 | oxford\_002904 | 0.882599 |
| 45 | oxford\_002985 | 0.434347 |
| 46 | oxford\_003335 | 0.171284 |
| 47 | oxford\_003410 | 0.699102 |
| 48 | pitt\_rivers\_000033 | 0.676313 |
| 49 | pitt\_rivers\_000058 | 1 |
| 50 | pitt\_rivers\_000087 | 0.85078 |
| 51 | pitt\_rivers\_000119 | 1 |
| 52 | pitt\_rivers\_000153 | 0.710125 |
| 53 | radcliffe\_camera\_000095 | 0.717263 |
| 54 | radcliffe\_camera\_000519 | 0.802418 |
| 55 | radcliffe\_camera\_000523 | 0.692497 |

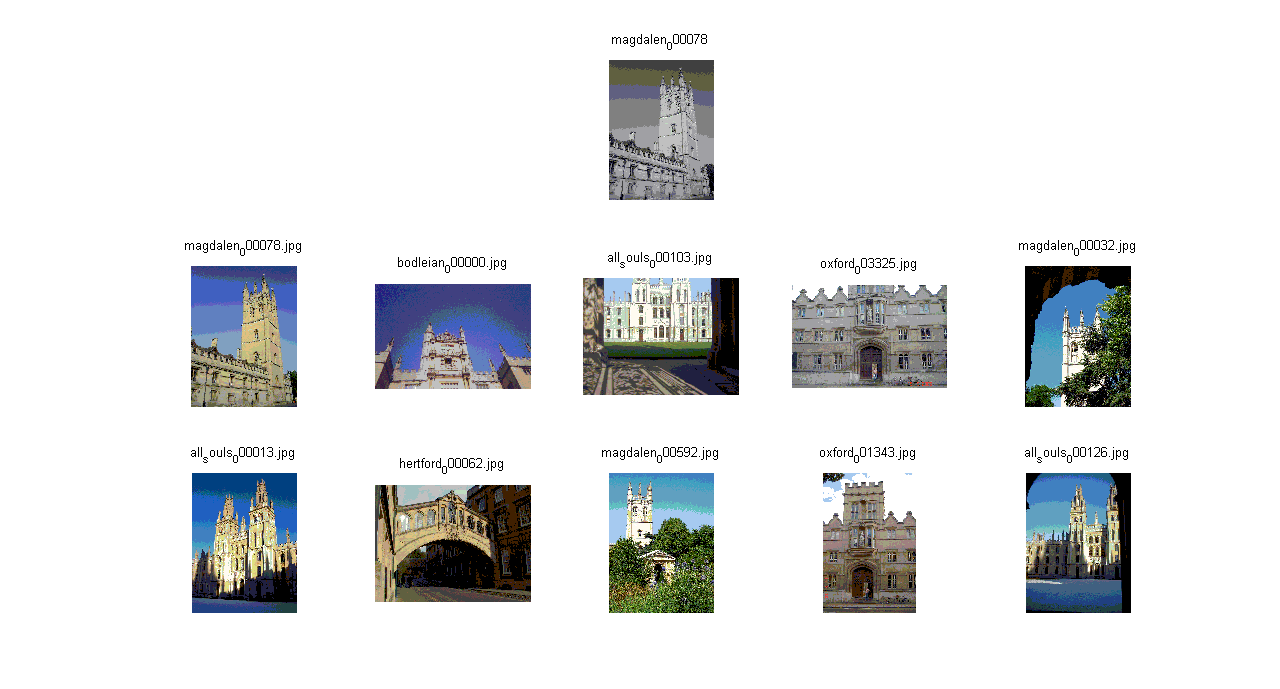
**Trung bình: 0.6296**

10 ảnh query có độ chính xác thấp nhất được liệt kê trong Bảng 2.

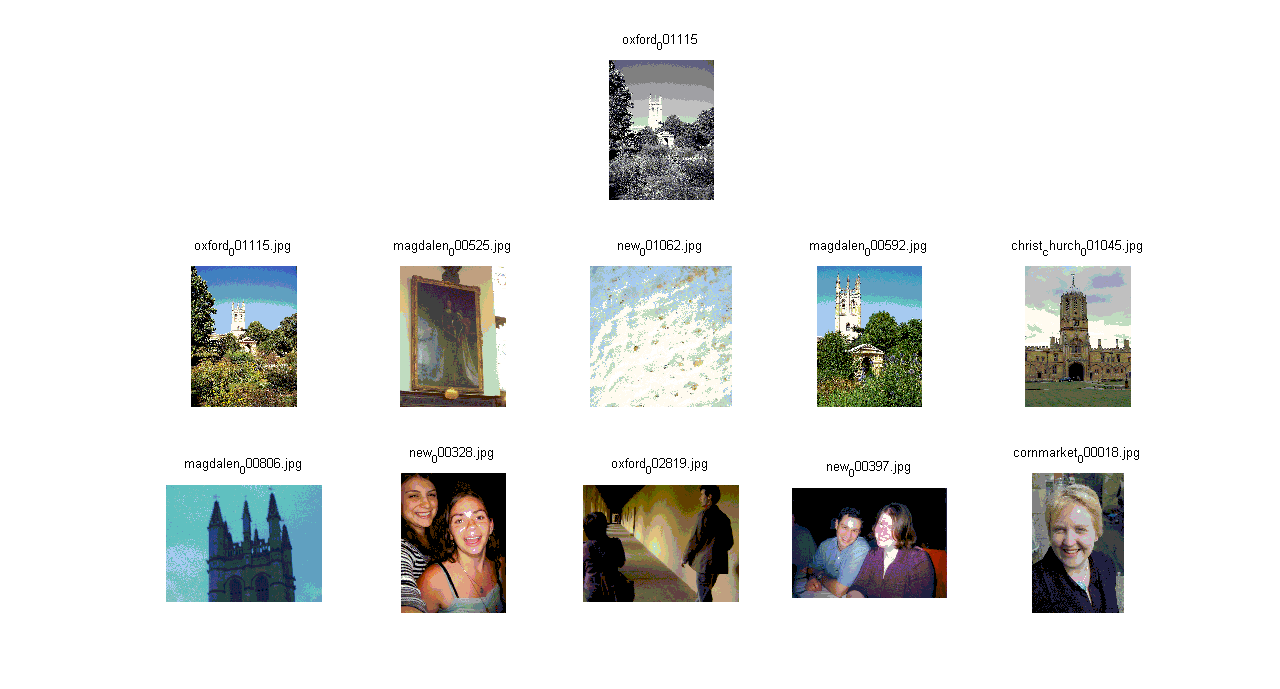
Bảng . 10 ảnh truy vấn có độ chính xác thấp nhất

|  |  |  |
| --- | --- | --- |
| **STT** | **Query image** | **Accuracy** |
| 34 | magdalen\_000078 | 0.0488831 |
| 38 | oxford\_001115 | 0.0510465 |
| 46 | oxford\_003335 | 0.171284 |
| 10 | balliol\_000167 | 0.174537 |
| 35 | magdalen\_000560 | 0.197525 |
| 33 | magdalen\_000058 | 0.236529 |
| 2 | all\_souls\_000026 | 0.300251 |
| 11 | balliol\_000187 | 0.304682 |
| 23 | cornmarket\_000105 | 0.333794 |
| 45 | oxford\_002985 | 0.434347 |

Một số ảnh kết quả cho trường hợp truy vấn đạt kết quả chính xác thấp:



Hình 4.Ảnh truy vấn magdalen\_00078 và 10 ảnh có score khoảng cách gần nhất



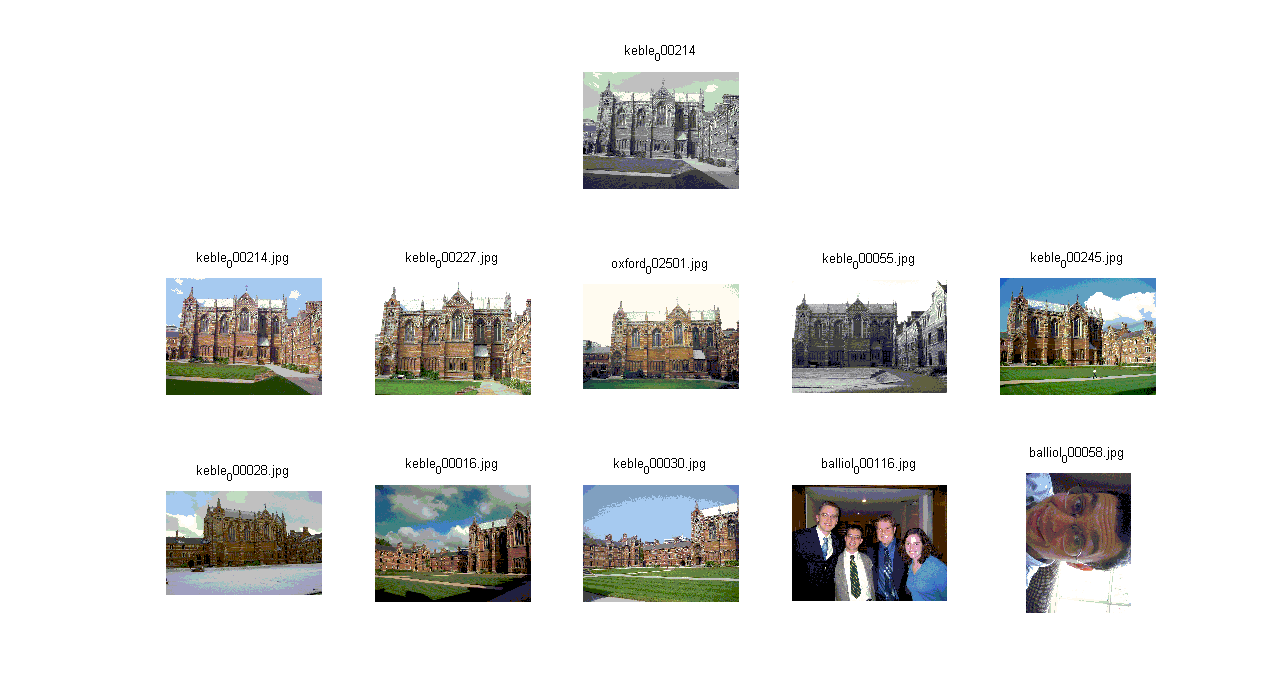
Hình 5.Ảnh truy vấn oxford\_001115 và 10 ảnh có score khoảng cách gần nhất

10 ảnh query có độ chính xác cao nhất được liệt kê trong Bảng 3.

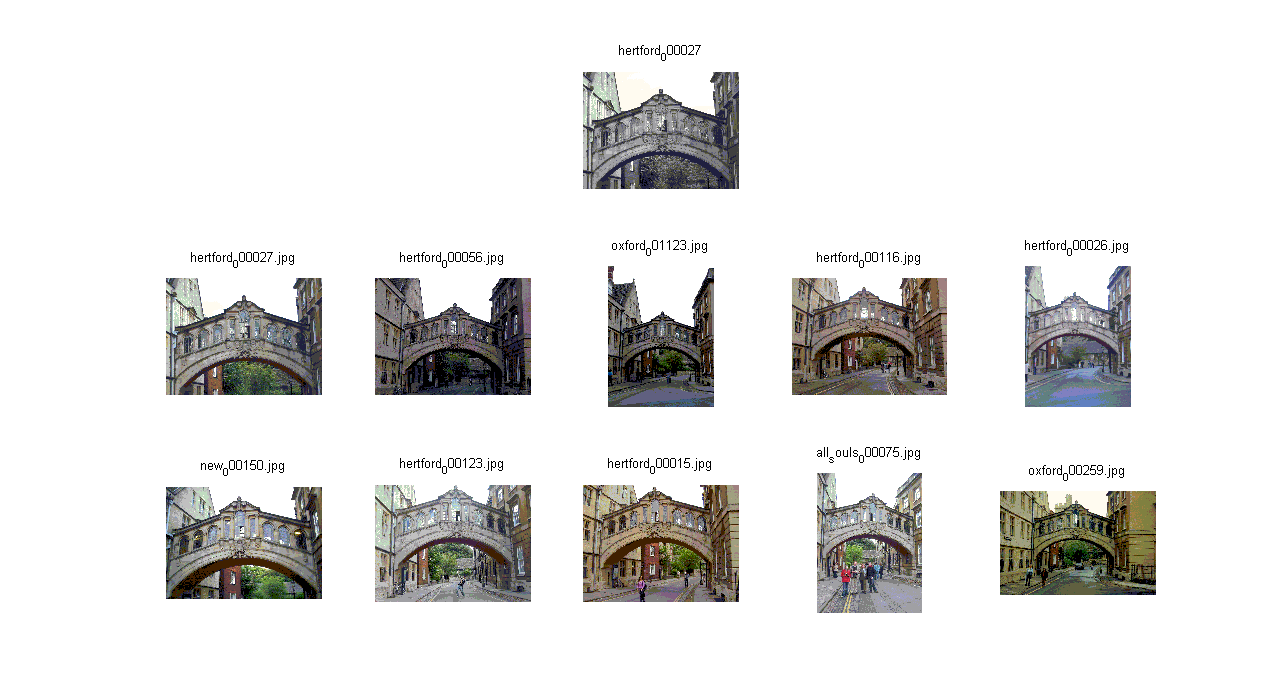
Bảng . 10 ảnh truy vấn có độ chính xác cao nhất

|  |  |  |
| --- | --- | --- |
| **STT** | **Query image** | **Accuracy** |
| 30 | keble\_000214 | 1 |
| 31 | keble\_000227 | 1 |
| 32 | keble\_000245 | 1 |
| 49 | pitt\_rivers\_000058 | 1 |
| 51 | pitt\_rivers\_000119 | 1 |
| 29 | keble\_000055 | 0.887559 |
| 26 | hertford\_000027 | 0.883455 |
| 44 | oxford\_002904 | 0.882599 |
| 39 | oxford\_001752 | 0.861477 |
| 27 | hertford\_000063 | 0.857492 |

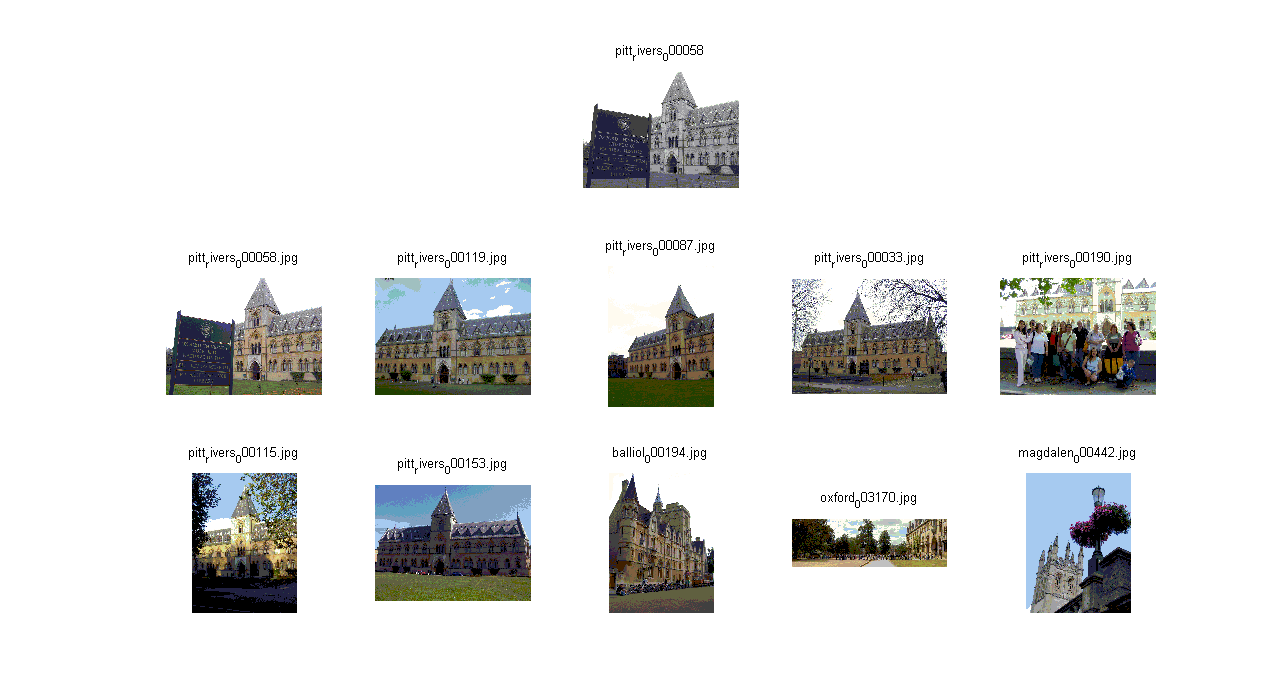
Một số ảnh kết quả cho trường hợp truy vấn đạt kết quả chính xác cao:



Hình . Ảnh truy vấn keble\_000214 và 10 ảnh có score khoảng cách gần nhất



Hình . Ảnh truy vấn hertford\_00027 và 10 ảnh có score khoảng cách gần nhất



Hình . Ảnh truy vấn pitt\_rivers\_00058 và 10 ảnh có score khoảng cách gần nhất