

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

Lab 03

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

| | | |
|----|---------------------------|---|
| 1. | Xây dựng mã nguồn | 3 |
| 2. | Yêu cầu 1 | 3 |
| a. | Quá trình thực hiện | 3 |
| b. | Kết quả và đánh giá..... | 4 |
| 3. | Yêu cầu 2 | 8 |
| a. | Quá trình thực hiện | 8 |
| b. | Kết quả và đánh giá..... | 8 |

1. Xây dựng mã nguồn

- Mã nguồn được xây dựng theo kiến trúc hướng đối tượng (OOP) với các lớp chính là **Feature**, **Dataset** và **BagOfWord_Model**. Trong đó:
 - o **Feature**: là lớp có thuộc tính là các đặc trưng (về histogram và descriptor) của ảnh, cũng như các phương thức được xây dựng để tính toán các đặc trưng đó.
 - o **Dataset**: là lớp có thuộc tính là tập ảnh (tập ảnh huấn luyện và kiểm thử) và các đặc trưng của tập ảnh tương ứng. Lớp được xây dựng có các phương thức để tính toán các đặc trưng cho tập ảnh.
 - o **BagOfWord_Model**: là lớp có thuộc tính là các từ trực quan (visual words) cùng với các phương thức để xây dựng visual word và tính toán giá trị histogram cho tập visual words đó.
- Chương trình được chia làm 6 file chính là, trong đó
 - o **main.cpp**: chứa mã nguồn chính để thực thi chương trình.
 - o **bow_model.cpp, dataset.cpp, feature.cpp**: chứa các nội dung xây dựng các phương thức của các lớp tương ứng (Feature, Dataset, BagOfWord_Model).
 - o **support_function.cpp**: chứa các hàm hỗ trợ trong quá trình xây dựng mã nguồn (như tạo canvas để hiển thị các ảnh trong cùng một cluster ở yêu cầu 1).
 - o **function.h**: chứa các khai báo thư viện, hàm cần thiết cũng như định nghĩa các lớp có trong mã nguồn.
- Chương trình được thực thi trên Terminal dưới lệnh sau:
`<ten_file_thuc_thi.exe> <yeu_cau_thuc_hien> <train_file_path><test_file_path><k_cluster>`

Trong đó,

- o `<ten_file_thuc_thi.exe>`: là file thực thi
- o `<yeu_cau_thuc_hien>`: đại diện cho yêu cầu muốn thực hiện (1 hoặc 2)
- o `<train_file_path>`: là đường dẫn đến nơi lưu trữ ảnh tập huấn luyện
- o `<test_file_path>`: là đường dẫn đến nơi lưu trữ ảnh tập thử nghiệm
- o `<k_cluster>`: giá trị K cụm tương ứng
- Ví dụ, để chạy chương trình, ta tiến hành mở terminal, truy cập đến thư mục chứa mã nguồn, và thực hiện gọi lệnh:
 - o Nếu ta muốn sử dụng dữ liệu từ webcam, ta sẽ thực thi lệnh sau

`21127090_Lab03.exe 1 ../training_images/*.jpg ../TestImages/*.jpg 5`

- Trong đó:
 - o 21127090_Lab02.exe: là file thực thi chương trình
 - o 1: là yêu cầu được thực hiện
 - o `../training_images/*.jpg`: là đường dẫn đến nơi lưu trữ hình ảnh của tập huấn luyện, lấy tất cả các ảnh có đuôi .jpg
 - o `../TestImages/*.jpg`: là đường dẫn đến nơi lưu trữ hình ảnh của tập kiểm thử, lấy tất cả các ảnh có đuôi .jpg
 - o 5: là số lượng cụm K tương ứng.

2. Yêu cầu 1

a. Quá trình thực hiện

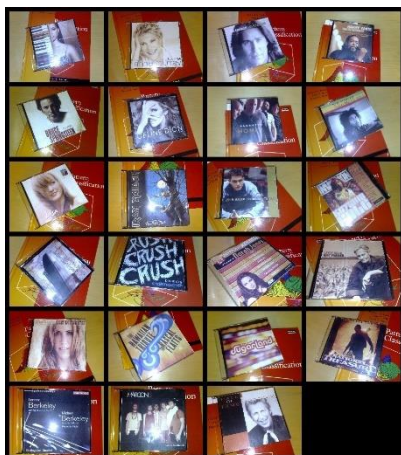
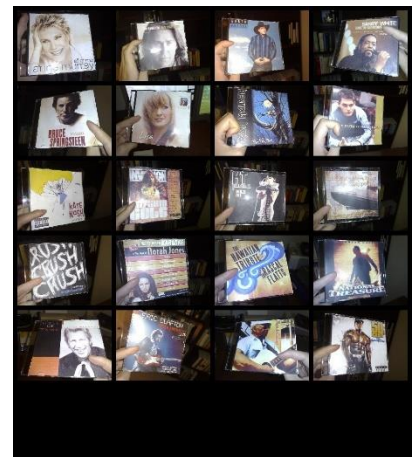
- Yêu cầu này được thực hiện tuần tự theo các bước dưới đây:
 - o Bước 1: Tính toán histogram đối với mỗi ảnh trong tập huấn luyện của bộ dữ liệu CD. Do các ảnh trong bộ dữ liệu đều là ảnh màu, nên việc tính toán histogram của ảnh cũng được tính toán

theo 3 kênh màu khác nhau (đỏ, xanh lục, xanh lam), đem đi chuẩn hoá lại và sau đó tổng hợp lại thành một ma trận.

- Để phù hợp với dữ liệu đầu vào của thuật toán Kmeans, các kết quả của histogram của tập ảnh sẽ được nối với nhau, tạo thành một ma trận lớn, đồng thời chuyển đổi kiểu dữ liệu của nó sang một loại khác là CV_32F
- Bước 2: Áp dụng thuật toán K-means với kết quả ở bước trên với số lượng cụm K tùy ý (5, 10, 15). Sử dụng nhãn *label* mà thuật toán trả ra để thực hiện tiếp tục.
- Bước 3: Gom các ảnh có cùng nhãn lại với nhau, và hiển thị chúng trên cùng một cửa sổ. Đồng thời, mã nguồn cũng đã lưu lại các ảnh trong cùng một cụm lại với nhau theo tên **k=<số cụm>_Cluster<thứ i>.jpg**
 - Giả dụ ta thực hiện gom nhóm với số cụm K=5, kết quả của cluster thứ 1 sẽ là: **k=5_Cluster1.jpg**

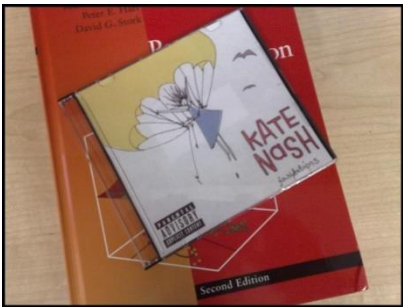
b. Kết quả và đánh giá

- K = 5

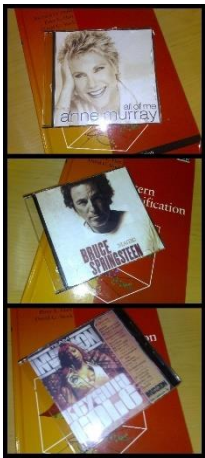


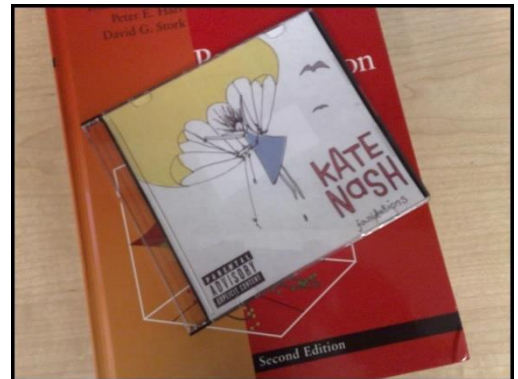
- K = 10





- K = 15





- **Đánh giá:** Với kết quả đạt được và được trình bày ở trên với giá trị K tương ứng, có thể thấy rằng việc lựa chọn giá trị K có ảnh hưởng mạnh mẽ đến kết quả gom nhóm mà ta thu được.
 - o Với giá trị K quá nhỏ ($K=5$), các ảnh được gom về cùng một nhóm dù vẫn khá giống nhau về màu sắc, tuy nhiên chỉ ở mức tương đối. Điều này có thể dễ dàng quan sát bằng mắt thường.

- Với giá trị K lớn hơn (K=10), kết quả có sự cải thiện tốt hơn giữa các ảnh trong cùng một cluster. Các ảnh trong cùng một cluster đã giảm bớt sự khác biệt so với kết quả trước đó.
- Với giá trị K = 15, kết quả có phần rõ ràng hơn. Trong cùng một cluster, các ảnh có màu sắc tương tự nhau khá nhiều, ít có sự khác biệt và khó có thể quan sát bằng mắt. Tuy nhiên, sẽ tồn tại những cụm chỉ có một đến hai phần tử.
- **Kết luận:** Từ kết quả thực nghiệm và nhận xét được nêu ra, có thể thấy rằng việc chọn giá trị K lớn cho ta kết quả tốt trong việc phân bổ ảnh vào các cụm dựa vào histogram của nó. Tuy nhiên, đi kèm với kết quả tốt sẽ là việc tiêu tốn tài nguyên cho quá trình tính toán phân cụm, đây là sự đánh đổi kèm theo.

3. Yêu cầu 2

a. Quá trình thực hiện

- Yêu cầu này được thực hiện tuần tự theo các bước sau:
 - Bước 1: Rút trích đặc trưng của ảnh bằng thuật toán SIFT, thu được giá trị descriptor tương ứng cho mỗi ảnh.
 - Sau khi lướt qua tất cả ảnh trong tập huấn luyện, ta sẽ thu được một tập các descriptor tương ứng. Tương tự như ở yêu cầu 1, để phù hợp với đầu vào của thuật toán Kmeans, ta tiến hành nối các descriptor lại thành một ma trận lớn.
 - Bước 2: Áp dụng thuật toán Kmeans lên ma trận lớn ở bước 1 với K tùy ý (50, 100, 200). Sử dụng các điểm trung tâm *centers* mà thuật toán trả ra để tiếp tục thực hiện.
 - Các điểm *centers* này chính là *visual word* mà thuật toán tìm được.
 - Bước 3: Để có thể đưa ra nhận xét khi thay đổi K, ta cần thực hiện thêm bước tính toán histogram của các *visual word* trong một ảnh (ảnh ở đây ta sẽ sử dụng 5 ảnh trong bộ ảnh kiểm thử của tập dữ liệu CD).
 - Sau khi rút trích đặc trưng của bộ ảnh kiểm thử, ta thực hiện tính toán histogram với chúng, thu được kết quả được trình bày ở mục tiếp theo.

b. Kết quả và đánh giá

Sử dụng 5 ảnh trong tập ảnh thử nghiệm của bộ dữ liệu CD, kết quả dưới đây là giá trị histogram thu được với số bin K tương ứng

- **K = 50**

```
Image1
[87, 47, 48, 54, 83, 47, 84, 157, 43, 43, 145, 34, 45, 48, 88, 145, 98, 54, 45, 38, 79, 48, 52, 38, 56, 61, 25, 69, 68, 47, 56, 68, 106, 45, 96, 126, 59, 74, 36, 17, 144, 59, 68, 46, 59, 38, 68, 56, 41, 32]

Image2
[59, 55, 16, 19, 24, 15, 18, 183, 6, 24, 86, 19, 28, 23, 2, 198, 19, 24, 19, 19, 16, 59, 17, 20, 27, 19, 17, 188, 69, 27, 15, 23, 177, 28, 22, 46, 2, 7, 26, 18, 12, 89, 34, 14, 25, 25, 22, 32, 8, 48, 19]

Image3
[59, 58, 38, 28, 26, 20, 28, 185, 33, 27, 51, 23, 33, 29, 145, 124, 34, 29, 15, 15, 32, 32, 15, 19, 19, 29, 12, 22, 31, 19, 35, 38, 48, 24, 49, 36, 37, 24, 47, 7, 168, 15, 36, 38, 14, 22, 28, 18, 57, 27]

Image4
[125, 75, 63, 94, 78, 55, 82, 191, 69, 63, 142, 72, 133, 46, 145, 167, 53, 92, 63, 55, 81, 52, 56, 96, 54, 67, 68, 117, 188, 95, 69, 49, 284, 68, 99, 154, 89, 136, 55, 47, 294, 84, 53, 85, 76, 59, 77, 82, 163, 75]

Image5
[59, 30, 44, 44, 45, 23, 28, 184, 24, 47, 67, 31, 34, 21, 57, 53, 25, 58, 25, 27, 49, 18, 26, 54, 35, 23, 49, 38, 62, 26, 39, 48, 68, 79, 49, 132, 7, 5, 58, 35, 8, 94, 45, 54, 48, 36, 38, 52, 43, 41, 36]
```

- Số lượng visual words phổ biến ở các ảnh khá ít (khoảng 3 đến 5 bin) đạt giá trị cao trong một ảnh.
- Tại một số bin (như bin thứ 7 và 15), giá trị của chúng ở 5 ảnh gần như cao, chứng tỏ rằng chúng là đặc trưng chung, phổ biến, xuất hiện hầu hết trong các ảnh.

- Việc phân bố các đặc trưng ở các ảnh có sự chênh lệch, trong khi các ảnh như 1 và 3 có sự phân bố các đặc trưng đều nhau, thì ở ảnh 2, các đặc trưng lại không được phân bố như vậy. Nó gần như phân bố ở các đặc trưng chính khá nhiều.

- **K = 100**

```

Image1
[21, 36, 35, 45, 73, 18, 48, 22, 26, 23, 17, 27, 28, 19, 26, 53, 21, 17, 19, 20, 27, 99, 22, 19, 101, 17, 23, 25, 110, 18, 29, 20, 51, 22, 25, 31, 2, 5, 32, 17, 24, 39, 25, 51, 95, 14, 34, 58, 32, 22, 13, 17, 20, 26, 38, 22, 26, 44, 14, 19, 23, 23, 29, 35, 39, 32, 10, 38, 19, 115, 46, 28, 39, 43, 36, 31, 24, 11, 37, 37, 49, 35, 24, 29, 30, 36, 26, 17, 64, 39, 29, 21, 11, 21, 16, 37, 45, 21, 27, 7, 0]

Image2
[11, 17, 14, 5, 156, 4, 12, 15, 9, 4, 11, 15, 38, 12, 9, 16, 16, 13, 8, 7, 12, 68, 14, 12, 168, 7, 68, 12, 79, 10, 13, 9, 2, 46, 17, 8, 15, 9, 8, 19, 8, 10, 12, 34, 21, 8, 54, 14, 27, 13, 11, 3, 7, 8, 16, 6, 57, 20, 12, 14, 9, 4, 17, 9, 6, 6, 4, 16, 83, 12, 7, 18, 11, 8, 15, 13, 12, 16, 15, 12, 44, 19, 15, 6, 8, 12, 1, 13, 5, 11, 25, 9, 17, 10, 6, 6, 13, 6, 11, 0]

Image3
[19, 30, 17, 15, 31, 11, 38, 12, 7, 18, 9, 10, 24, 20, 28, 24, 15, 15, 14, 34, 8, 15, 24, 8, 82, 10, 7, 5, 133, 14, 11, 19, 115, 23, 15, 18, 17, 12, 5, 16, 53, 11, 23, 17, 21, 17, 23, 16, 34, 15, 6, 10, 17, 13, 9, 18, 19, 22, 17, 7, 12, 11, 25, 12, 24, 5, 19, 10, 144, 37, 11, 7, 9, 20, 5, 18, 25, 3, 7, 2, 21, 19, 12, 11, 36, 14, 5, 23, 14, 19, 2, 10, 11, 7, 10, 21, 8, 28, 2, 0]

Image4
[30, 37, 53, 37, 152, 31, 45, 33, 32, 25, 40, 40, 26, 23, 26, 57, 24, 58, 36, 42, 56, 75, 52, 29, 110, 29, 56, 37, 152, 15, 33, 38, 84, 25, 50, 27, 44, 32, 32, 31, 67, 26, 57, 117, 53, 28, 78, 31, 98, 37, 27, 18, 52, 36, 35, 39, 91, 22, 28, 44, 32, 62, 49, 37, 29, 72, 70, 45, 119, 67, 27, 38, 27, 59, 35, 53, 30, 31, 40, 36, 69, 28, 49, 22, 65, 22, 21, 49, 39, 47, 48, 33, 45, 27, 99, 25, 25, 51, 40, 2]

Image5
[20, 42, 32, 14, 50, 13, 49, 31, 15, 23, 17, 28, 9, 19, 19, 12, 25, 29, 25, 32, 14, 39, 16, 15, 26, 17, 14, 22, 66, 23, 27, 14, 42, 10, 17, 7, 16, 1, 7, 15, 23, 36, 13, 22, 93, 21, 10, 29, 24, 19, 24, 12, 20, 17, 19, 10, 21, 48, 7, 18, 22, 17, 29, 19, 15, 29, 3, 20, 35, 66, 53, 10, 15, 27, 20, 14, 32, 22, 15, 11, 19, 21, 12, 23, 5, 16, 32, 22, 26, 18, 31, 24, 17, 36, 16, 26, 8, 30, 14, 1, 0]

```

- Số lượng visual words phổ biến ở các ảnh được tăng lên nhiều hơn (khoảng 10 bin hoặc hơn).
- Trong kết quả này, nó cũng tồn tại một số visual word phổ biến, có tần suất xuất hiện lớn trong cả 5 ảnh (như bin thứ 25, 27, 29, ...). Điều này cho thấy rằng đây là những đặc trưng chung của bộ dữ liệu.
- So với kết quả thu được từ **K=50**, kết quả của giá trị **K=100** cho ra các phân bố đều nhau hơn giữa các visual words trong cùng một ảnh.

- **K = 200**

```

Image1
[10, 23, 17, 8, 24, 11, 19, 19, 15, 20, 7, 20, 11, 13, 2, 13, 14, 25, 33, 60, 27, 44, 13, 4, 25, 93, 5, 12, 22, 8, 24, 8, 27, 22, 24, 12, 5, 18, 18, 9, 11, 9, 5, 9, 13, 8, 25, 11, 12, 8, 24, 9, 18, 13, 9, 5, 20, 11, 8, 6, 18, 3, 9, 15, 15, 12, 18, 19, 53, 19, 10, 6, 10, 13, 29, 8, 10, 9, 9, 1, 4, 8, 13, 74, 19, 13, 17, 18, 20, 8, 16, 14, 14, 7, 18, 16, 7, 10, 6, 34, 5, 4, 9, 15, 18, 24, 29, 2, 14, 7, 16, 14, 7, 18, 8, 13, 23, 7, 53, 7, 2, 17, 12, 16, 31, 15, 6, 60, 10, 27, 10, 5, 13, 9, 23, 19, 12, 17, 31, 7, 8, 24, 17, 24, 7, 3, 18, 8, 18, 18, 7, 1, 17, 17, 34, 26, 27, 11, 7, 12, 18, 15, 12, 5, 14, 9, 11, 16, 14, 7, 12, 9, 2, 16, 26, 23, 23, 13, 16, 15, 69, 30, 14, 24, 12, 13, 12, 7, 19, 10, 17, 9, 9, 25, 14, 6, 8, 31, 6, 14]

Image2
[4, 16, 8, 5, 10, 3, 6, 3, 3, 3, 1, 20, 5, 8, 4, 10, 7, 8, 12, 115, 4, 11, 1, 4, 2, 54, 2, 6, 0, 4, 42, 8, 3, 8, 10, 4, 12, 0, 17, 0, 4, 4, 6, 3, 18, 7, 3, 2, 0, 6, 3, 7, 49, 4, 2, 4, 6, 13, 5, 4, 9, 2, 10, 4, 8, 4, 1, 5, 136, 10, 6, 7, 24, 3, 4, 0, 4, 5, 5, 7, 3, 3, 68, 3, 6, 7, 3, 6, 0, 1, 6, 6, 12, 29, 5, 3, 9, 6, 9, 2, 5, 9, 12, 10, 3, 6, 9, 3, 3, 7, 6, 4, 5, 4, 2, 5, 2, 12, 12, 3, 3, 10, 5, 4, 10, 2, 21, 6, 2, 16, 9, 4, 3, 3, 9, 3, 7, 2, 5, 8, 6, 5, 10, 2, 14, 2, 5, 14, 9, 6, 1, 8, 6, 36, 3, 4, 9, 6, 5, 5, 13, 10, 3, 7, 5, 0, 14, 8, 7, 4, 5, 4, 2, 8, 34, 17, 9, 8, 1, 69, 7, 8, 5, 4, 6, 12, 8, 8, 3, 7, 20, 2, 9, 1, 2, 33, 5, 3, 4]

Image3
[6, 32, 2, 10, 22, 5, 8, 24, 10, 16, 4, 11, 8, 8, 3, 6, 17, 22, 30, 94, 7, 9, 9, 2, 4, 22, 12, 2, 51, 17, 4, 3, 3, 12, 4, 10, 34, 12, 2, 8, 1, 1, 8, 1, 29, 14, 62, 6, 7, 6, 4, 8, 11, 5, 10, 3, 7, 5, 12, 3, 2, 1, 11, 9, 1, 13, 5, 7, 20, 16, 4, 11, 7, 7, 1, 9, 1, 3, 15, 11, 4, 4, 115, 3, 2, 1, 9, 10, 8, 6, 7, 10, 10, 14, 7, 4, 0, 7, 16, 4, 6, 0, 4, 1, 24, 12, 0, 4, 2, 6, 13, 4, 9, 10, 9, 4, 7, 13, 15, 1, 4, 12, 6, 11, 3, 1, 9, 10, 9, 9, 12, 8, 7, 8, 7, 6, 5, 5, 6, 1, 5, 7, 10, 10, 3, 9, 19, 11, 10, 8, 0, 0, 8, 7, 10, 13, 15, 1, 0, 6, 3, 7, 2, 5, 2, 16, 0, 5, 3, 13, 6, 0, 6, 12, 14, 1, 6, 9, 8, 112, 11, 2, 18, 18, 5, 9, 11, 14, 3, 6, 11, 4, 6, 0, 2, 17, 11, 6, 18]

Image4
[14, 14, 18, 13, 20, 16, 20, 22, 22, 39, 32, 40, 12, 23, 1, 12, 22, 39, 37, 79, 12, 52, 20, 13, 27, 61, 20, 13, 47, 9, 42, 11, 37, 18, 35, 23, 13, 1, 1, 23, 15, 7, 17, 21, 14, 9, 29, 18, 43, 20, 19, 14, 8, 24, 78, 22, 22, 15, 23, 14, 13, 18, 14, 7, 15, 12, 13, 23, 16, 11, 92, 33, 14, 15, 8, 20, 13, 20, 17, 1, 22, 14, 8, 30, 89, 13, 52, 14, 15, 23, 11, 11, 24, 16, 13, 27, 34, 22, 14, 18, 28, 13, 13, 2, 8, 60, 44, 34, 5, 17, 13, 19, 16, 13, 14, 39, 31, 19, 11, 47, 19, 7, 15, 9, 37, 32, 15, 20, 55, 68, 38, 43, 31, 14, 18, 39, 23, 18, 19, 29, 17, 7, 28, 19, 18, 14, 15, 27, 54, 27, 23, 17, 18, 17, 15, 29, 16, 9, 32, 18, 10, 12, 14, 17, 10, 25, 25, 20, 15, 18, 54, 144, 11, 5, 26, 38, 37, 9, 4, 34, 37, 74, 27, 18, 35, 20, 29, 15, 31, 11, 1, 3, 20, 14, 16, 43, 9, 10, 11, 27, 10, 28]

Image5
[8, 11, 8, 11, 12, 10, 6, 14, 11, 15, 17, 12, 7, 6, 1, 0, 11, 10, 49, 18, 11, 37, 7, 3, 7, 29, 8, 22, 10, 9, 16, 14, 7, 15, 5, 0, 3, 17, 14, 10, 6, 17, 11, 16, 9, 38, 10, 34, 15, 4, 12, 18, 7, 14, 10, 7, 7, 6, 4, 1, 13, 9, 2, 43, 10, 3, 5, 1, 15, 9, 17, 7, 4, 12, 7, 6, 7, 3, 1, 5, 7, 15, 6, 18, 11, 14, 0, 16, 5, 6, 13, 8, 12, 3, 10, 3, 0, 3, 7, 9, 8, 2, 7, 8, 8, 11, 4, 12, 9, 13, 12, 4, 24, 6, 1, 5, 10, 12, 7, 10, 11, 46, 7, 34, 30, 7, 21, 15, 27, 11, 21, 0, 12, 0, 13, 19, 7, 10, 7, 6, 9, 11, 14, 10, 9, 0, 4, 13, 12, 5, 4, 19, 5, 15, 8, 22, 7, 2, 7, 9, 13, 14, 11, 11, 7, 8, 0, 0, 3, 20, 13, 3, 3, 10, 12, 29, 12, 6, 11, 22, 13, 6, 9, 10, 9, 16, 7, 9, 23, 11, 11, 3, 0, 4, 7]

```

- Có thể thấy, tần suất xuất hiện của các visual words trong cùng một ảnh được phân bố đều nhau, không có sự chênh lệch quá lớn như kết quả thu được ở hai giá trị K trước. Điều này dẫn đến việc số lượng visual words phổ biến trong một ảnh cũng được tăng lên rất nhiều.
- Với kết quả phân bố các visual words như trên, ta dễ dàng đưa ra nhận xét về các đặc trưng trong ảnh.

- **Kết luận:**

- Với số lượng K nhỏ, histogram thường có xu hướng đơn giản và ít chi tiết hơn so với nhưng giá trị K lớn. Với cách phân bố như thế, các đặc trưng được thu thập và biểu diễn có khả năng không tốt và không đáng kể.
- Với giá trị K lớn, đồng nghĩa với việc histogram cho ra sẽ chi tiết và phức tạp hơn rất nhiều, tuy nhiên, nó lại cung cấp một lượng thông tin chi tiết và đa dạng hơn về các đặc trưng có trong ảnh.
- Tuy nhiên, đánh đổi với kết quả có được tốt như thế, việc thực hiện xây dựng mô hình Bag Of Word bằng Kmeans với lượng K lớn đòi hỏi nguồn tài nguyên và thời gian thực hiện khá lớn.