

Image Similarity Analysis Based on Color Histograms and Dominant Colors

AACHABI Mohammed
Computer Science Department
Faculty of Sciences and Technology, Tangier, Morocco

Abstract

Image similarity analysis is a key domain in computer vision, used in applications such as content-based image retrieval, image library management, and visual recommendation systems. This project explores and compares images based on color characteristics using two main approaches: color histograms and dominant color extraction. The similarity is computed using histogram-based Bhattacharyya distance, dominant color distance, and a combined global metric.

Contents

1	Introduction	2
2	Project Structure	3
2.1	Objectives	3
3	Methodology	4
3.1	Step 1: Loading and Preprocessing Images	4
3.2	Step 2: Color Histogram Calculation	4
3.3	Step 3: Dominant Color Extraction	5
3.4	Step 4: Similarity Distance Calculations	5
4	Results	7
5	Conclusion	8

Chapter 1

Introduction

The primary goal of this project is to analyze and compare a set of images based on their color characteristics to determine their similarity. The color features used for this analysis include:

- Color histograms for each color channel (Red, Green, Blue) to quantify color distribution.
- Dominant colors extracted using KMeans clustering, focusing on significant colors only.

Chapter 2

Project Structure

The project is organized as follows:

- **DVI1:** Folder containing 16 reference images for similarity comparison, categorized by types such as animals, flowers, ocean, and sky.
- **DVI2:** Folder with 4 test images, one per category, used for similarity testing.

2.1 Objectives

The specific objectives of this project include:

- Extracting color characteristics (histograms and dominant colors) for each image.
- Comparing images based on Bhattacharyya distance for histograms, Euclidean distance for dominant colors, and a global metric combining both.
- Visualizing and ranking the most similar image pairs for each metric.

Chapter 3

Methodology

3.1 Step 1: Loading and Preprocessing Images

Images from the specified directory are loaded and harmonized in format for analysis.



Figure 3.1: Sample image from DVI1 folder

3.2 Step 2: Color Histogram Calculation

Histograms are calculated for each color channel (Red, Green, Blue) using OpenCV, then normalized to enable consistent comparisons across images.

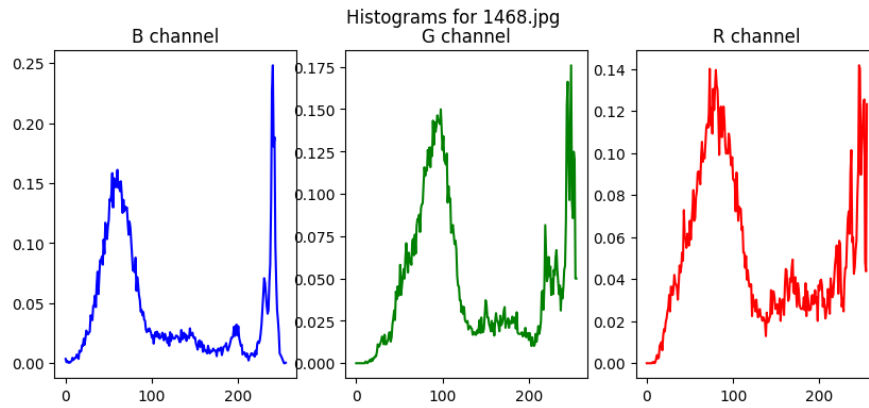


Figure 3.2: Color Histogram for an image

Listing 3.1: Python Code for Calculating Color Histograms

```

1 def calculate_color_histogram(image_path):
2     img = cv2.imread(image_path)
3     chans = cv2.split(img)
4     histograms = {color: cv2.normalize(cv2.calcHist([chan], [0],
5         None, [256], [0, 256]), hist).flatten() for chan, color in
6         zip(chans, ('b', 'g', 'r'))}
7     return histograms

```

3.3 Step 3: Dominant Color Extraction

KMeans clustering is applied to each image to extract dominant colors, with a threshold of 5% to retain only significant colors, filtering out noise and preserving the most representative colors.

Listing 3.2: Python Code for Finding Dominant Colors

```

1 def find_dominant_colors(image_path, k=20, threshold=0.05):
2     img = cv2.imread(image_path)
3     img = img.reshape((-1, 3))
4     kmeans = KMeans(n_clusters=k, random_state=0).fit(img)
5     dominant_colors = [kmeans.cluster_centers_[idx].tolist() for
6         idx, count in Counter(kmeans.labels_).items() if count /
7         total_pixels > threshold]
8     return dominant_colors

```

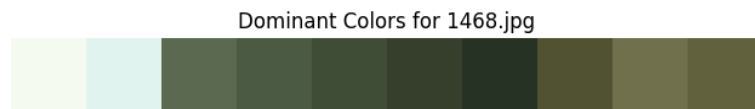


Figure 3.3: Dominant Colors for an image

3.4 Step 4: Similarity Distance Calculations

Similarity calculations are based on:

- **Bhattacharyya Distance** between histograms of each color channel.
- **Euclidean Distance** between dominant colors.
- **Global Metric** combining normalized distances from histograms and dominant colors.

Listing 3.3: Python Code for Calculating Bhattacharyya Distance

```
1 def bhattacharyya_distance(hist1, hist2):  
2     return cv2.compareHist(hist1, hist2, cv2.HISTCMP_BHATTACHARYYA  
    )
```


Chapter 4

Results

The results show the similarity between images using three different metrics: histogram similarity, dominant color similarity, and global metric similarity. Below is a comparison based on the global metric.

Top Similar Images by Global Metric for ./DV12\3486.jpg



Figure 4.1: Comparison of Similar Images Based on Global Metric

Chapter 5

Conclusion

This project demonstrates the effectiveness of color-based features for image similarity analysis. While histograms provide a broad sense of color distribution, dominant colors capture the main tones, and combining both provides a balanced similarity metric.