Multimedia Computing and Applications

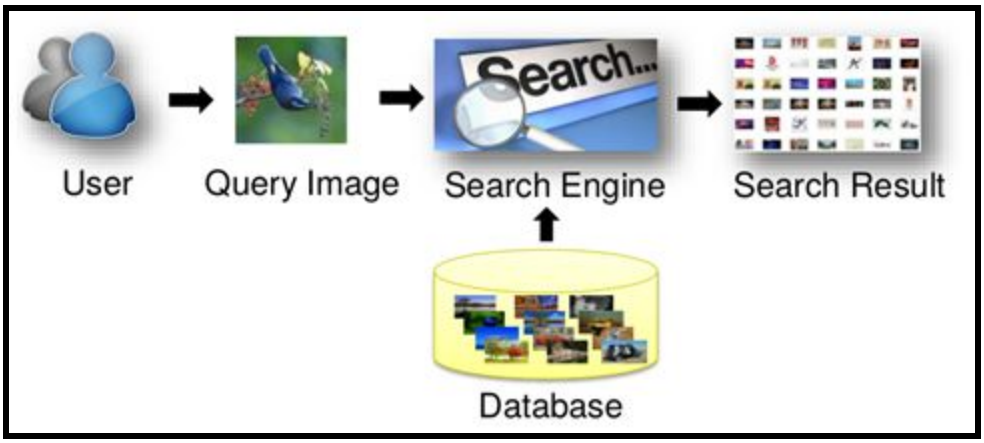
Assignment 1

Content Based Information Retrieval

Surabhi S Nath
2016271

Problem Statement

Image Search: Given a query image, rank images in database according to relevancy.



Dataset

The dataset provided contained:

- Database Images - 1050 images
- Query Images - 81 images
- Ground truth text file containing relevance information - relevance.txt

The images included day to day scenes, objects, animals, birds, persons, abstract art, sceneries, scenic locations, hoarders/sign boards etc. The images are of varying sizes and are in .jpg format.

In the relevance.txt file, relevant images were provided for each query image. Few queries has several relevant images (around 300), while other queries had lesser relevant images (around 50). On an average, every query had 160 relevant images from among the database images.

Approaches Used with Justification

Two approaches were used for performing the relevance based image search:

1. Colour Histogram Model
2. Bag of Visual Words Model

Colour Histogram

Colour histogram is a global image representation based feature. It is one of the easiest to calculate and visualize. The image pixel intensities are mapped to fixed number of predefined intensity bins and a histogram is generated for every image.

This was selected as the initial approach for a baseline understanding of the dataset and achievable accuracies.

Colour histogram presents the following advantages and drawbacks -

Advantages:

- Easy to calculate
- Purely based on pixel intensities
- Generates a fixed size vector

Drawbacks:

- Does not capture spatial information
- Does not capture local level features

OpenCV provides a ready implementation of Colour Histogram cv2.**calcHist** and Histogram Comparisons **cv2.compareHist**. The following steps were executed:

- The images were read in coloured form
- A 256-dimensional histogram was created was every image in the database
- Distance was calculated from every database image for every query image
- Nearest/best K database images were selected
- Precision @ K was evaluated by comparing with ground truth information

The following results for **Average Precision @ K** (average over all query images) were obtained:

K/Distance Metric	**Correlation**	**Chi Squared**	**Intersection**	**Bhattacharya**
K = 20	0.156	0.199	0.165	0.186
K = 15	0.163	0.207	0.174	0.192
K = 10	0.159	0.218	0.188	0.203
K = 5	0.167	0.246	0.224	0.235

Shortcomings:

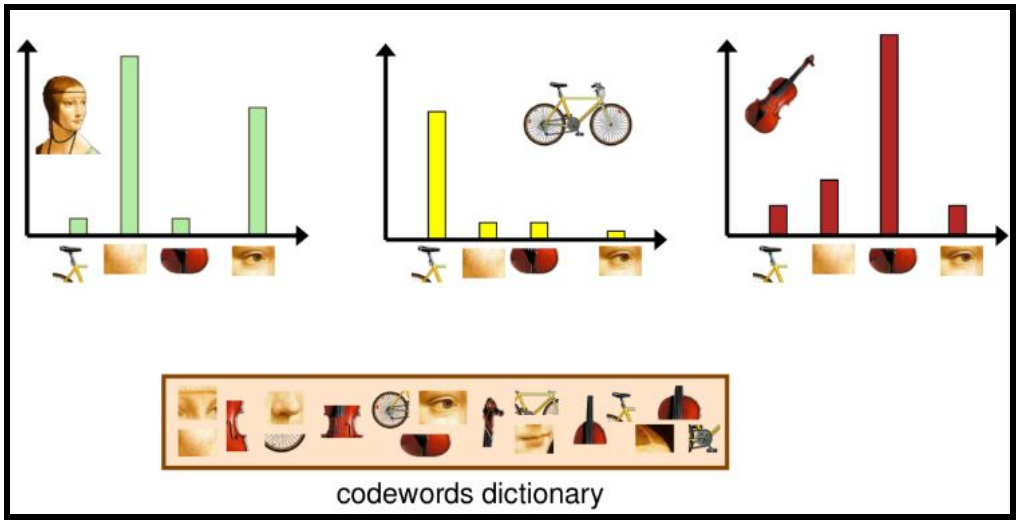
On studying the query image and the associated relevant database images, it was observed that the similarity is not based purely on colour but rather based on the content of the image. For example, the image on the right is considered most relevant to the left query image:

These images are related since they both have a lady in them. Thus the content is to be analyzed for arriving at the relevance score. This implies that features such as colour or texture may not be the most suitable and instead one must adopt local or visual features.

Bag of Visual Words

In an attempt to overcome the shortcomings of Colour Histogram model, Big of Visual Words was implemented next.

Bag of Visual Words is commonly used to find matching and similar images based on visual features. It a supervised technique wherein SIFT features are evaluated for each database image. Following, a mapping is generated between the feature and the object in the image. A vocabulary is generated and clustering is performed to group similar features. A reverse indexing is maintained. Now when a new query image appears, the visual features are decoded and then images having those features are concluded to be relevant to this query image.



The following steps were implemented:

- Database images were read, SIFT features were computed, clusters were formed and vocabulary was generated with dimension (1 x num_clusters)
- Vocabulary was next generated for query image
- Distance was calculated from every database image for every query image
- Distance metric was simply the sum of the absolute difference in corresponding indices
- Nearest/best K database images were selected
- Precision @ K was evaluated by comparing with ground truth information

The following results for **Average Precision @ K** (average over all query images) were obtained:

K/Num Clusters	100
K = 20	0.202
K = 15	0.221
K = 10	0.227

Shortcomings:

It is observed that despite using local and visual features, the relevance prediction is not very efficient. This could be either due to improper keypoint matching or unsuitability of SIFT features for the nature of data. Deep learning based solutions can perhaps yield more promising results in such a scenario.

Analysis Performed

Colour Histogram

Analysis was performed based on the following factors:

1. Number of Channels - The image was read in both coloured and grayscale and results were compared. It was found that coloured histograms lead to better results.

2. Distance metric - Various distance metrics were tested as shown in results above. These included:

    a. Correlation

$$d(H_1, H_2) = \frac{\sum_I (H_1(I) - \bar{H_1})(H_2(I) - \bar{H_2})}{\sqrt{\sum_I (H_1(I) - \bar{H_1})^2 \sum_I (H_2(I) - \bar{H_2})^2}}$$

    b. Chi Squared

$$d(H_1, H_2) = \sum_I \frac{(H_1(I) - H_2(I))^2}{H_1(I)}$$

    c. Intersection

$$d(H_1, H_2) = \sum_I \min(H_1(I), H_2(I))$$

    d. Bhattacharya

$$d(H_1, H_2) = \sqrt{1 - \frac{1}{\sqrt{\bar{H_1}\bar{H_2}N^2}} \sum_I \sqrt{H_1(I) \cdot H_2(I)}}$$

3. Vary K - Multiple possibilities of K value were tested and compared

Bag of Visual Words

Analysis was performed based on the following factors:

1. Vary KMeans iterations - Number of KMeans iterations were varied to obtain a balance between accuracy and time spent in training

2. Vary number of Clusters - Number of KMeans clusters was varied to incorporate a suitable number of visual features

3. Vary K - Multiple possibilities of K value were tested and compared

References

Code for Bag of Visual Words was adopted from github https://github.com/kushalvyas/Bag-of-Visual-Words-Python and modified as per requirement.