Visual Search of an Image Collection

Coursework - 1

Submitted in partial fulfillment of requirement
for the module of

Computer Vision & Pattern Recognition

Submitted by

MANISH PANDA(6614056)

Submitted to

PROF. Miroslaw Bober

University of Surrey

School of Computer Science and Electronic Engineering

Department of Electrical and Electronic Engineering

United Kingdom

November 2023

Abstract

This study uses a collection of 591 images to investigate the field of visual search. Our methodology is based on the concept of the 'query image,' which is a single image selected from the collection to serve as the reference for comparison. Three distinct distance measures allow for a rigorous descriptor computation process to be applied to every image in the collection, including the query image. This computation yields a unique collection of 591 descriptors for each distance measure, laying the groundwork for an extensive comparative analysis. The spatial grid systems and color histograms approaches form the basis of the visual search method. These methods must be applied in order to measure each image's quantifiable similarity to the query image inside the dataset. The result of this method is a ranking of the images based on how similar they are to the query image. This ranking illustrates the practical uses of visual search technology in the processing and analysis of big image collections, in addition to highlighting the relative efficacy of various distance measures and descriptor computation methods. Our goal with this work is to investigate and gain a deeper understanding of the complex dynamics of photo similarity and the effectiveness of different visual search techniques.

Contents

Chapter 1

Introduction

1.1 Visual Search

The process of searching through a collection of data for a certain image or visual item is known as visual search. The image that was chosen is known as the query image. Next, for every image in the data set, an image descriptor—a representation of the image—is created. The query image and each image descriptor in the data set are then compared, and the results are ordered based on how comparable they are determined by applying distance measurement methods, such as the Euclidean, Mahalanobis, and Cosine distances. Greater the rank, the higher the similarity quotient.

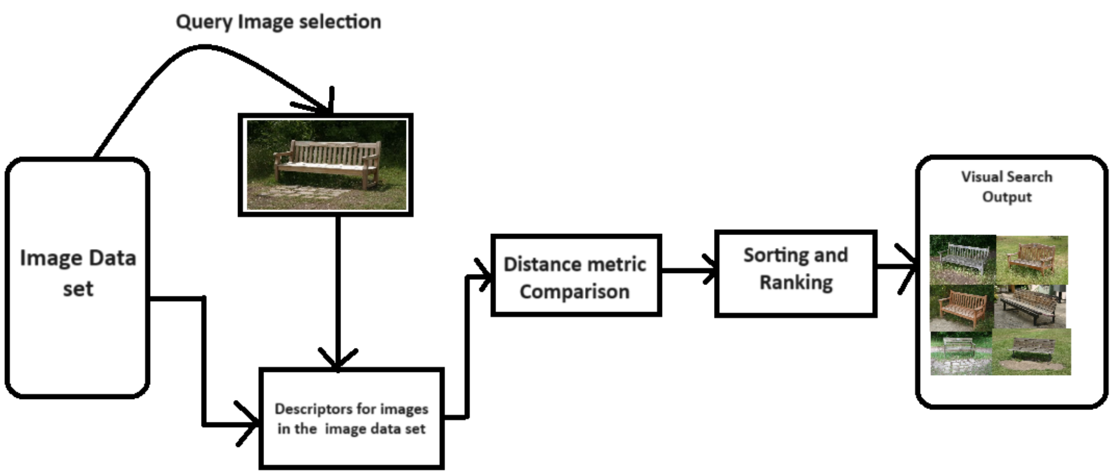


Figure 1.1: Visual Search Process.

Chapter 2

Image Descriptor

2.1 Global Colour Histogram

The distribution of colors in an image is generally represented by a global color histogram. RGB stands for red, green, and blue. These three colors combine to create a color image. By creating bins in the color space and counting the number of pixels that fall into each bin, the histogram may be calculated. RGB values are represented as points in three dimensions for each pixel. The space is then quantized in each of the three planes, and a standard histogram is produced. The values are then normalized by dividing the histogram's values by their sum, which makes it independent of the image dimensions.

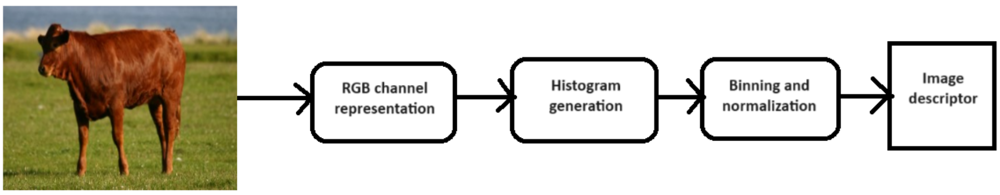


Figure 2.1: Global Colour Histogram.

2.2 Spatial Grids

2.2.1 Color Grids

Spatial color grids are used in computer vision and image processing to analyze the color distribution inside an image. By using this technique, the image is split up into several smaller yet identically sized grid cells. Within each cell, many color metrics are evaluated, such as the average color, color distribution, and conspicuous colors. Color quantization is often used in this, simplifying hues by combining them into larger sets. A comprehensive color descriptor is produced by combining the data from every grid cell to represent the spatial arrangement of colors in the image. These grids are exceptionally useful for tasks like object detection, content-based picture retrieval, and image recognition where the spatial distribution of colors is important. In photographs where color is important, such as in landscapes or artistic creations, the technique is useful since it captures the color arrangement of the image efficiently. But, in images where color isn't emphasized or isn't evenly distributed, its impact may be diminished. Spatial color grid effectiveness also depends on the grid size and color quantization level selected. There must be a balance between the two: too small a grid or color binning could be overly sensitive to slight color changes, while too coarse a grid or color binning could overlook important details.

2.2.2 Texture Grids

The image texture and color information can be used to create the descriptor. One important factor that makes this possible is the presence of edges in the image. Furthermore, these margins are formed in part by the image's uneven brightness levels. The process of identifying discontinuities in an image is known as edge detection in the world of image processing. One technique that can be used to identify edges in photos is the Sobel filter. Utilizing texture and color information in image descriptors allows for a more thorough examination of a picture's properties. The structure of the image can be understood by the texture, which consists of the pattern and uniformity of the pixel intensity. Even when two items or settings have identical color schemes, this is very helpful in differentiating between them. In order to show the boundaries between various items or locations, edges are important in images. These edges are frequently characterized by notable shifts in hue or brightness, which are sometimes made more noticeable by changes in lighting. When it comes to the form and arrangement of items in an image, the existence and direction of edges might reveal important details.

2.3 Sobel Filter

One popular edge detection technique in image processing is the Sobel filter. It works by convolutionally filtering the image through two 3x3 kernels, one of which calculates the image's vertical gradient and the other its horizontal gradient. The output of these convolutions is then combined to estimate the gradient's magnitude at each position in the image. This gradient measures the rate at which brightness changes and is used to identify edges. Each pixel in the directed image has a potential value ranging from 0 to 360 degrees. On this directed image, a grid segmentation technique is also used, much like in earlier methods. Each of these grids is subjected to angular quantization, which separates the angles into distinct bins.The histograms generated from each grid are then combined to provide the overall image description. Lower edge strength pixels are ignored in order to minimize noise in the descriptor. Because little directional changes can have a large effect on the descriptor, a lower threshold indicates a higher likelihood of errors in the data. Conversely, a high threshold may lead to complete disregard for the interesting thing.The accompanying image displays the mask along with its various threshold levels.

Chapter 3

Distance Estimation

3.1 Euclidean Distance

The length of the shortest path in euclidean space between two points is known as the euclidean distance. It is also referred to as L2 norm. It is often employed as a distance metric to measure how similar or different vectors that depict patterns or images are from one another. This distance measure may occasionally be used as a standard for dimensionality reduction methods like Principal Component Analysis (PCA). In mathematical notation, it is expressed as: [1]

$$\text{Euclidean Distance} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

3.2 Mahalanobis Distance

The distance between a point vector and a distribution can be calculated using the efficient multivariate distance metric known as the Mahalanobis distance. Since it converts the columns into uncorrelated variables, Mahalanobis distance differs computationally from Euclidean distance. After scaling each column to have a variance of one, the euclidean distance is finally estimated. For this reason, it's referred to as an L2 norm extension. The following formula yields the Mahalanaobis distance between a point and a distribution:

$$D_M = \sqrt{(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\mu})}$$

where, $\mathbf{x}$ is the vector of the point's coordinates.
$\boldsymbol{\mu}$ is the mean vector of the distribution.
$\mathbf{C}$ is the covariance matrix of the distribution.
$\mathbf{C}^{-1}$ is the inverse of the covariance matrix.

3.3 Cosine Distance

The cosine distance, which is the cosine angle between two vectors, is a measurement used to assess the similarity of vectors in an inner product space that are usually non-zero. The formula for calculating cosine similarity between two vectors is as follows: [1]

$$\text{Cosine Similarity}(\mathbf{A}, \mathbf{B}) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|}$$

where,

$$\mathbf{A} \cdot \mathbf{B} = a_1 \cdot b_1 + a_2 \cdot b_2 + \ldots + a_n \cdot b_n$$

$$\|\mathbf{A}\| = \sqrt{a_1^2 + a_2^2 + \ldots + a_n^2}$$

$$\|\mathbf{B}\| = \sqrt{b_1^2 + b_2^2 + \ldots + b_n^2}$$

Cosine distance is computed by subtracting cosine similarity from 1.

$$\text{Cosine Distance} = 1 - \text{Cosine Similarity}$$

The scale invariance and high dimensionality performance of the cosine distance measure are its primary advantages.

Chapter 4

Evaluation Approach

4.1 Precision

It serves as a barometer for the model's predictive accuracy. The ratio of all positive forecasts to actual positive forecasts is used to compute it. When a model predicts a positive occurrence, a high precision indicates that it is most likely correct. The precision formula is as follows: [1]

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

4.2 Recall

Recall is a measure of a model's capacity to capture all relevant instances of a positive class. Recall is calculated as the ratio of true positive predictions to the total number of real positive cases. A high recall number suggests that the model may be sensitive to good occurrences and that a sizable portion of positive examples may be caught by it. The Recall formula is as follows: [1]

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

4.3 Average Precision

In computer vision, average precision is a useful metric to assess the trade-off between accuracy and recall in models that produce ranked predictions. Frequently utilized in the assessment of object recognition algorithms, it provides a broader viewpoint on a model's effectiveness than do more basic measurements such as accuracy.

$$\text{AP} = \frac{\sum_{n=1}^{M} \text{P(n)} * \text{rel(n)}}{\text{Images present in the relevant class}}$$

4.4 Mean Average Precision

Metric Average Combining precision with recall yields precision, which is an important variable. Recall is how many true positives are correctly identified out of all true positives, whereas precision measures how many accurately recognized positive events there are to all positives the model detects. Two conditions for an algorithm are to minimize false positives (false identifications) and to accurately identify objects (true positives). This component of algorithmic performance is captured by a single metric that is provided by mAP.

$$\mathrm{mAP} = \frac{1}{C}\sum_{c=1}^{C} \mathrm{AP}_c$$

4.5 Precision-Recall Curve

It is a graphical illustration of the trade-off between precision and recall at different thresholds.

Chapter 5

PCA for dimensionality reduction

Principal component analysis, or PCA, is a widely used technique in computer vision to reduce dimensionality. It is particularly helpful when working with high-dimensional datasets since it transforms the data into a lower-dimensional space while maintaining as much of the original variability as is practical. In order to arrive at a lower-dimensional representation, principle component analysis (PCA) entails a series of procedures that include centering the data, eigen decomposition, computing the covariance matrix, choosing the principal components based on eigenvalues, and ultimately projecting the original data onto the selected principal components. This technique keeps the most important information while reducing the number of dimensions in the data.

The eigenvalues gave information about the level of variation in the data during the Principal Component Analysis (PCA) step of dimensionality reduction. More specifically, the first N eigenvectors are chosen so that the sum of their corresponding eigenvalues accounts for 99% of the variation in the dataset. This deliberate choice was made to condense the high-dimensional feature space into a more comprehensible form. Next, using the chosen eigenvectors, the original high-dimensional descriptor is projected into the eigen space. Because fewer dimensions are used in this projection, the data may be shown more succinctly and clearly while retaining much of its volatility.

Chapter 6

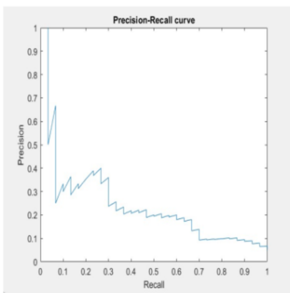
Observation

6.1 Global Colour Histogram

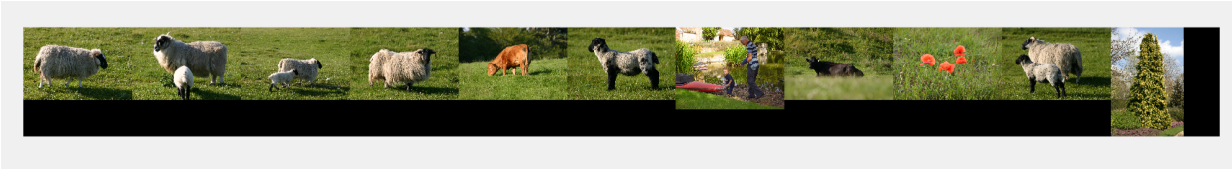
The inability to distinguish between images is made more difficult by the low resolution and small quantization factor, which result in decreased mean average precision values. It's also noticed that the color tones were similar to the image in the inquiry. The results of the histogram method are not reliable when objects have the same color but different shapes. Another drawback is that, because the descriptor is dependent on the overall content of the image, it will be impacted by image rotation or jigsaw puzzle-style rearranging. Depending on the particular issue being treated, this effect may or may not be advantageous.

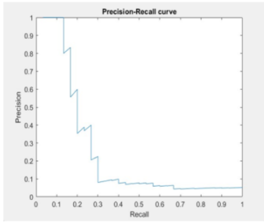
Observation 1:

```
Name of image 16_16_s.bmp
QueryImage : 187
Elapsed time is 0.005093 seconds.
Precision: 0.272727
Recall: 0.100000
Average Precision: 0.163612
```



Observation 2: Because these methods construct the image descriptor by focusing on the complete image, the majority of the photographs in observation 1 have brown hues, while the majority of the images in observation 2 have green and white colors.

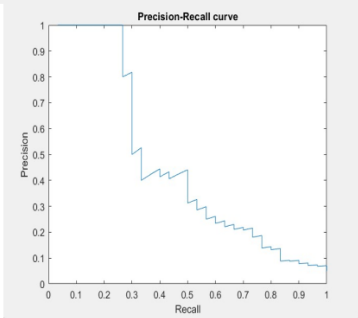





```
Name of image 9_6_s.bmp
QueryImage : 588
Elapsed time is 0.005042 seconds.
Precision: 0.545455
Recall: 0.200000
Average Precision: 0.259597
```

As a result, these color data are used to build the picture descriptors. Nonetheless, there are situations in which distinct objects have the same color, resulting in image descriptors that are similar. This descriptor similarity may lead to imprecise results when conducting visual searches.


```
Name of image 13_13_s.bmp
QueryImage : 100
Elapsed time is 0.004097 seconds.
Precision: 0.818182
Recall: 0.300000
Average Precision: 0.472603
```

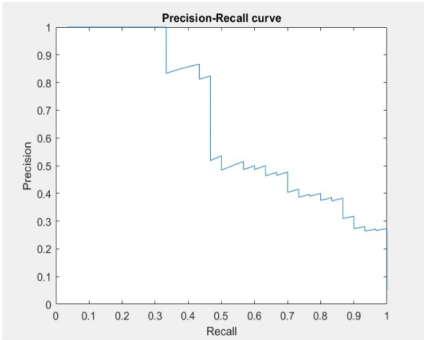


This last point relates only to the category of books, which does better than dogs, chairs, sheep, faces, birds, and cars, to name a few. The Global Color Histogram technique efficiently compares the color content of the query image with other book images when it comes to books. Two photos that are not from books are on the list because of their strong use of the colors red and black. The Global Color Histogram approach shows promise in some situations where particular color intensities are important consideration, even if it may not always be the best option.

6.2 Spatial Colour Grid

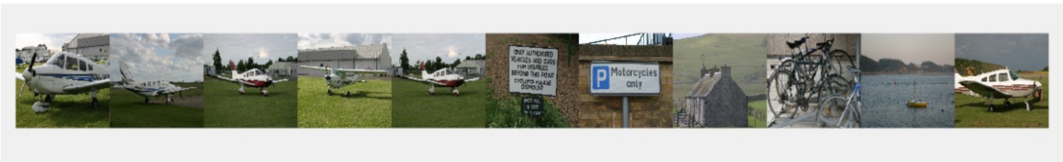
Observation 1: According to Observation 1, there are similarities between the images:

```
Name of image 4_17_s.bmp
QueryImage : 419
Elapsed time is 0.004418 seconds.
Precision: 0.909091
Recall: 0.333333
Average Precision: 0.667261
```

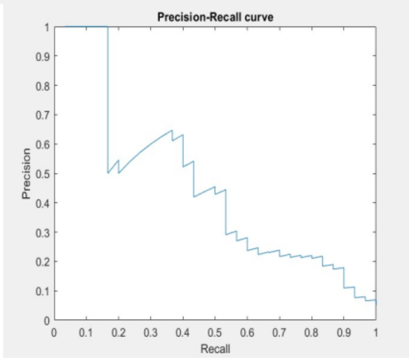


the white and blue elements from the sky and the plane, respectively, and the center of each image is centered on a single item, a plane. The ranking includes photographs of houses because of this similarity. One significant drawback is that descriptors are frequently assigned to photographs with the same color content. The cosine distance metric is the one used in this instance.

Observation 2: In the second observation, it is especially clear in the plane category



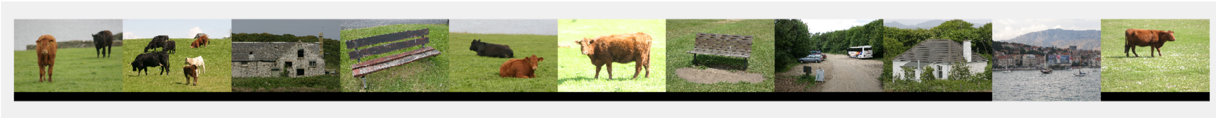

```
Name of image 4_17_s.bmp
QueryImage : 419
Elapsed time is 0.012729 seconds.
Precision: 0.545455
Recall: 0.200000
Average Precision: 0.461738
```



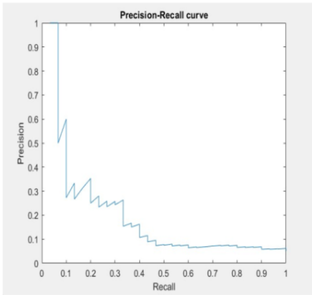
that the utilization of the Mahalanobis distance metric improves the visual search's accuracy. The reemergence of the image of the house was explained by the search taking into account colors like brown, green, blue, and white. These findings clearly show that the Mahalanobis distance metric performs better than the cosine distance metric. Moreover, it is observed that better accurate results are obtained when an image descriptor is created utilizing color data in a 3 x 3 grid.

6.3 Spatial Texture Grid

Observation 1:



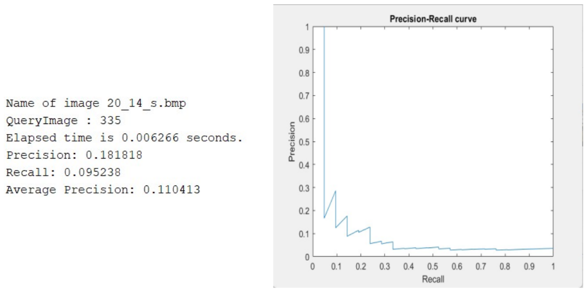
```
Name of image 1_21_s.bmp
QueryImage : 313
Elapsed time is 0.004438 seconds.
Precision: 0.272727
Recall: 0.100000
Average Precision: 0.210738
```



While the color content of each image varies, it has been observed that they all have one thing in common: a horizontal line that is centered in every picture. Furthermore, the query image has a high degree of textural complexity, which caused photographs with distinct edge features to be included in the list. Certain photos have moderate edge features, but the high entropy effect is clearly visible because of noise.In these pictures, the texture is more vivid at the top and less noticeable at the bottom. It is feasible to identify which photographs in the list have a smoother lower half and a more textured top half by using grids and texture analysis.The image of the cow, where the bottom half is fuzzy and has a smoother texture, is a good illustration of this.

Observation 2: Similar conclusions to observation 1 are found in this observation

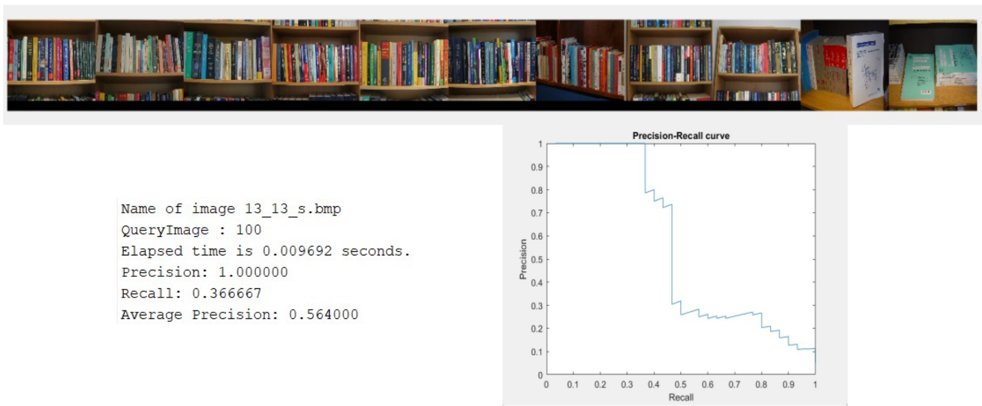
as well, especially with relation to color content and edge information. Because of the query image's abundance of textural characteristics, pictures with distinct edge details were chosen for the list. Certain photos show reasonably strong edge features, however noise causes a notable high entropy effect in these images.



6.4 Spatial Colour and Texture Grid

While the spatial texture grid highlights edge details, the spatial color grid concentrates on color data. An image descriptor that combines color and edge information will probably produce a greater accuracy rate for improved visual search performance.

Observation 1: According to Observation 1, the images are distinguished by dis-



tinctive, variable horizontal or vertical lines, like to those seen in books. Because of their noticeable vertical margins, most of the photos are recognized and sorted correctly. When an image is ranked incorrectly, it is observed that although the color content is somewhat recorded, the edge information is not sufficiently represented. In addition, the technique used in this analysis is the cosine distance measure.

Observation 2: Regarding the second observation, the pictures always have high entropy backgrounds and features around the horizontal edges. It is clear that each image's center is primarily white, green and brown. In this instance, the Mahalanobis distance metric is applied.

Taking into account the findings from both investigations, it is evident that images with distinct and contrasting horizontal or vertical lines work best when using the spatial color and texture method. Furthermore, compared to the Mahalanobis distance metric, the cosine distance metric has been found to perform better.

Figure 6.1: Observation 2

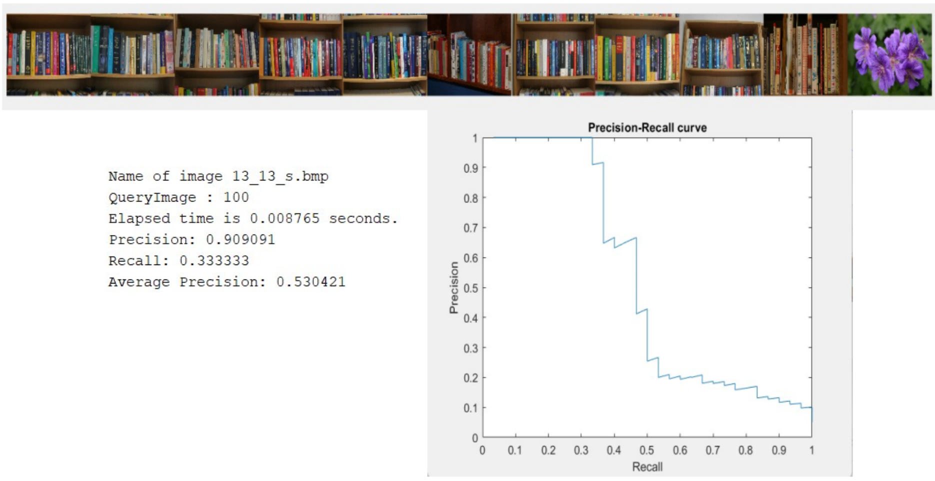


Figure 6.2: Observation 3

Observation 3 : To measure differences in color and texture properties, the spatial color and texture grid approach uses Euclidean distance, which has advantages. Nevertheless, this metric's effectiveness depends on selecting the right features, determining the grid's size, and taking into account the spatial arrangement of those elements. In case of books class, euclidean and cosine distance metric worked very well.

Chapter 7

Conclusion

Figure 7.1: Distance metric comparison

After a mean average precision study was carried out on twenty random datasets with three distinct distance metrics, it was observed that the cosine distance and Euclidean distance (also referred to as L2 norm) were the most efficient metrics. In general, the L2 norm and cosine distance outperformed the Mahalanobis distance, albeit it did well on some datasets, such as ones involving sheep. The Mahalanobis distance still has value and can be very helpful in some situations, thus this observation does not necessarily mean that these two metrics will always be the best option for all datasets.

Owing to its comprehensive assessment of both color and texture features, a spatial color and texture grid usually produces better results in visual searches. Still, the ideal approach varies according to the features of the photos and the specific requirements of the search. A spatial color grid or global color histogram, for instance, might be sufficient when color is the only differentiating factor. Yet, a combined strategy employing a spatial color and texture grid is typically more beneficial in more complex searches when both color and texture are important elements in distinguishing images.

References

[1] A. Jeya Christy,K. Dhanalakshmi, Research of Global Feature of 3D Color Histogram with Similarity and Dissimilarity Metrics for Image Processing. IJITEE ISSN: 2278-3075, Volume-9 Issue-3S, January 2020.

[2] Qu, Yongming & Ostrouchov, George & Samatova, Nagiza & Geist, Al. (2002). Principal Component Analysis for Dimension Reduction in Massive Distributed Data Sets. Knowledge and Information Systems - KAIS

[3] Wang X Y, Wu J F, Yang H Y. Robust image retrieval based on color histogram of local feature regions[M]. Kluwer Academic Publishers, 2010