Visual Search of an Image Database

Advaith Vellanki

First Semester Coursework for the Master of Science in Artificial Intelligence from the University of Surrey

Department of Electronic Engineering
Computer Vision and Pattern Recognition
EEE3032

Autumn 2023

Contents:

1 Abstract

This coursework consisted of performing Visual Search of an Image from the Microsoft Research MSRCv2 Dataset using Content Based Image Retrieval (CBIR). The accuracy of various descriptors were experimented on and assessed to gauge how it affected the results. First the simple Global Colour Histogram descriptor was computed, and visual search was executed using the Euclidean Distance followed by the computation of a Colour Moments descriptor. Then, the more complex Spatially Gridded Descriptors were computed and the results were compared and evaluated using Precision-Recall metrics.
This was followed by the implementation of Principal Component Analysis (PCA) on the descriptors to reduce their dimensionality and check how it affected the search process. After PCA, Mahalanobis Distance was calculated and its accuracy was evaluated.
Finally, additional distance measures were calculated such as the L1 or Manhattan Distance, Cosine Similarity and Pearson's Correlation Coefficient to see if they led to any improvements in performance and the experiments were documented.

2 Introduction

One of the larger applications of Computer Vision involves visual search, which is the ability to deduce patterns or features in an image and compare it to other images to measure the similarity. This is used in several areas such as but not limited to finger-print security systems, facial recognition systems on mobile phones and cancer detection during medical examinations.
Visual search can be performed by creating descriptors of the images which store features using colour and edge/shape information and then analysing the similarity between them to determine whether the images they represent are the same or not.

3 Description of Visual Search Techniques Implemented

Visual search of the query image was carried out using a total of 5 different descriptors along with 5 different distance measures which produced results of varying accuracy. A detailed description of the descriptors and distance measures used is given below.

3.1 Descriptors

3.1.1 Global Colour Histogram

A global colour histogram is a representation of the distribution of colours in an image. The global colour histogram is created by quantizing the RGB space and then building a histogram based on the quantized values.
Quantization is the process by which the continuous colour space is categorized into a number of bins. This quantization is applied to each channel of the image i.e. Red, Green and Blue. This is done to reduce the complexity of the colour space. It is done by dividing each pixel value by 256 to normalize it to the range [0,1] and then multiplying it by a 'Quantization Factor' which is then floored to obtain integer values in the range [0,Q-1].
Then, the descretized RGB values are combined to create a single integer value for each pixel. This is achieved by summing the product of each RGB value with the corresponding power of Q. The resulting single integer value is considered as the bin index for the histogram. This bin is now a 2D image where each pixel contains an integer value. This bin is reshaped into a long vector for further processing.
After quantization, a histogram is calculated using Matlab's 'hist' command. The 'hist' command requires a 1D array of values to create the histogram which is why the bin was reshaped [1]. This histogram has Q^3 bins and represents the frequency distribution of the bins.
Finally, the histogram is normalised to ensure that the sum of all frequencies is equal to 1. This makes the histogram independent of the size of the image and facilitates easier comparison between the images.

After calculating the Global Colour Histogram descriptor, the similarity measure is computed using the Euclidean distance. The measured Euclidean distance of the query image is compared with the database images results in order to bring out the similar images.

3.1.2 Colour Moments Descriptor

The colour moments descriptor first involves loading an image and separating it into its RGB colour channels. We then calculate the three colour moments which are the mean, standard deviation and skewness for each colour channel. This implies that an image is characterized by 9 moments in total – 3 moments for each of the 3 colour channels. [2]
The mean which represents the average colour intensity of an image is classified as the First Order Moment.
The standard deviation which indicates the variation or dispersion in colour intensities is classified as the Second Order Moment.
The skewness which defines the degree of asymmetry in the distribution is classified as the Third Order Moment.
Finally, after calculating all of these moments, we create a vector to store them. This vector contains the colour moments descriptor for the image.

The Colour Moments descriptor was then used to execute visual search of the query image using Euclidean Distance as the distance measure.

3.1.3 Spatially Gridded Colour Histogram

A spatially gridded colour histogram descriptor involves a feature extraction technique that first divides an image into a grid structure and then computes colour related information for each cell. The main goal is to capture the distribution of colours in different regions of the image.

The spatially gridded colour histogram computed in this coursework systematically divides the image into a grid specified by the GRIDFACTOR, computes a global RGB histogram for each grid, and concatenates them into a single feature vector. Each element of the vector represents a certain aspect of colour distribution in a specific region. The code for computing the descriptor has been made efficient by calling the function to compute the global RGB histogram mentioned above but instead of sending a normal image as an argument, it sends the sub image corresponding to a specific grid to compute the global RGB histogram for it.
The grid size is flexible and can be adjusted to capture details at different scales which leads to improved matching. A result of dividing the image into grids is that the descriptor becomes more discriminative. Spatially localised histograms can better capture details which leads to improved accuracy due to the capturing of variations in colour within different parts of an image.
Since the spatial arrangements of colour is significant for visual search in the Microsoft Research MSRCv2 Database, the Spatially Gridded Colour Histogram should, in theory, work well.

3.1.4 Spatially Gridded Edge Orientation Histogram

A Spatially gridded edge orientation histogram is a representation of the distribution of edge orientations in different regions of an image. The goal is capture information about the directions of edges within localised grids, which provides a more detailed description of the image's structural features.
The Spatially Gridded Edge Orientation Histogram in this coursework is obtained by first performing spatial gridding on an image and then using it to compute the Edge Orientation Histogram. This is performed by first converting the input RGB grid to grayscale. Then, Sobel filters are applied in both horizontal and vertical directions to compute the image gradients.
It is worth noting that the Sobel Operator performs a 2D spatial gradient measurement on an image and so emphasizes regions of high spatial frequency that correspond to edges.[3]
After the image gradients have been computed, their magnitude is calculated to obtain edge strength. This is followed by the subsequent normalisation of the edge magnitude values to the range [0,1]. The pixels with high normalised magnitude (greater than 0.20) are considered to be significant edges. The orientations of these significant edges are retrieved and normalised. Then, the normalised orientations are quantised into 'Q' bins which makes this histogram robust to small variations in edge orientations. Finally, these quantised edge orientations are used to create a histogram which is normalised to form the Spatially Gridded Edge Orientation Histogram.
The use of histograms helps in summarising the predominant orientation of edges in the image and quantisation parameter (Q) controls the granularity of the histogram, affecting its sensitivity to small changes in edge orientation.
This descriptor captures detailed information about the edge orientations in different regions of the image and provides a structural description of how edges are distributed across various parts of the scene. It is particularly beneficial in visual search tasks where understanding the distribution of edges in localised regions contributes to better accuracy.

3.1.5 Spatially Gridded Colour and Edge Orientation Histogram

This descriptor involves dividing an image into a grid of cells, computing the global colour histogram and the edge orientation histogram of those cells and finally concatenating both the histograms to form a Spatially Gridded Colour and Edge Orientation Histogram. In this coursework, to compute this descriptor we first compute both the Spatially Gridded Colour Histogram and the Spatially Gridded Edge Orientation Histogram and finally concatenate them both. The resulting histogram is then normalised to ensure that it is invariant to changes in image size and intensity. The final descriptor is hence a feature vector that encodes the colour and edge orientation information in a spatially gridded manner.
This descriptor is particularly useful in scenarios where different regions of an image might have distinct colour or texture characteristics. It allows for a more localised analysis of both colour and edge features and hence is an efficient descriptor to be used for visual search.

3.2 Distance Measures

After the computing of all of the descriptors, Visual search is executed by measuring the distance between the query image and the rest of the images in the database. There are several ways to measure this distance and ones used in this coursework are explained below.

3.2.1 L2 Norm or Euclidean Distance

Euclidean or L2 Norm Distance is a measure of the straight line or shortest distance between two points in space. This distance metric is derived from the Pythagorean theorem and measures the 'as a crow flies' distance between two points.
For two points (x_1 , y_1 , . . . z_1) and (x_2, y_2, . . . z_2) in an n-dimensional space, the Euclidean Distance *d* is calculated using the formula

$$d = \sqrt{}((x_2 - x_1)^2 + (y_2 - y_1)^2 + . . . + (z_2 - z_1)^2)$$

3.2.2 L1 Norm or Manhattan Distance

Manhattan or L1 Norm Distance is the sum of the absolute differences between the coordinates of corresponding units. For two points (x_1 , y_1 , . . . z_1) and (x_2, y_2, . . . z_2) in an n-dimensional space, the Euclidean Distance *d* is calculated using the formula

$$d = |x_2 - x_1| + | y_2 - y_1| + . . . + | z_2 - z_1|$$

3.2.3 Mahalanobis Distance

Mahalanobis distance is a measure of the distance between a point and a distribution. It takes into account the correlations between variables. It also considers the scale and orientation of the data. For a point *x* in a multivariate distribution with mean μ and covariance matrix Σ , the Mahalanobis distance is calculated as

$$D_M (x) = \sqrt{}((x - \mu)^T \Sigma^{-1}(x - \mu))$$

Where,
X is the vector representing the point,
μ is the mean vector of the distribution,
Σ is the covariance matrix of the distribution and,
Σ^{-1} is the inverse of the covariance matrix.

Mahalanobis distance is used to calculate the distance when Principal Component Analysis has been implemented on the descriptors. Therefore, when the covariance matrix Σ undergoes eigen decomposition, the Eigen Vectors U and the Eigen Values V can be used to calculate Mahalanobis distance with

$$D_M (x) = \sqrt{}|V^{-1} U^{-1} (x - \mu)|$$

3.2.4 Pearson's Correlation Coefficient

This distance measure measures the Linear correlation between two variables. When used as a distance measure it quantifies how much two sets of data deviate from a perfect linear relationship. In the context of visual search, it can be used to compare the similarity between two feature vectors. The coefficient *r* is calculated as

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}}$$

Where,
The numerator is the covariance of the two vectors and,
The denominator is the square root of the product of their standard deviations.
From the Pearson Correlation Coefficient *r,* the distance is calculated as *d = 1 – r* .

3.2.5 Cosine Similarity

Cosine Similarity calculates the cosine of the angle between two non-zero vectors. The formula for cosine similarity between two vectors A and B is given by:

$$Cos(A,B) = \frac{A \cdot B}{||A||\,||B||}$$

Where,
A . B is the dot product of the vectors A and B,
||A|| is the magnitude or Euclidean Norm of the Vector A and,
||B|| is the magnitude or Euclidean Norm of the Vector B.
From the Cosine Similarity, the distance is calculated by *d = 1 – Cos(A,B).*

3.3 Principal Component Analysis (PCA)

In order to interpret reasonably large datasets, methods are required to drastically reduce their dimensionality in an interpretable way, such that most of the information in the data is preserved [4]. PCA is one such technique that is widely used in the field of computer vision and involves reducing the dimensionality of the dataset, while preserving as much variability in the data.
PCA begins by computing the covariance matrix of the original dataset which represents the relationship between the different features in the dataset.
This is followed by Eigen Decomposition which involves finding the eigenvectors and eigenvalues of the covariance matrix. The eigenvectors represent the principal components in which the data varies the most, and the eigenvalues indicate the magnitude of the variance in these principal components.
The next step involves sorting the eigenvectors based on their corresponding eigenvalues. The eigenvectors with the highest eigenvalues (Principal Components) capture the most variance in the data.
Finally, the dataset is projected onto the subspace followed by the selected principal components which results in a lower-dimensional representation of the data.

In this coursework code, these steps have been implemented clearly. First the eigenvectors and eigenvalues are computed for each pre-computed image descriptor representing the images of the database and are simultaneously sorted in the Eigen_Build function. Then the third parameter of the Eigen_Deflate function decides the dimensions to which you want to reduce each of the descriptors. Finally, the projection of the dataset onto the subspace is carried out by the Eigen_Project function.

PCA is useful in filtering out noise as the first few principal components often capture the dominant patterns in the data while later components may represent noise. Removing these later components can lead to denoised data. It is also used to address the curse of dimensionality which states that greater the dimensions of the data, more the number of points that are needed to form meaningful patterns or relationships.

In this coursework, after performing PCA, the dimensionally reduced descriptors were stored in different folders in order to make comparisons of results easier. Mahalanobis distance was used as an additional distance measure after PCA during visual search to check if performance and accuracy increased. The experiments performed relating to this are documented further into this report.

4 Test Method

4.1 Precision and Recall

Precision and recall are two key metrics used to evaluate a visual search system. They are particularly relevant in scenarios where the goal is to retrieve relevant items from a larger set of items as is the case in this coursework.

4.1.1 Precision
Precision is the ratio of relevant instances retrieved by the system to the total instances retrieved. It is used to measure the accuracy of the system when it claims to have found something. The formula for precision is

Precision = (Number of true positives) / (Number of true positives + Number of false positives)

A higher precision indicates that the majority of the retrieved images are relevant. An ideal system response would be a precision of 1.

4.1.2 Recall
Recall is the ratio of relevant instances retrieved by the system to the total amount of relevant instances in the dataset. It measures the system's ability to find all the relevant instances. The formula for recall is

Recall = (Number of true positives) / (Number of true positives) + (Number of False Negatives)

A high recall indicates that the system can find a large proportion of the relevant items.

Hence, Precision and Recall are often in tension with each other. Increasing precision may decrease recall and vice versa.
In the context of visual search, precision and recall help to gauge the system's effectiveness in returning relevant images while minimising false positives or negatives.
In this coursework, Precision and Recall were computed based on distances and file paths. The class information from the file path is extracted and it considers the information present in the file name before the first underscore as the reference class. Then, the true positives are assumed to be the instances belonging to the reference class.

4.2 Precision – Recall Curve (PR Curve)
A precision-recall curve is a graphical representation that can be used to evaluate the performance of a classification or information retrieval system. The X-Axis represents the recall which in this case is a percentage of actual positive instances that were retrieved. The Y-Axis represents the Precision which is the percentage of retrieved instances that are actually positive.
The top-left corner of the P-R Curve indicates an ideal scenario where the system retrieves many positive instances and those retrieved instances are highly relevant.
The bottom-right corner of the P-R Curve indicates a scenario where the system retrieves few instances and those retrieved instances are not very relevant.
The curve shape illustrates the trade-off between precision and recall. A large area under the P-R Curve generally indicates better performance.

5 Experimentation

In all of the experiments, the results are assessed using Average Precision and Average Recall along with the Precision – Recall Curve to provide a detailed analysis of the accuracy of the descriptors. It is worth noting that the first image in the results is always the query image and that is why the Precision value in the Precision – Recall Curve always begins from 1.
In the following experiments, Euclidean Distance or/and Manhattan Distance are used at first followed by a separate section consisting of the other distance measures used on the same query images.
The experimentation for the more complex descriptors was done by performing hyperparameter tuning, by changing the levels of quantization along with the grid factors.
There is a conclusion section included after each experiment which analyses the results to obtain deductions.

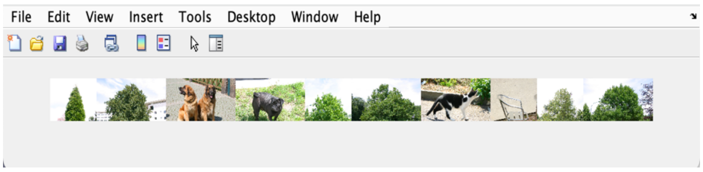
It is important to note that the figures below each Visual Search experiment are the (1) The top 10 most relevant results, (2) The Precision – Recall Curve for the results obtained and (3) The Average precision and Recall values of the Visual Search.

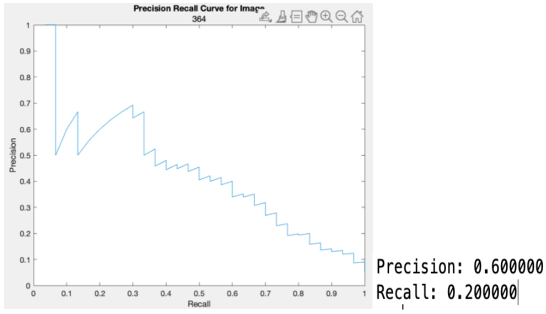
5.1 Global Colour Histogram

In this experiment, we are using the Global Colour Histogram as the descriptor and Euclidean Distance as the Distance Measure. We select the query image in prior in order to get an accurate comparison of the results with different levels of RGB quantization. We will evaluate the results using Precision – Recall and the PR Curve.

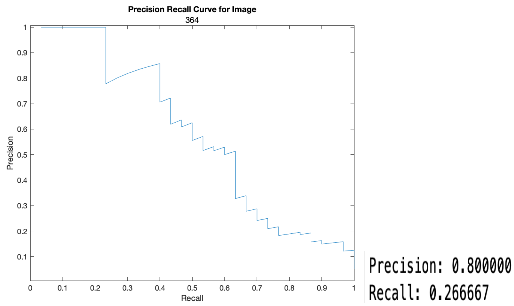
Query Image: 364

Visual Search Results with Quantization = 4 with Euclidean Distance

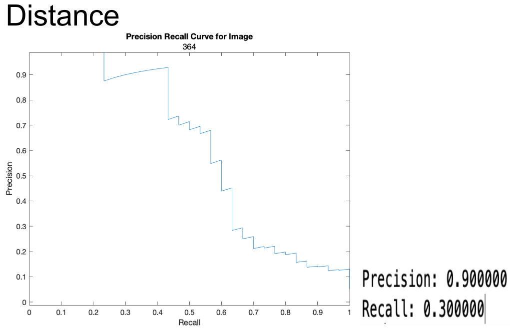


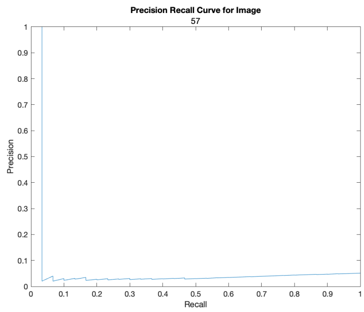


Visual Search Results with Quantization = 5 with Euclidean Distance



Visual Search Results with Quantization = 7 with Euclidean Distance





Precision: 0.100000
Recall: 0.033333

5.1.1 Conclusion from the experiment

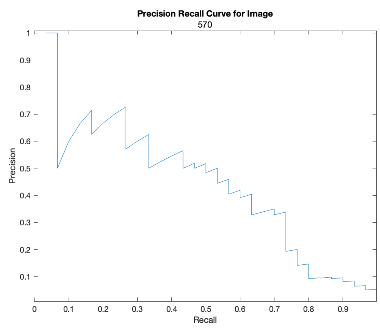
Here, It can be inferred that for a Global Colour Histogram, as the quantization increases, the accuracy of the search increases as well, with the system giving 90% precision with quantization = 7.

It was also observed that the global colour histogram worked very efficiently in retrieving images having the same distributions of colour globally but poorly when the object was the same but the colour distributions were different. This is noted in the first 3 experiments as well and the results are great because the colour patterns of all images are the same (they include blue around the top of the image and green in the centre). This descriptor gave great results for some classes of images such as farm animals and trees but gave unsatisfactory results for other classes of images where colour distribution was not an important factor such as bicycles and sign boards as shown in the final visual search experiment in this subsection.

5.2 Colour Moments Descriptor

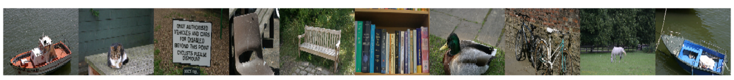
In this experiment we are using Colour Moments as the descriptor and Euclidean Distance as the distance measure at first. We will be experimenting with different query images and noting observations. We will evaluate the results using Precision – Recall and the PR Curve.

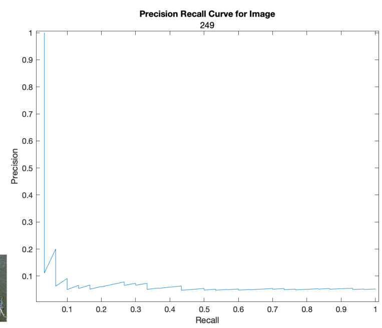
Visual Search Results 1 with Euclidean Distance



Precision: 0.700000
Recall: 0.233333

Visual Search Results 2 with Euclidean Distance





Precision: 0.200000
Recall: 0.066667

5.2.1 Conclusion from the experiment

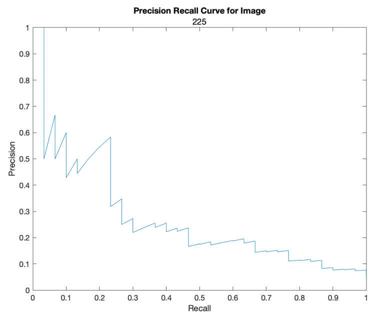
From the above two experiments, it is clear that the colour moments descriptor works well when colour distribution is a significant factor in describing the content. They are particularly useful in scenes with diverse and vibrant colours such as landscapes or objects with rich colour variations such as a sheep standing in a green field. However, this descriptor performs poorly when it comes to scenes having limited colour diversity and intricate texture details such as a boat in water or a house with complex textures.

5.3 Spatially Gridded Colour Histogram

In this experiment, we are using the Spatially Gridded Colour Histogram which in theory should work better than the previous two descriptors. This descriptor has a quantization value as well as a Grid Factor (G) where G x G= number of grids. The distance measure used in this experiment is the Euclidean Distance and Manhattan Distance. The query image for this experiment was pre-decided. We will evaluate the results using Precision – Recall and the PR Curve.

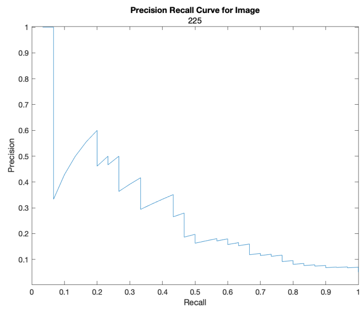
Query Image: 225. It consists of a road and buildings.

Visual Search Results when Quantization = 4, Grid Factor = 4 with Euclidean Distance



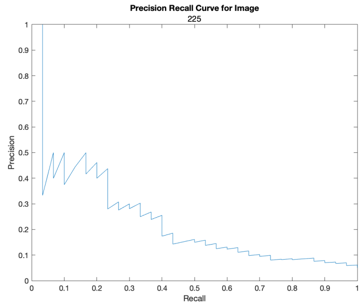

```
Precision: 0.500000
Recall: 0.166667
```

Visual Search Results when Quantization =6, Grid Factor =4 with Euclidean Distance



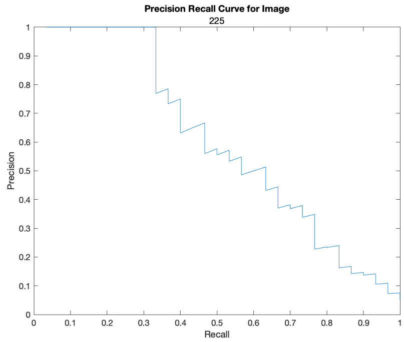

```
Precision: 0.600000
Recall: 0.200000
```

Visual Search Results when Quantization = 6, Grid Factor =5 with Euclidean Distance




```
Precision: 0.500000
Recall: 0.166667
```

Visual Search Results when quantization = 6 and grid factor = 4 with Manhattan Distance




```
Precision: 1.000000
Recall: 0.333333
```

5.3.1 Conclusion from the experiment

From the above experiments, it is seen that the results of this descriptor works well for images which have distinctive regions with different colour characteristics, for example, buildings and a road in the centre. The ideal quantization was deduced to be 6 and further increasing of quantization led to poorer accuracy. It can be seen from above that the ideal Grid Factor was 4 and that increasing it beyond 4 led to a loss in accuracy.
It was also seen that when Manhattan Distance was used instead of Euclidean Distance, the results had a precision of 1.00 and were ideal.
Spatially Gridded Colour histograms do have a weakness when it comes to images having low spatial colour variations or homogenous colour distributions such as a simple background.

5.4 Spatially Gridded Edge Orientation Histogram

In this experiment we are using the Spatially Gridded Edge Orientation Histogram which has a quantization value as well as a grid factor. The distance measure used in this experiment was Euclidean Distance. The query image for the experiment was pre-decided. We will evaluate the results using Precision – Recall and the PR Curve.

Query Image:

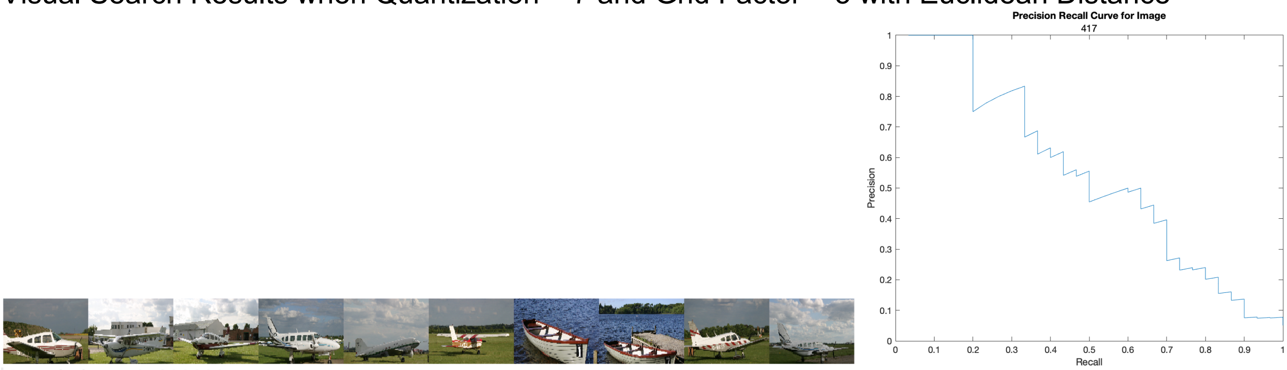
Visual Search Results when Quantization = 4 and Grid Factor =4 with Euclidean Distance

Precision: 0.400000
Recall: 0.133333

Visual Search Results when Quantization = 5 and Grid Factor =4 with Euclidean Distance

Precision: 0.700000
Recall: 0.233333

Visual Search Results when Quantization = 7 and Grid Factor = 5 with Euclidean Distance



Precision: 0.800000
Recall: 0.266667

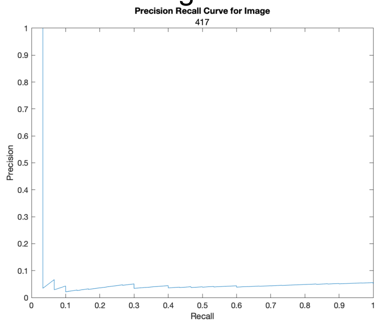
5.4.1 Conclusion from the experiment

From the above experiments it is evident that Spatially Gridded Edge Orientation Histograms are impressive at detecting objects having clear edges in an image. It is also seen that the initial Precision of 0.400 jumps to an impressive 0.700 when we increase the quantization. The planes are recognised even though there are edge variations in them. This is due to the spatial gridding which increases the discrimination of such edges. Upon further increasing the quantization to 7 and increasing the Grid Factor to 5, it is seen that the already accurate precision of 0.700 increases to 0.800 and that the visual search is less affected by changes in global image transformations.

5.5 Spatially Gridded Colour and Edge Orientation Histogram

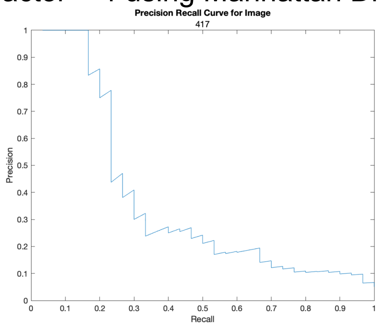
In this experiment we use the Spatially Gridded Colour and Edge Orientation Histogram which is a combination of the previous two descriptors. Since the previous two descriptors were already performing well, this descriptor should, in theory, outperform all of the above mentioned descriptors as it is gridded and considers both colour and edges in an image for a more thorough analysis. The distance measures used in this experiment are Euclidean Distance and L1 Norm or Manhattan Distance. The query image was pre-decided to be the same as the one for the previous experiment to show the efficiency of this descriptor when used with L1 Norm Distance. We will evaluate the results using Precision – Recall and the PR Curve.

Visual Search Results when quantization = 4 and grid factor = 4 using Euclidean Distance



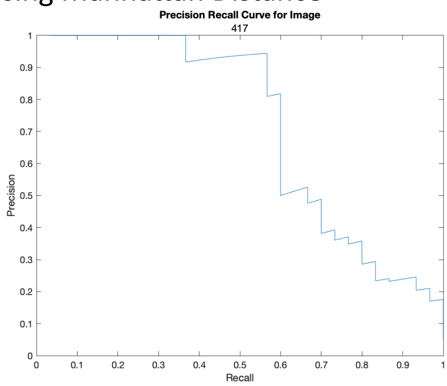
Precision: 0.100000
Recall: 0.033333

Visual Search Results when quantization = 4 and grid factor = 4 using Manhattan Distance



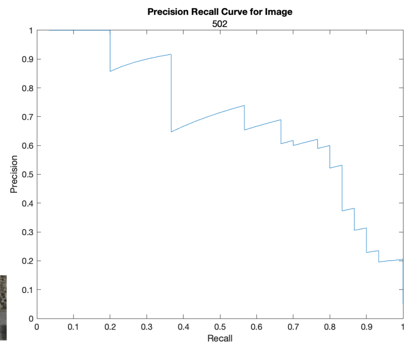
Precision: 0.700000
Recall: 0.233333

Visual Search Results when quantization = 5 and grid factor = 5 using Manhattan Distance



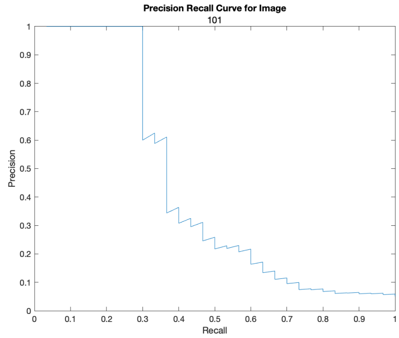
Precision: 1.000000
Recall: 0.333333

Visual Search Results with a different image when quantization = 7 and grid factor = 4 with Manhattan Distance



```
Precision: 0.900000
Recall: 0.300000
```

Visual Search Results with a different image when quantization = 6 and grid factor = 4 with Manhattan Distance




```
Precision: 0.900000
Recall: 0.300000
```

5.5.1 Conclusion from the experiment

From the above results it is observed that this descriptor works very poorly while using Euclidean Distance with quantization = 4. However, if you use Manhattan Distance the precision increases drastically to 0.700 with the same level of quantization. Finally if you change the quantization to 5 and the grid factor to 5 it is observed that the visual search with Manhattan distance returns with precision = 1.00. This is an ideal result. Further experiments were carried out with different query images and it was deduced that this descriptor outperforms the previous ones when used with Manhattan Distance as the distance measure.

The penultimate experiment i.e. the visual search with the car as the query image, shows that this descriptor is robust enough to work accurately even when the object undergoes rotation or if the object belongs to the same class but is of a different shape or colour. The one inaccurate image in the said experiment can be due to the fact that the row of cycles look similar to the shape of a car.

5.6 Principal Component Analysis (PCA)

From the above experiments, it was deduced that the Spatially Gridded Colour and Edge Orientation Histogram performed accurately most of the times. In order to test out Principal Component Analysis, this descriptor is used as the subject for PCA with the aim of reducing the dimensionality of each of the image descriptors computed for the images in the database. We are taking a set quantization of 6 and Grid factor = 4 for the descriptor which leads to its dimensions being 1 x 3552. We will perform PCA and reduce the dimensions multiple times and use the Mahalanobis Distance as the distance measure to perform visual search. As in the previous experiments, the performance will be evaluated using Precision – Recall and the PR Curve. The query image will be the final used for the previous experiment i.e. The Bookshelf, in order to compare the results to the results of the descriptor before PCA.

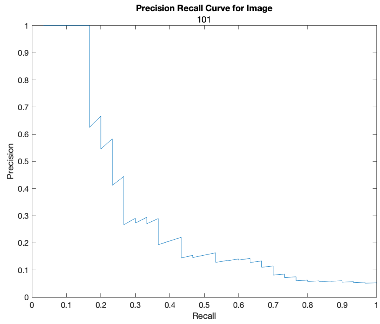
First the original dimensions will be displayed

Name ▲	Size	Bytes	Class
F	1x3552	28416	double

Visual Search with the dimensions reduced to 1 x 1400 using Mahalanobis Distance

Reduced Dimensions:

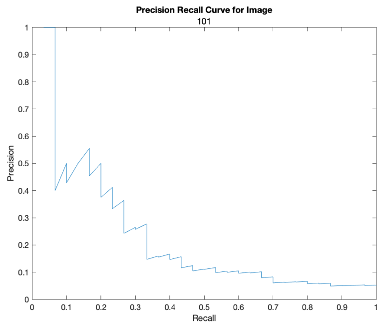
Name ▲	Size	Bytes	Class
F	1x1400	11200	double



Precision: 0.600000
Recall: 0.200000

Visual Search with the dimensions reduced to 1 x 850 using Mahalanobis Distance
Reduced Dimensions:

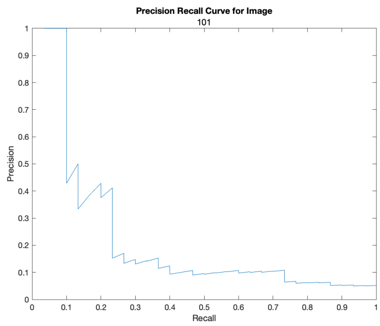
Name	Size	Bytes	Class
F	1x850	6800	double



Precision: 0.500000
Recall: 0.166667

Visual Search with the dimensions set to 1650 using Mahalanobis Distance
Reduced dimensions:

Name	Size	Bytes	Class
F	1x1650	13200	double



Precision: 0.400000
Recall: 0.133333

5.6.1 Conclusion from the experiment

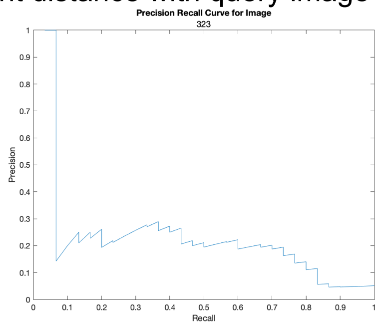
From the above experiments on the Bookshelf query image, several things were observed and deduced. Upon reducing the dimensions to the lowest size (1 x 850) the precision value obtained was 0.500. This low precision is probably due to the fact that a lot of the principal components or key features were lost by reducing the dimensions of the image this far. Next, upon reducing the dimensions down to the relative highest size (1 x 1650) the precision value obtained was 0.400. Again, the precision was lower than acceptable and this is likely due to fact that there is a lot of noise left in the image after dimensionality reduction and the noise was preserved instead of the principal components or features. Finally, when the dimensions were reduced to 1 x 1400, the precision value obtained was 0.600. Although this value was lesser than the precision of the descriptor before PCA (obtained in the previous experiment), it was still higher than the other two visual searches executed in this experiment. This is due to the fact that in this dimension, noise was lesser than the higher dimension but the principal components or features were greater than in the lower dimension.

Upon the completion of this experiment and others, it was noted that the overall performance of the system decreased after PCA when compared to the Spatially Gridded Colour and Edge Histogram descriptor.

5.7 Experiments conducted using the Pearson's Correlation Coefficient and Cosine Similarity as the distance measures

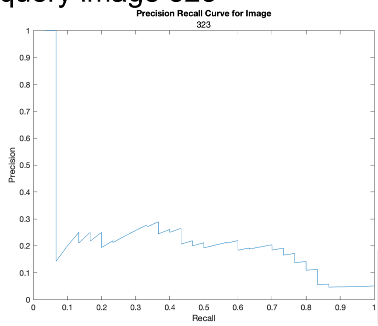
In this experiment, we are using the Spatially Gridded Colour and Edge Histogram descriptor with quantization = 6 and Grid factor = 4. We will be using the Pearson's Correlation Coefficient Distance, the Cosine Similarity distance and the Manhattan distance as the distance measures and evaluate the results using Precision – Recall and the PR Curve. The same query image will be used for each visual search with each of the distances.

Visual Search using Pearson's Correlation Coefficient distance with query image 323



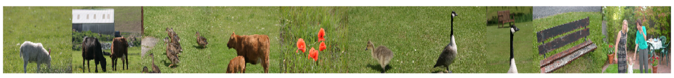
Precision: 0.200000
Recall: 0.066667

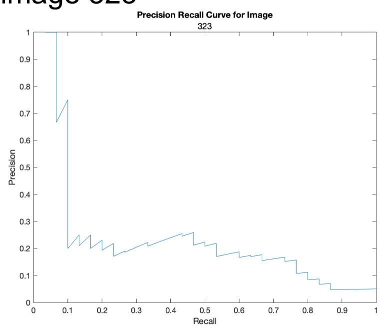
Visual Search using Cosine Similarity distance with query image 323



Precision: 0.200000
Recall: 0.066667

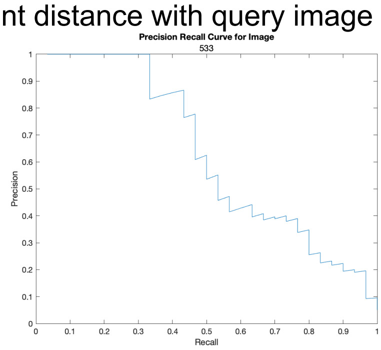
Visual Search using Manhattan distance with query image 323





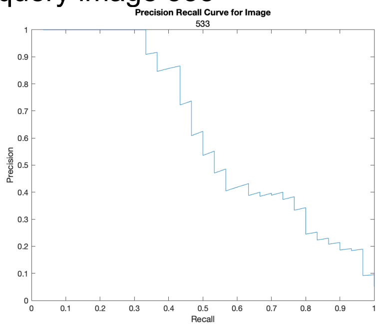
Precision: 0.300000
Recall: 0.100000

Visual Search using Pearson's Correlation Coefficient distance with query image 533



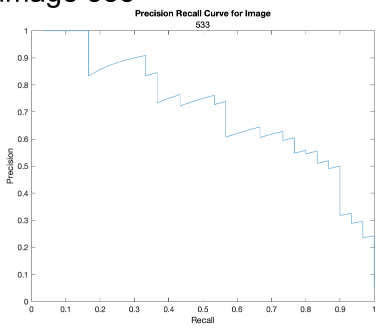
Precision: 1.000000
Recall: 0.333333

Visual Search using Cosine Similarity distance with query image 533



Precision: 1.000000
Recall: 0.333333

Visual Search using Manhattan distance with query image 533



Precision: 0.900000
Recall: 0.300000

5.7.1 Conclusion from the experiment

From the above experiment, it was noted that both the Pearson's Correlation Coefficient and the Cosine Similarity distances worked better than the Manhattan distance for images that can be searched accurately using the Spatially Gridded Colour and Edge Orientation Histogram descriptor and worse than the Manhattan Distance for images that cannot be searched accurately using this descriptor.

Hence, from the above experiment and several others which were performed, it was noted that while both these distances worked extremely similar to each other for most images, they always performed well or worse together based on how accurate the descriptor itself was for the respective query image. Therefore, it can be concluded that the Pearson's Correlation Coefficient and the Cosine Similarity distances are good options to be used to increase the accuracy of an already accurate descriptor.

6 Overall Summary and Conclusion

From the experiments performed in this coursework, it was noted that each descriptor performed differently for different query images. The descriptors performed exceptionally well when the image had characteristics that could be identified by the descriptors but also poorly due to their inability to identify key components in some images. Hence, it is imperative to decide the choice of the descriptor based on the type of image that has to be analysed and compared to the dataset. Doing so would ensure that different types of images can be accurately analysed to produce the visual search results with highest precision. The Global Colour Histogram and the Colour Moments descriptors worked well when the main component of the image that was to be analysed was the colour. The Spatially Gridded Edge Orientation Histogram and the Spatially Gridded Colour and Edge Orientation Histograms worked well when the main components were edges that were present locally in the image.

The use of PCA resulted in the accuracy of the visual search process dropping either due to the fact that too much information was lost during dimensionality reduction or that too much noise was present instead of the principal components after dimensionality reduction.

Finally, it was noted that the best distance measure to be used was the L1 Norm or Manhattan Distance. The Pearson's Correlation Coefficient distance and the Cosine Similarity distance were powerful tools to further increase the accuracy of the Manhattan distance when the accuracy was already high.

7 References and Bibliography

[1] *hist*, R2023B. [Online]. Available: https://www.mathworks.com/help/matlab/ref/hist.html

[2] V. Anusha, V. Usha Reddy and T. Ramashri, *Content Based Image Retrieval Using Color And Texture.* [Online]. Available: https://www.ijert.org/research/content-based-image-retrieval-using-color-moments-and-texture-IJERTV3IS21051.pdf

[3] Jolliffe IT, Cadima J. 2016, *Principal component analysis: a review and recent developments*.*Phil. Trans. R. Soc. A 374*: 20150202. [Online]. Available: https://royalsocietypublishing.org/doi/10.1098/rsta.2015.0202

4. ChatGpt was used to understand concepts.