# *VISUAL SEARCH OF AN IMAGE COLLECTION*

EE003032 - Computer Visions and Pattern Recognition Assignment

*Orogun, Tega O (PG/T - Comp Sci & Elec Eng)*

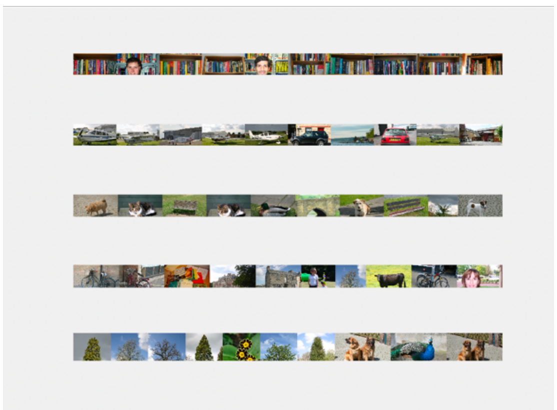
*University of Surrey | Stag Hill, University Campus, Guildford GU2 7XH*

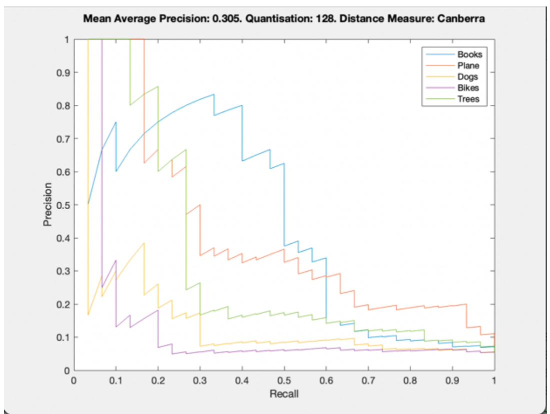
# ABSTRACT

This project aims to show different methods that can be used to search an image collection visually. Each method accepts an image as a query and returns a list of images that best fit the query image. Different image descriptors and distance measurements are used in each method to differentiate and pick the best images that are similar to the given query image.

## Table of Contents







# MAIN VISUAL SEARCH TECHNIQUES

## GLOBAL COLOUR HISTOGRAM

A global colour histogram is the first descriptor used to implement the visual search. This is the most common descriptor used in image processing. A colour histogram represents the overall colour distribution of an image(1). Generally, digital images apply this in the RGB (Red, Green, Blue) or HSV (Hue, Saturation, Value) colour space (for this assignment, this was applied in the RGB colour space).

To calculate the global colour histogram for an image, given a pixel in RGB space, we quantise each colour into a certain amount of levels (I will assume the quantisation level is equal to Q here). So when quantising a pixel: r' = floor(r*Q/256) for the red colour space, g' = floor(g*Q/256) for the green colour space and b' = floor(b*Q/256) in the blue colour space. We can then calculate a single integer value for the pixel RGB value by: $r' * Q^2 + g' * Q + b'$.

Afterwards, once the global histogram for the images has been calculated, the query image is then represented in a feature space. A distance measure is then applied to help the comparison with the query image with the other images in the dataset. Similar images will appear closer to the query image in the feature space. The distance measures used in this assignment are explained in more detail further on in this report.

Although this method is relatively easy to implement, as you will see in the results later on, this is not a completely accurate descriptor as it does not provide any spatial data or information for the query image.

## SPATIAL COLOUR GRID

This is a based image descriptor that does take spatial colour information into account, unlike the global colour histogram. This descriptor is calculated by dividing the image into equal cells using a grid. We then calculate each cell's average colour(RGB) in the grid and concatenate the values to form the image descriptor(1).

Once the image descriptor for the query image is formed, we then represent the image in a feature space and find the images that are the most similar to it using distance measures e.g. Euclidean distance measure or the Mahalanobis distance measure.

## EDGE ORIENTATION HISTOGRAM (EOH)

The Edge Orientation Histogram is an image descriptor that focuses on analysing the texture of an by examining the differences in edge orientation. Unlike the colour-based descriptors, the Edge Orientation Histogram is more concerned with edges of an image. When computing the EOH, an image is typically converted to grayscale. This is because the descriptor is more interested in intensity rather than changes in colour. With a grayscale image, the edge detection process is simplified(2).

In this project, the Sobel Filter was used as the edge detection technique. It computes the gradient of an image intensity to find the direction of a sharp change. A sharp change would usually correspond to an edge of some sort. This process of using the Sobel filter typically requires using two convolution operations: Horizontal Edge Detection and Vertical Edge Detection. With horizontal edge detection, the grayscale image is convolved with the following 3 x 3 matrix: [1 0 -1; 2 0 -2; 1 0 -1]. The same process is done with vertical edge detection but with the following matrix: [1 2 1; 0 0 0; -1 -2 -1;]

After the application of both filters, the results would then represent the gradient in both the horizontal and vertical direction. Subsequently, the magnitude and direction of the gradient can be used to calculate for each pixel using:

$$G = \sqrt{G_x^2 + G_y^2}$$

Where G is the gradient and Gx and Gy correspond to the horizontal and vertical edge detections respectively.

As for the direction, the equation is as follows:

$$\theta = \arctan\left(\frac{G_y}{G_x}\right)$$

Where Θ is the gradient direction and Gx and Gy correspond to the horizontal and vertical edge detections respectively.

The orientations are then quantised to a specific number of bins (this reduces the complexity of the histogram). Similar to the Spatial Colour Grid, a grid size is specified, although in this grid, it is the histogram of edge orientations that is computed rather than colour. This makes the Edge Orientation Histogram very suitable for images where, considering we are looking for images for similar textures, very useful.

## SPATIAL COLOUR GRID AND EDGE ORIENTATION HISTOGRAM

As stated previously, the Spatial Colour Grid and Edge Orientation Histogram have a similarity with both descriptors using a grid size but for different purposes. Combining both descriptor methods requires concatenating the RGB the average colour values in the grid and also computing the histogram of edge orientations in the same grid. The final feature descriptor value for each cell is concatenation of both results. This feature value encompasses both colour and texture information, which enables a more comprehensive analysis of an image's features.

## ADDITIONAL IMAGE DESCRIPTORS

The above highlighted descriptors are the main descriptors experimented with in this project. However, additional descriptors were experimented with to test the results with the dataset.

### LOCAL BINARY PATTERN (LBP)

The Local Binary Pattern is another texture descriptor that was experimented with in this project. LBP works by comparing every pixel with its encompassing neighbours. Similarly to the EOH descriptor, the first step is to convert the image to grayscale as we are dealing with intensity. Following that, a specific radius is defined for every individual pixel to compare the pixel to its encompassing neighbours. The typical size for this is a 3 x 3 radius(which is the only radius shown in the results further below in this report). Each neighbour is compared with the pixel in the centre. If the neighbour's intensity is equal or greater than the pixel in the centre, a binary value of 1 is assigned to it, otherwise a value of 0 is assigned. These values are then concatenated to create a pattern (creating a binary number), before this eventual pattern is turned to a decimal number. This number is used in the image retrieval process(3,4).

There have been several variants of this descriptor created to increase its efficiency but those are not explored in the scope of this project.

### COLOUR CORRELOGRAM

This is a colour descriptor that summarises the spatial colour distribution in an image. This is done by conveying how the probability of finding a pixel of a particular colour changes depending on the distance from pixels of another colour. This descriptor main focus is the spatial relationship between the colours in an image. For every colour in an image(a quantised image version to reduce complexity) and for the specified distances, the probability of finding a pixel of a colour at the distance from another pixel of another colour is calculated. This process takes quite some time, which will be briefly discussed in the results section of this project(5,6).

# DISTANCE MEASURES

This section contains a brief description of the distance measures used in this project. However, it is important to go over the concept of Principal Component Analysis briefly.

## PRINICIPAL COMPONENT ANALYSIS (PCA)

Principal Component Analysis is process carried out when dealing with distance measures. PCA reduces the number of dimensions when dealing with high-dimensional data such as the pixels in the images in this project. This makes the distance calculations more efficient. It can also filter out the noise from the data. This can be done by collecting the principal components during the analysis.

In this project, the PCA was primarily only used coincided with the Mahalanobis distance measure, which is detailed below. Although in a few experiments PCA was conducted while using other distance measures. Helper functions to perform the PCA are included in the files used to perform this project.

## EUCLIDEAN DISTANCE

The Euclidean distance is a measure of the straight line distance between two points in a Euclidean space. The Euclidean distance between two points, a and b, in a n-dimensional space is:

$$d = \sqrt{\sum_{i=1}^{n} (b_i - a_i)}$$

## MAHALANOBIS DISTANCE

The Mahalanobis distance is a measure that measures the distance between points in units of standard deviation. Unlike the Euclidean distance, which is the straight line distance, the Mahalanobis is concerned more with the general spread, as it also takes the distribution of the data into account. The equation for the Mahalanobis is as follows:

$$d = \sqrt{\sum_{i=1}^{n} \frac{(b_i - a_i)}{E_i}}$$

Where E contains the eigenvalues of the concerned space.

## MANHATTAN DISTANCE (L1 NORM)

The Manhattan Distance (also known as the L1 Norm Distance Measure) is a distance in a grid-based path. The formula for this distance measure is as follows:

$$d = \sum_{i=1}^{n} |b_i - a_i|$$

## CANBERRA DISTANCE

The Canberra Distance is weighted version of the Manhattan Distance. This distance is more sensitive to small changes if the variable values are greater(7). The formula for this distance measure is as follows:

$$d = \sum_{i=1}^{n} \frac{|b_i - a_i|}{|b_i| + |a_i|}$$

## MINKOWSKI DISTANCE

The Minkowsi distance measure is a measure between two vectors in a normed vector space. It is a generalised distance form of both the Euclidean and the Manhattan distance measures(8). Its formula is as follows:

$$\left(\sum_{i=1}^{n} |b_i - a_i|^r\right)^{\frac{1}{r}}$$

Where r is the Minkowski metric. If r is set to 1 or 2, the result corresponds to the Manhattan Distance or the Euclidean Distance respectively.

In the scope of this project, the main distance measures are the Euclidean, Manhattan and the L1 distance measures.

## EVALUATION METHODOLOGY

During the period of this project, multiple files have been created. Typically, about three files were created for each chosen descriptor, excluding the functions for the different distance measures. Among these three files, a script was used to compute the descriptor. Subsequently, a single testing script, '*test.m',* was implemented. This script calls a descriptor's visual search function which passes in two queries which correspond to a query image and a distance query. Each descriptor is invoked five times with five different image queries that are declared at the beginning of the test file. These five images correspond to images in five different classes from the MSVRC-v2 dataset, which is the primary data which all the experiments and results were obtained from in this project. These images were chosen to show the variation across different classes in the dataset. It is important to note that the query images were limited to five classes as during experimentation, it was observed that any more than five classes obscured the visibility of the Precision-Recall Curve plotted by the test file which will be talked about later on.

When the test file calls one of the visual search functions (it is important to note that only one descriptor's visual search should be called at a time to maintain clarity), one of the selected images is passed in as the index for the query image. Simultaneously, one of the distance measures described above is also passed in to the visual search function (again it is important to note that using more than one distance would degrade the integrity of the Precision-Recall Curve plotted at the end therefore, it is important to only use one distance measure when running the search for all images). The visual search then carries out its routine, noting that it would only perform PCA if the distance measure passed in is the Mahalanobis measure. The 'SHOW' variable is assigned to the amount of images that would be returned to compare with the query image. During this project, this variable was changed between '10' and '591'. With '10', the images are more visible to the user as they are sub plotted by the test file for the user. With '591', the user will get a more accurate representation from the Precision-Recall Curve and also when calculating the Average Precision and Mean Average Precision which were the chosen metrics used in this project for analysis of each descriptor.

The files also contain a handful of helper functions for certain descriptors and process. Any file beginning with 'Eigen' was essential for carrying out the PCA process. 'colourPairsProb' is a helper function for the additional Correlogram descriptor and 'binaryToVector.m' is another self-explanatory helper function (MATLAB does have a built function to perform binary to decimal conversion but this was not performing well).

Regarding the naming convention used in this project, each file indicated its purpose. Descriptor computation files are suffixed with '_cvpr_computedescriptors'; files for the descriptor calculations are suffixed with '_extractRandom'; visual search files are suffixed with '_cvpr_visualsearch'; and distance comparisons are suffixed with '_cvpr_compare'.

## RESULTS

As stated above in the methodology, the experiments and evaluations in this project were carried out on the MSVRC-v2 dataset. For each descriptor, different quantisation levels, grids sizes and a combination of both were conducted to evaluate the differences when these variable were altered. A Precision-Recall Curve, plotted against all 591 images is shown alongside the top 10 images for each of the five query images that have been chosen. These images were picked to demonstrate a descriptors effectiveness in correctly returning the appropriate image alongside the query image, as a descriptor could be effective on a certain class but ineffective on another. The Average Precision was carried out for each query and with this value, the Mean Average Precision was then calculated and displayed with each graph plot. This shows the user the effectiveness of the plots between the different classes.

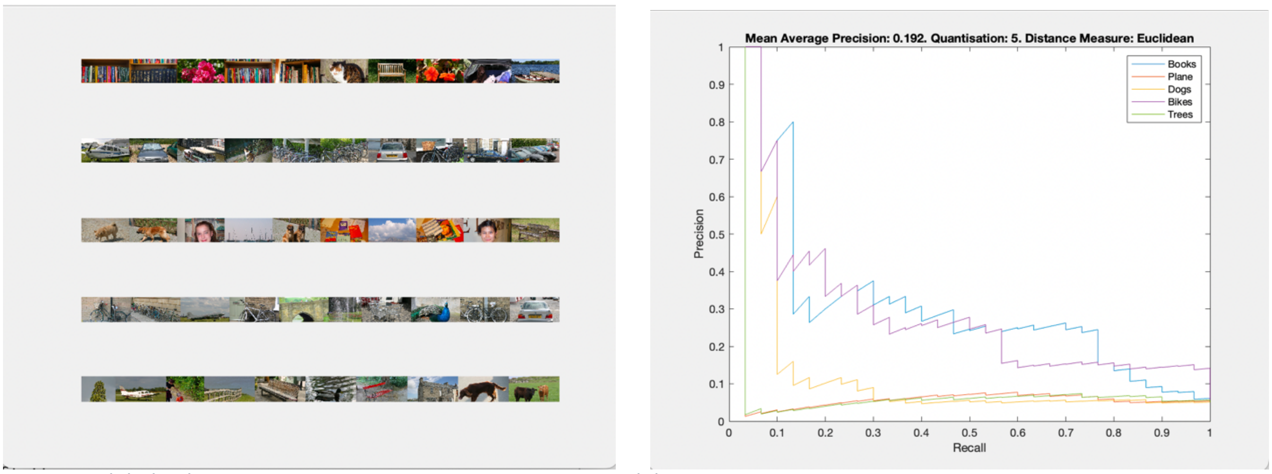



*Figure 1: Global Colour Histogram. Quantisation: 5. Distance: Euclidean*

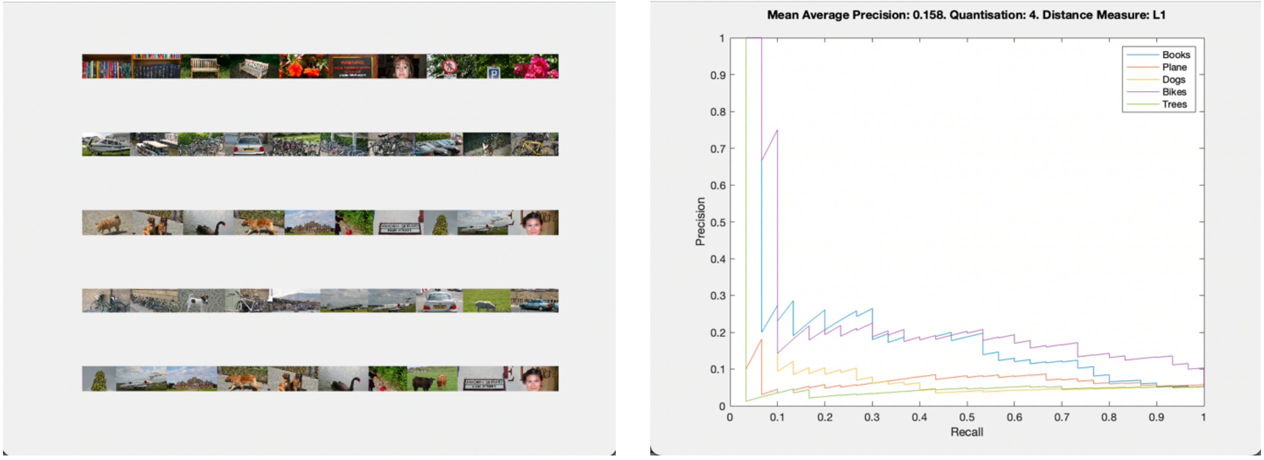



*Figure 2: Global Colour Histogram. Quantisation: 4. Distance: L1.*

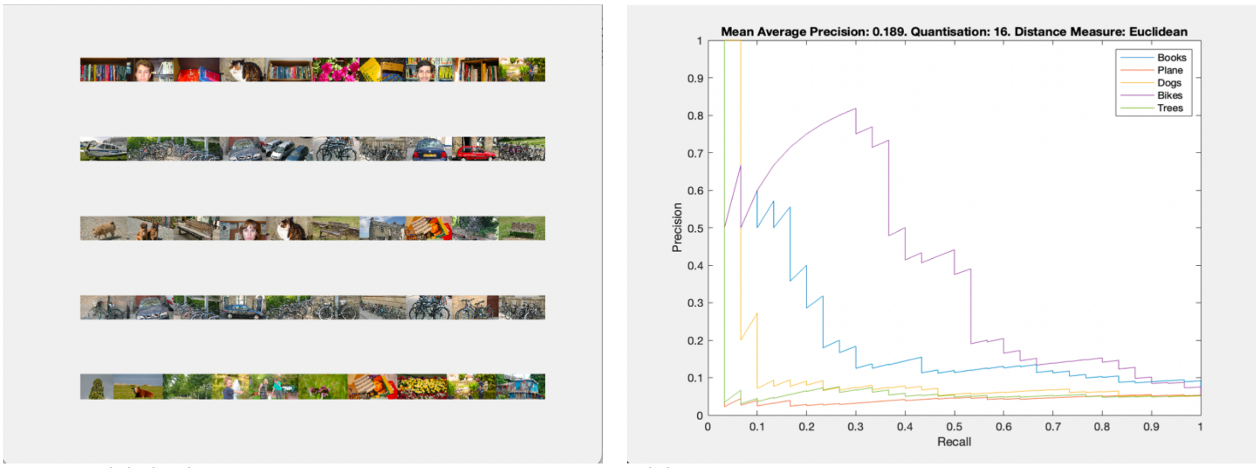



*Figure 3: Global Colour Histogram. Quantisation: 16. Distance: Euclidean*

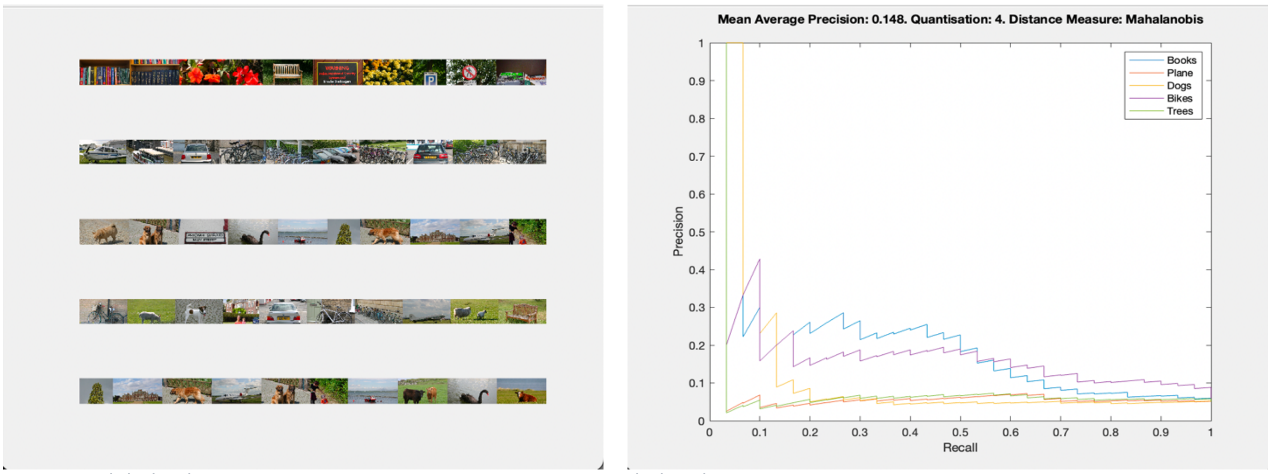



*Figure 4: Global Colour Histogram. Quantisation: 4. Distance: Mahalanobis.*

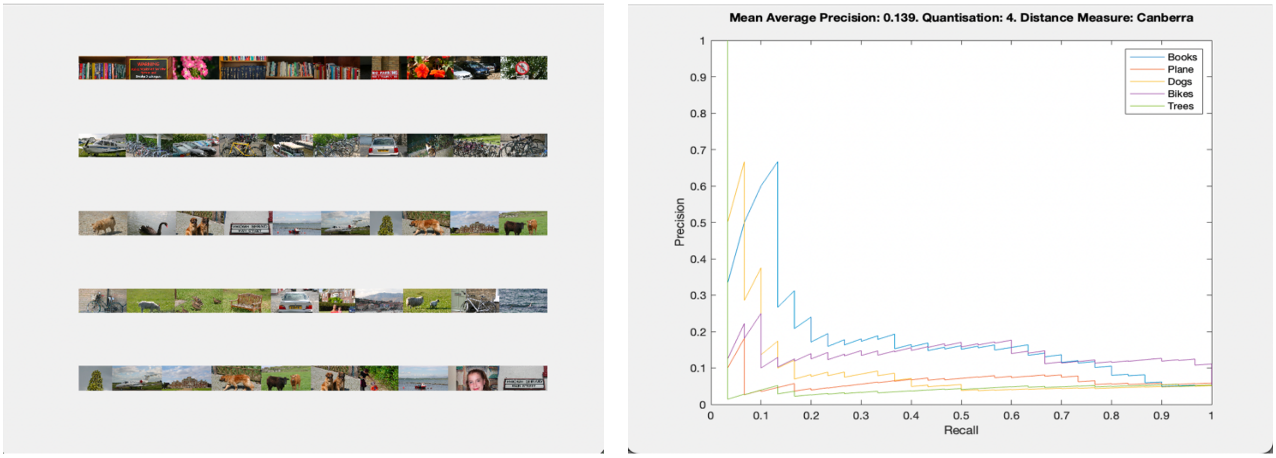



*Figure 5: Global Colour Histogram. Quantisation: 4. Distance: Canberra*

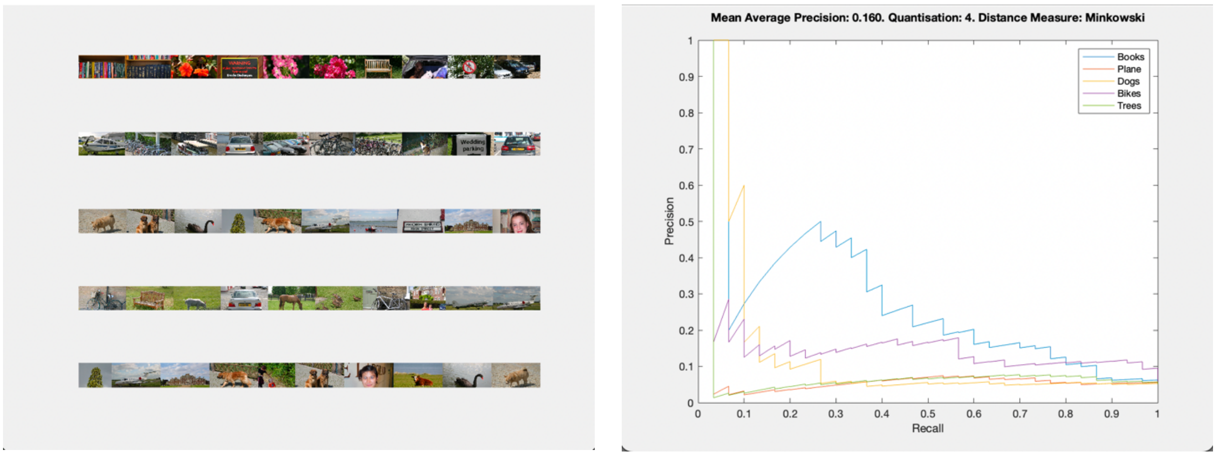



*Figure 6: Global Colour Histogram. Quantisation: 4. Distance: Minkowski.*

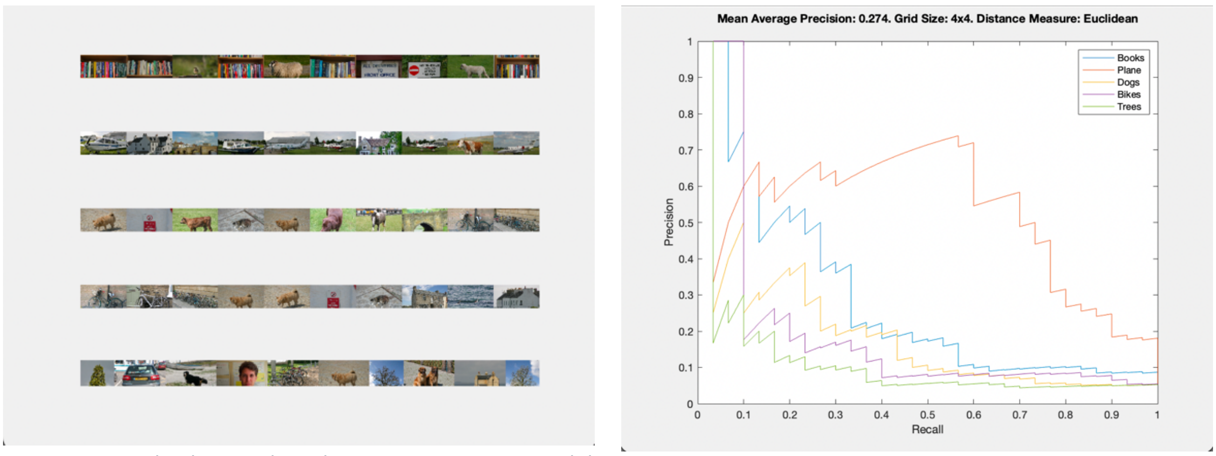



*Figure 7: Spatial Colour Grid. Grid Size: 4 x 4. Distance: Euclidean*

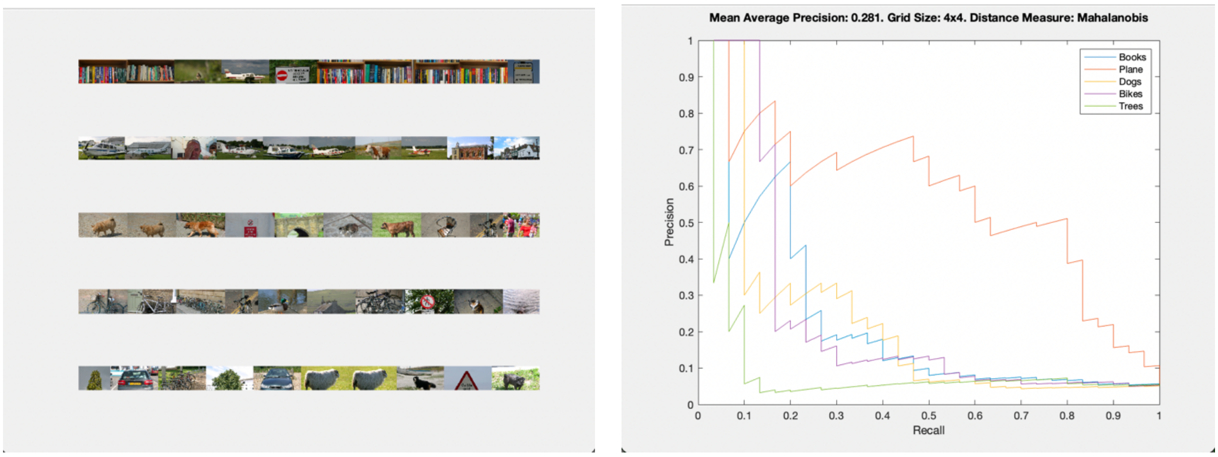


*Figure 8: Spatial Colour Grid. Grid Size: 4 x 4. Distance: Mahalanobis*

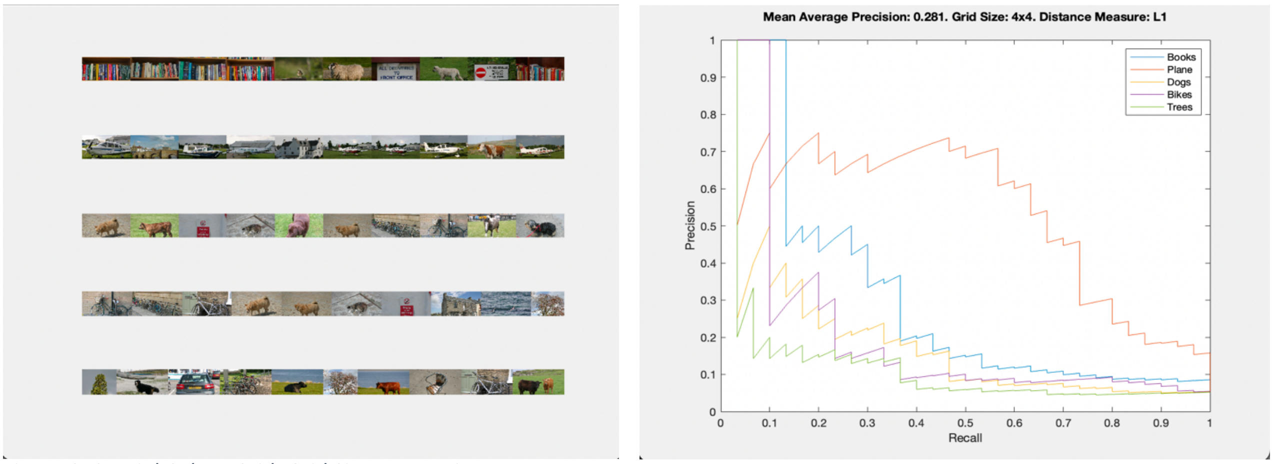


*Figure 9: Spatial Colour Grid. Grid Size: 4 x 4. Distance: L1*

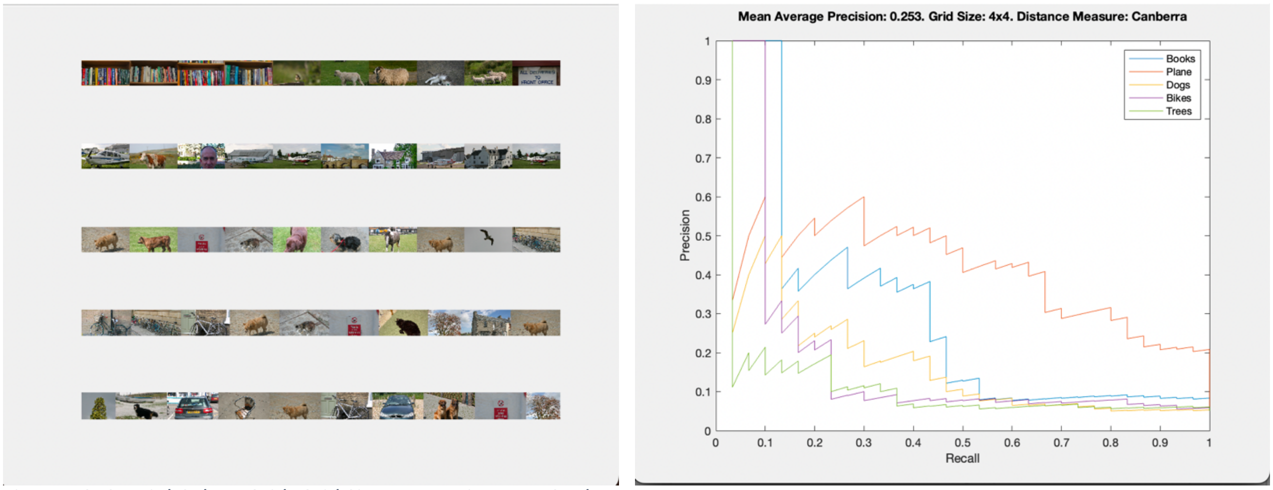


*Figure 10: Spatial Colour Grid. Grid Size: 4 x 4. Distance: Canberra*

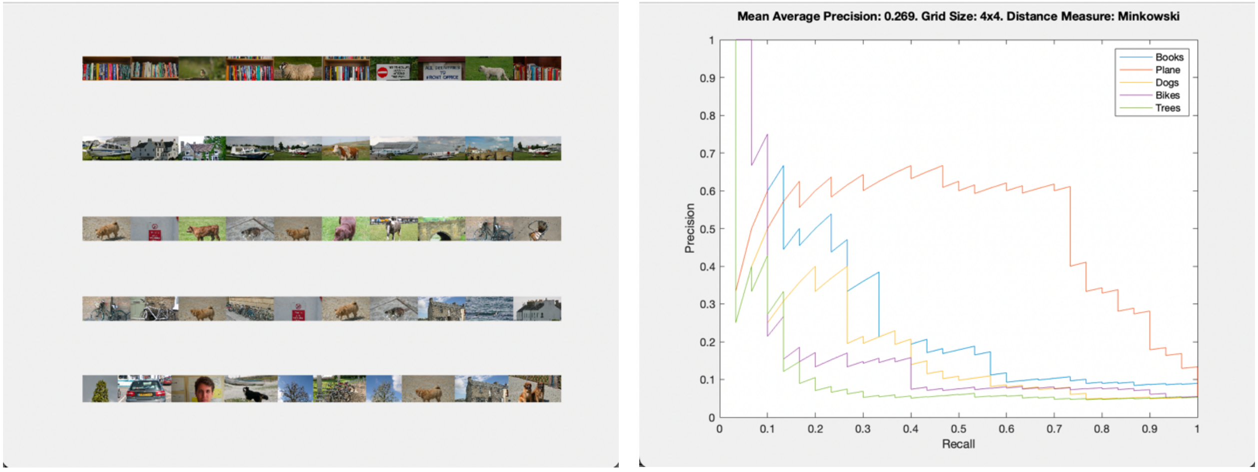


*Figure 11: Spatial Colour Grid. Grid Size: 4 x 4. Distance: Minkowski (Order 3)*

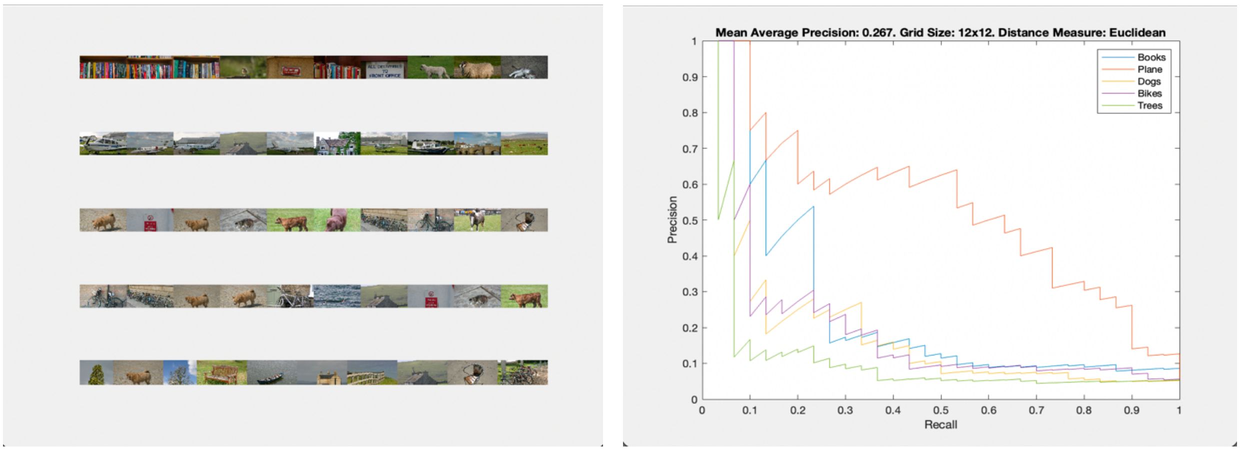


*Figure 12: Spatial Colour Grid. Grid Size: 12 x 12. Distance: Euclidean*

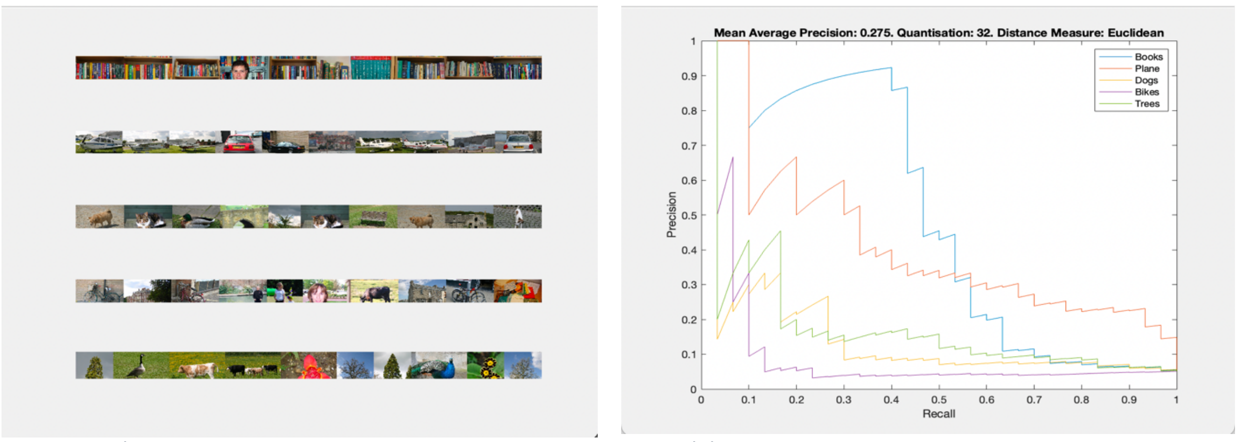


*Figure 13: Edge Orientation Histogram. Quantisation: 32. Distance: Euclidean*

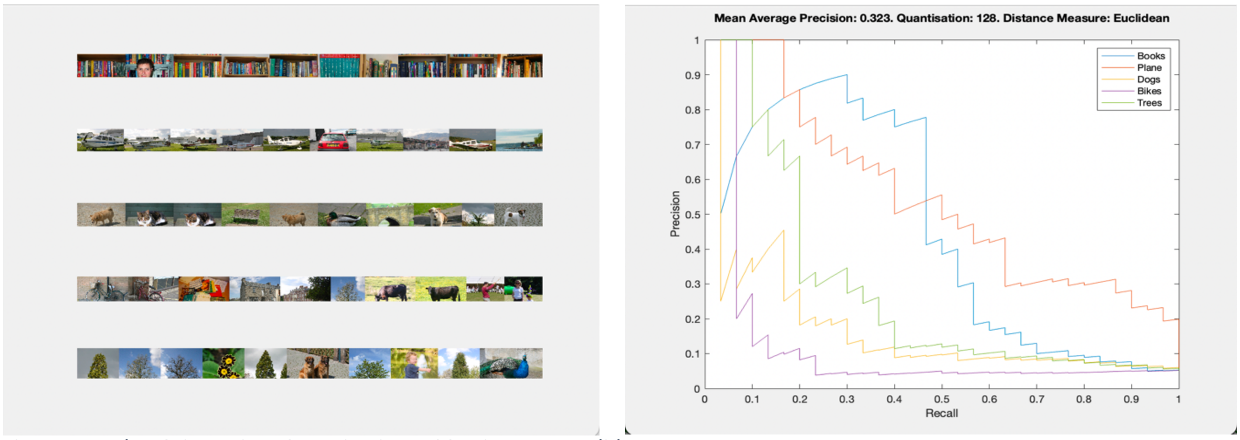



*Figure 14: Edge Orientation. Quantisation: 128. Distance: Euclidean*

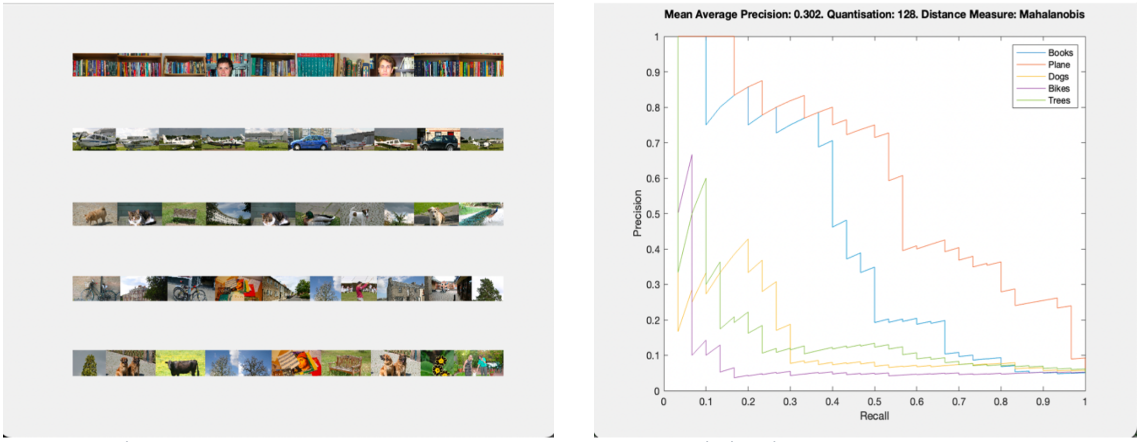



*Figure 15: Edge Orientation Histogram. Quantisation: 128. Distance: Mahalanobis*

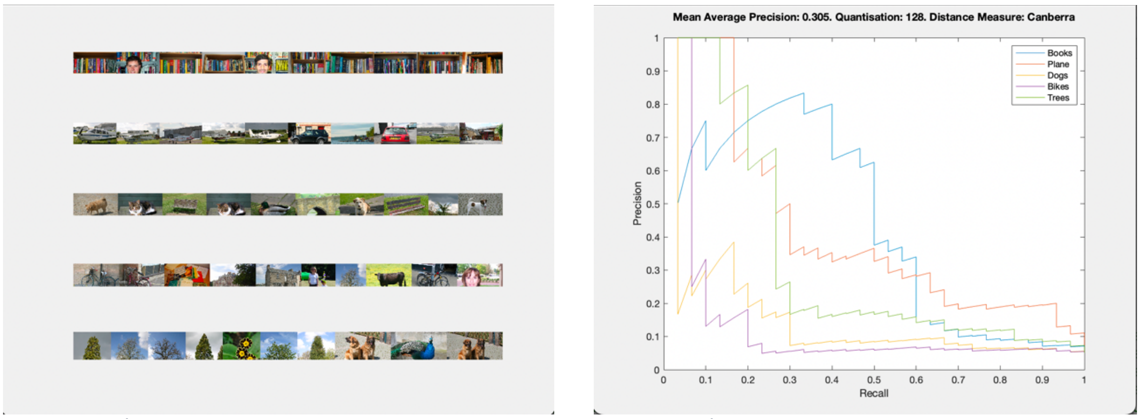



*Figure 16: Edge Orientation Histogram. Quantisation: 128. Distance: Canberra*

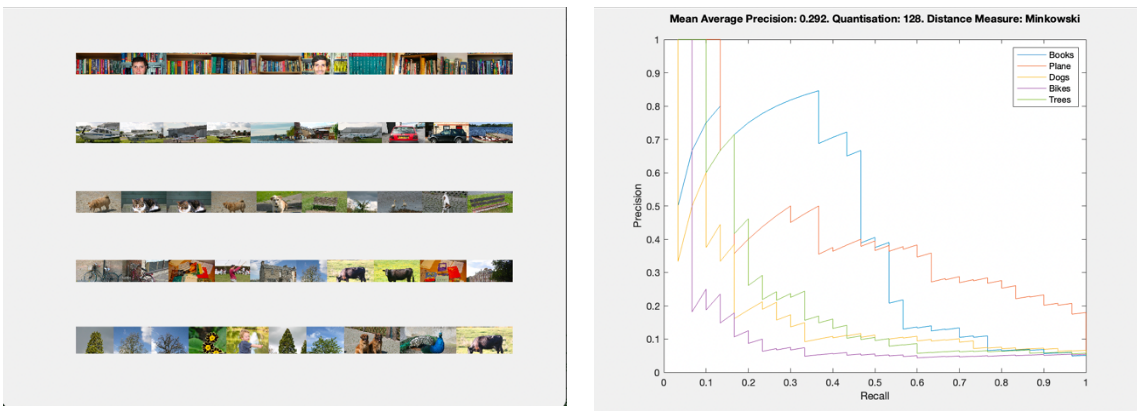


*Figure 17: Edge Orientation Histogram. Quantisation: 128. Distance: Minkowski (Order 3).*

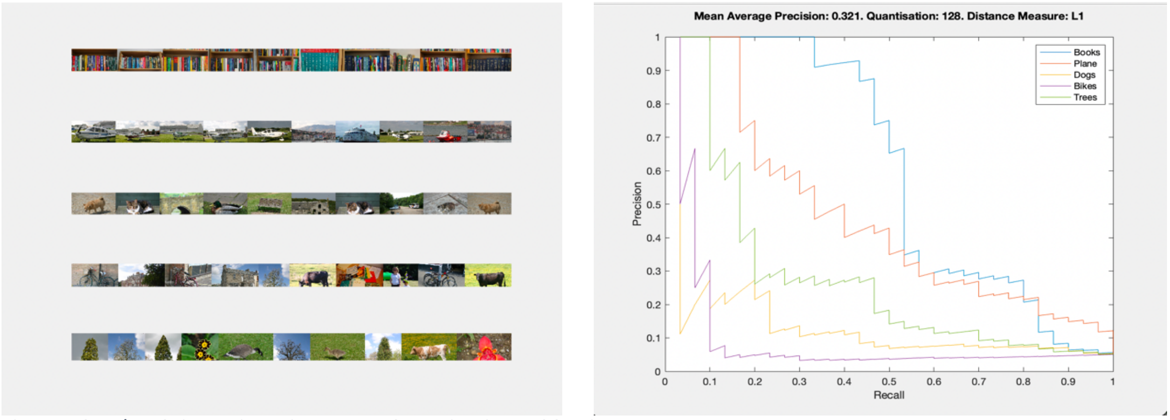



*Figure 18: Edge Orientation Histogram. Quantisation: 128. Distance: L1.*

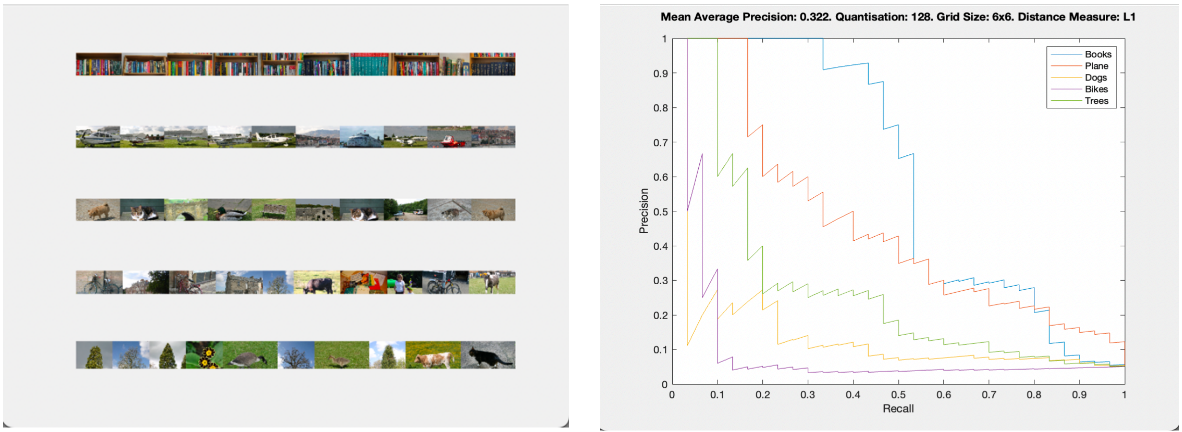



*Figure 19: Colour Grid & Edge Orientation Histogram. Grid Size: 6 x 6. Quantisation: 128. Distance: L1*

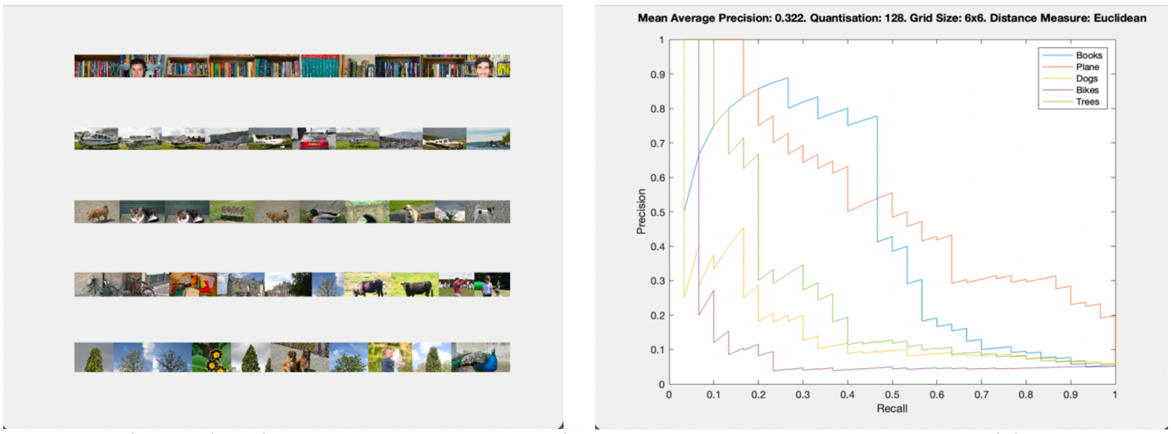



*Figure 20: Colour Grid & Edge Orientation Histogram. Grid Size: 6 x 6. Quantisation: 128. Distance: Euclidean*

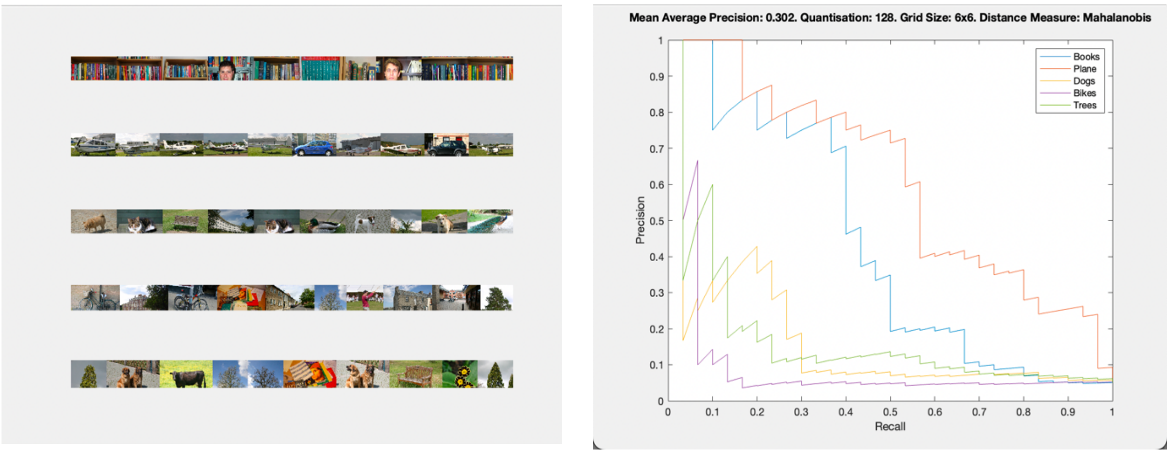



*Figure 21: Colour Grid & Edge Orientation Histogram. Grid Size: 6 x 6. Quantisation: 128. Distance: Mahalanobis*

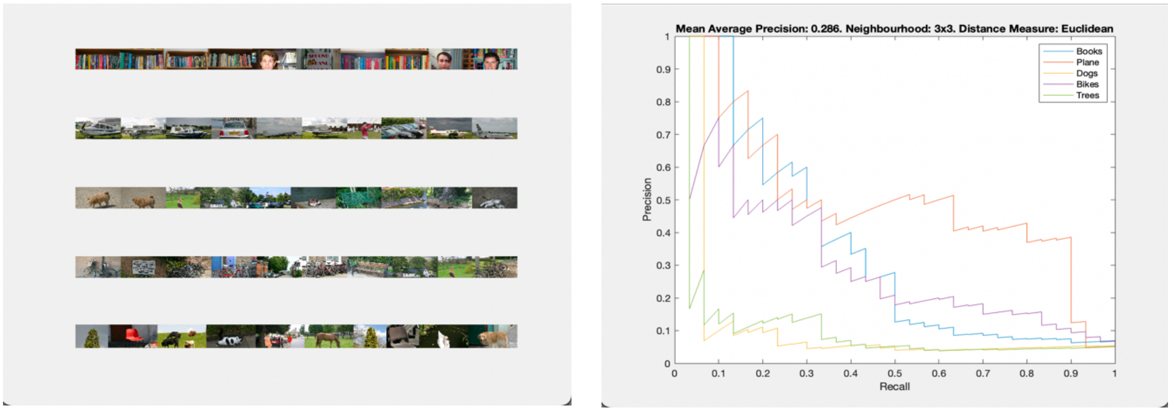



*Figure 22: Local Binary Pattern. Neighbourhood: 3 x 3. Distance: Euclidean*

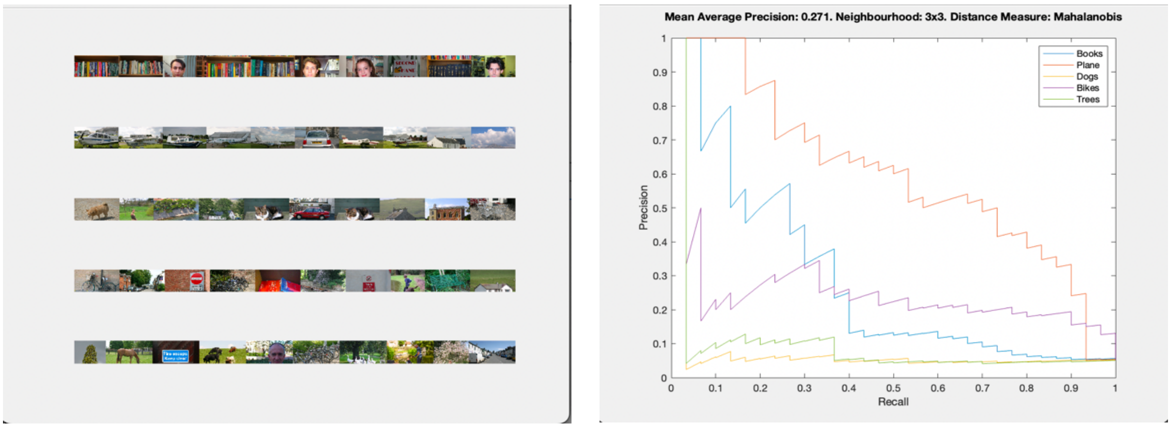



*Figure 23: Local Binary Pattern. Neighbourhood: 3 x 3. Distance: Mahalanobis.*

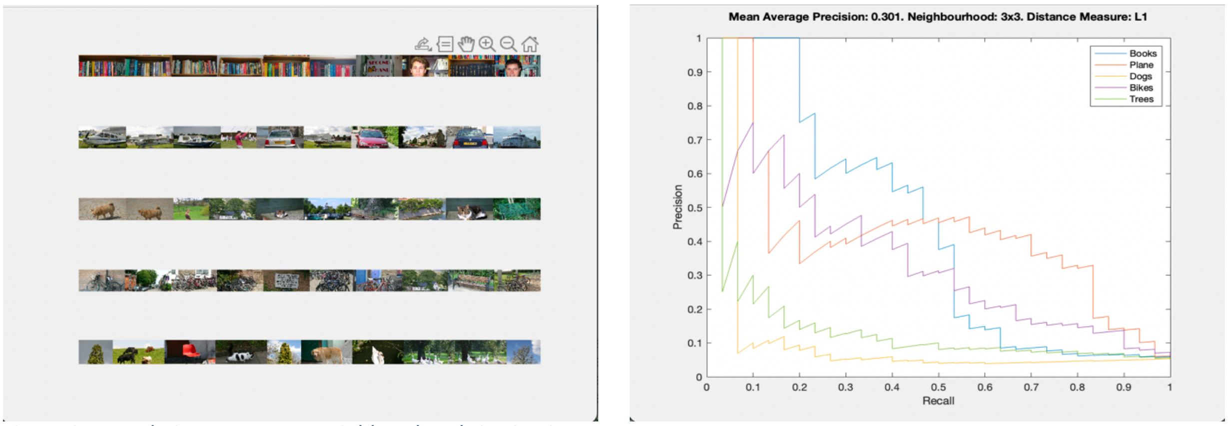


*Figure 24: Local Binary Pattern. Neighbourhood: 3 x 3. Distance: L1.*

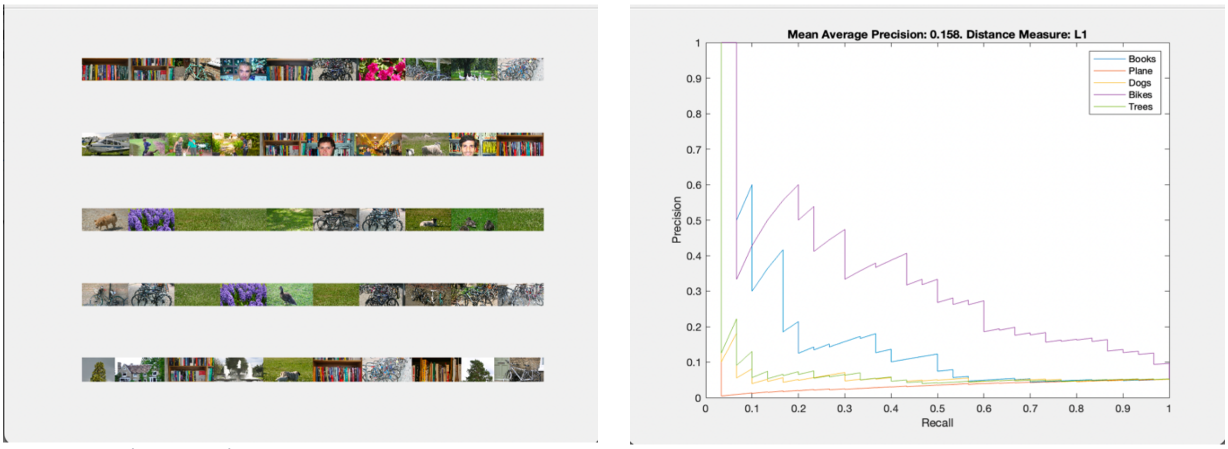


*Figure 25: Colour Correlogram. Distance: L1*

## DRAWN CONCLUSIONS FROM EXPERIMENTS

Multiple conclusions were drawn after conducting the experiments with the different descriptors and the distance measures in this project. With these conclusions, it is important to note that although we have used the Precision-Recall Curve and the Mean Average Precision as evaluations for the descriptors, these do not tell the whole truth of the efficiency of each descriptor. It is also worth noting most of these observations were for the five classes selected above (Books, Planes, Dogs, Bikes and Trees).

The Global Colour Histogram, as described earlier is a colour descriptor. It does not consider spatial information which was apparent from the top 10 images returned on all the distance measures and quantisation levels experimented with. Its highest Mean Average Precision (MAP) was observed at a quantisation level (Q) of 5. An original assumption when approaching this descriptor was that as the quantisation level got larger, the precision of the descriptor would improve. This was disproven, since a quantisation of 5 had a superior MAP to a quantisation level of 16 (although this quantisation level was better than that of Q = 4). With these results, a conclusion was drawn that an appropriate quantisation level would need to calculated or determined to get the best results from the dataset. This does not solve the issue of spatial information in the images. It performed relatively okay with the Books class, presumably because the books class contains a consistent  amount of colours among all its images. Beyond that, it did not perform that well as expected.

The Spatial Colour Grid is another one of the main descriptors used in this project. Spatial information is considered with this colour descriptor, so it was expected that the resulting images and the calculation for the MAP would be an improvement when compared to the Global Colour Histogram (GCH). This proved true as all the MAP with all distance measures were greater than the results from the GCH with the best results coming from a grid size of 4 x 4 with the distance measures of Mahalanobis and L1. In one of the experiments conducted with this descriptor, the grid size was increased to a size of 12x12 and although the MAP was slightly lower, the difference is negligible. With additional experiments, conclusions could be drawn that there exists also an appropriate grid size to use to conduct searches with this descriptor. It is also important to note that  the returned top images aligned with the descriptor's intent. That is to say that when the grid calculated the average colour in a certain space, it returned images that matched the calculated colour. So, although this also is not the best descriptor, its results are satisfactory.

The Edge Orientation Histogram is a texture based descriptor as described previously. The results from this were a significant improvement from the previous descriptors with its highest MAP almost double the lowest from the Global Colour Histogram. Its highest MAP was achieved with a quantisation level of 128(which would be the number of bins). With this descriptor, it is important to note that the MAP and results do not completely tell the whole truth of the accuracy of the descriptor on some classes, in this case the Books class. With Books class, some of the top images returned were from another class that contained people's faces. In these images, the descriptor detected the books behind the person's face which was very impressive as it detected the texture of the bookshelf even when hidden. This result highlights a flaw with the dataset as although as some classes can be combined or convolved with each other as their similarities are too great. The results from this surpassed expectations.

The next descriptor used was a combination of the Spatial Colour Grid and the Edge Orientation Histogram. The expectations for this descriptor was to use the intent of both descriptors and improve returned results as with using them both, we now consider both spatial positioning and texture. Its highest MAP was significantly better than that of the colour grid but did not outrank the best retuned MAP of the Edge Orientation Histogram on its own (combined highest MAP = 0.322, Edge Orientation Histogram highest MAP = 0.323). The conclusion drawn from this was that, the results and accuracy varied because of specific objects in each class. For example, for the Books class, the descriptor worked more than adequately for the books that were positioned on a shelf but would not prioritise the books that were on the tables. For cases like these, the shape of an object would suffice. Colour calculations would yield to incorrect results. Although it did not meet expectations, it still proved to be one of the better descriptors in this project.

For the additional descriptors experimented with, the Local Binary Patterns descriptor, another texture descriptor performed adequately when experimented with. As expected, it worked well with images where the colour intensity variance between pixels was not too great (Books, Planes). Its highest MAP being around some of the results from the Edge Orientation Histogram. As for the Colour Correlogram, another additional descriptor experimented with on this project, it performed the worst. The conclusion drawn from this was that the bin size used when computing this descriptor might have been inadequate but as discovered during experimentation, this descriptor is very computational intensive. A bin size of around 256 might yield better results, unfortunately, that would be too intensive for the equipment used to conduct these experiments.

In summary, I believe the texture-based descriptors performed better than the colour-based descriptors. It is important to note that this statement only pertains to the selected classes for this experiment and this could be different in some of the different classes in the datasets. It is suspected that if the Spatial Colour Grid or Global Colour Histogram was used on classes that contained grass(e.g. Cows, Ducks), the MAP from those classes would rival that of the Books and Planes that have been selected for this project. The classes selected for this project all had distinct edges, which is why the drawn conclusion was that the texture descriptors performed significantly better.

## FUTURE IMPROVEMENTS

For improvements that could be made to this project, selecting more classes in the dataset would help conclude on the best descriptor to use on it. Experimenting with the other classes would yield more results. Additionally, for evaluation a Confusion Matrix could be implemented using the ground truth images for accurate evaluation. There are also unexplored descriptor methods that were not experimented with in this project such as splitting the data set into a test set and training set using a supervised classification method. This would likely yield more accurate results.

## REFERENCES


1. Collomosse John Prof. Computer Vision and Pattern Recognition.
2. Pinheiro AMG. Image descriptors based on the edge orientation. SMAP ’09 - 4th International Workshop on Semantic Media Adaptation and Personalization. 2009;73–8.
3. Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng. 2010;22(10):1345–59.
4. Local Binary Pattern for texture classification — skimage 0.23.0rc0.dev0 documentation [Internet]. [cited 2023 Nov 14]. Available from: https://scikit-image.org/docs/dev/auto_examples/features_detection/plot_local_binary_pattern.html
5. Huang J, Kumar SR, Mitra M, Zhu WJ, Zabih R. Spatial color indexing and applications. Int J Comput Vis [Internet]. 1999 Dec [cited 2023 Nov 14];35(3):245–68. Available from: https://link.springer.com/article/10.1023/A:1008108327226
6. Image Indexing Using Color Correlograms [Internet]. [cited 2023 Nov 14]. Available from: https://www.researchgate.net/publication/2445753_Image_Indexing_Using_Color_Correlograms
7. Fadhillah Pulungan A, Zarlis M, Suwilo S. Performance Analysis of Distance Measures in K-Nearest Neighbor. 2020;
8. Minkowski Distance Formula in Data Analysis Guide [Internet]. [cited 2023 Nov 14]. Available from: https://techlitistic.com/minkowski-distance-formula/