Visual Search of an Image Collection

EEE3032 – Computer Vision and Pattern Recognition Coursework Assignment

Autumn 2021

Faiz Ahmad Khan

University of Surrey

# Abstract

The report includes description of techniques used and experimental observations made when visual search algorithm was applied on MSRCv2 dataset. Descriptors and distance measures of varying complexity were experimented upon. Additionally, the dimensionality of descriptors was reduced using principal component analysis and results were noted. Finally, a support vector machine was developed to classify a subset of images from original dataset.

Table of Figures

[Figure 1: L1 vs L2 Norm 6](https://d.docs.live.net/6b5a9591954bfa8f/Documents/Visual%20Search%20of%20an%20Image%20Collection.docx#_Toc87970709)

[Figure 2: Colour vs Texture Grid 7](https://d.docs.live.net/6b5a9591954bfa8f/Documents/Visual%20Search%20of%20an%20Image%20Collection.docx#_Toc87970710)

[Figure 3: No PCA vs PCA 8](https://d.docs.live.net/6b5a9591954bfa8f/Documents/Visual%20Search%20of%20an%20Image%20Collection.docx#_Toc87970711)

# 1 Introduction

Visual search is a remarkably convenient means to explore digital image collections. This is especially true for large collections such as those provided by Google and Amazon. These organisations have leveraged visual search technology to differentiate their products from competitors. The convenience of searching for items without describing their appearance through text is very desirable for applications such as text translation and online shopping. Visual search is an application of content-based image retrieval which identifies unique features in an image to develop a descriptor that is used for comparison and ranking similarity of images.

# 2 Description of Techniques

This section will describe the visual search techniques along with the descriptors that were evaluated.

## Global Colour Histogram

A Global Colour Histogram (GCD) is used to represent the overall colour distribution of an image. This distribution is then used as a descriptor. First, the image is represented as points in 3D colour space ,i.e., red, green, and blue intensity values of each pixel. The colours are quantized into (q) divisions Each colours intensity (Val) is converted into an integer bin value. This floating point is converted into an integer by truncating all decimal values. The result is the (bin value).

Thus, the RGB colour space has been quantised to a three-colour histogram. To convert to a descriptor, each point is reduced to a single number and represented by a global histogram. Each pixel is augmented into a single number in base (q) by taking decimal bin integers and then concatenating them. The histogram is normalised to counter the effect of varying resolution of images.

The descriptor lots an image as a point in (q3) dimension feature space and similarity can be calculated through a distance measure. The advantage of this descriptor is that objects can be compared by the groups of colours that represent them. The disadvantage with this descriptor is that it holds no spatial information for the colour groups and hence two different images will be matched if they just have similar colour composition.

## Spatial Colour

This descriptor utilizes spatial colour information by dividing the image space into grids within which the average colour composition is calculated and concatenated to form a descriptor. Within the grids GCD could be utilized as well but at a much greater computational cost. A spatial colour descriptor should perform better than a GCD as it encodes the distribution of colours in the image.

## Edge Orientation

This descriptor divides the image space into grids as well but measures high frequency information, i.e., textures and edges. This is achieved by converting the image to grey scale and utilizing a Sobel filter for each axis. This approximates the gradient of the pixel intensity of an image. The edge magnitude and angle of the edge are then concatenated to develop the descriptor. Only edge magnitude values of certain threshold are included in the descriptor to strike a balance between computational cost and accuracy.

## Principal Component Analysis

In simple terms, Principal Component Analysis (PCA) is used to condense information by applying a transformation on the data. The reasoning for this can be traced to the ‘curse of dimensionality’. Ideally, the dimensions are kept as low as possible while maintain an appropriate level of accuracy. This is done as each additional dimension exponentially increases the volume of feature space.

Thus, a higher dimension descriptor requires additional time and processing power. Multiple dimensions have minimal variation and thus do not contribute much to decision making and can be discarded. For PCA, the covariance (spread) and mean (position) is first calculated and then decomposed into eigen vectors and eigen values that define the direction and magnitude of variation respectively.

## Support Vector Machine

Support Vector Machines (SVM) is a supervised learning model that is widely used for binary classification problems. This is achieved by calculating the best possible supporting vector or dividing line between two group of observations. This model can also be utilized for classification tasks involving more than two categories by constructing multiple classifiers where each one trains data from two classes and graphically represented by plane or (hyperplane).

# 3 Distance Measure

This section will describe the distance measures that were evaluated. Distance measures are used to make comparisons between image descriptors plotted in a feature space. This is done by calculating the vector between two points in the feature space. The method evaluating this vector is known as the Norm.

## L1 Norm

Also know as the Manhattan distance, this is the sum of the absolute values of the vector.

## L2 Norm

Also known as the Euclidean distance, this is the shortest distance between two points in space. In mathematics it is substituted for the magnitude of a vector. This is the most commonly used distance measure.

## Mahalanobis Distance

The Mahalanobis is essentially the Euclidean distance, measured in terms of standard deviation. In machine learning this measure better accounts for a data model through PCA. It is calculated by measuring how many standard deviations away from the models mean a query point is. The Mahalanobis distance becomes the Euclidean distance when moving from the root frame to the frame defined through PCA.

# 4 Evaluation Techniques

This section will present a brief of the evaluation techniques that have been utilised.

## Precision and Recall

After ranking the images in a dataset in comparison to the similarity with a query image, the precision and recall statistics can be calculated for n images of the dataset.

Precision represents the percentage of returned results that are relevant to the query image. High precision indicates better system performance and in an ideal system precision is 1 as the value of n increases and eventually all relevant values have been returned, so the precision drops steadily to a minimum value given by the fraction of relevant results in a dataset.

The recall at n represents how many relevant results have been returned up to that point. Higher values a n indicate that a system has high confidence and can recall relevant results faster with low false positives. At the final value of n, recall is equal to 1 as all results have been returned. A system with high recall but low precision at n will indicate that the system is effectively able to retrieve all relevant documents eventually however there will be false positives within the results. On the contrary, a system with high precision but low recall at n would indicate that the system is very confident but may indicate an increase in false negatives when the system cannot correctly identify a relevant result.

One way to visualise the response of a visual search system is to calculate both precision and recall for all values of n and plot a PR curve. The recall is measures along the x axis and precision along y axis. The resulting curve plots the system performance over normalised time to retrieve the query’s relevant results.

## Mean Average Precision

The Mean Average Precision is calculated over a range of queries. These methods are best suited to evaluate level 1 of similarity. This report is limited to L1 comparisons as the dataset is not properly labelled and considerable work would be required.

# 5 Experimental Results

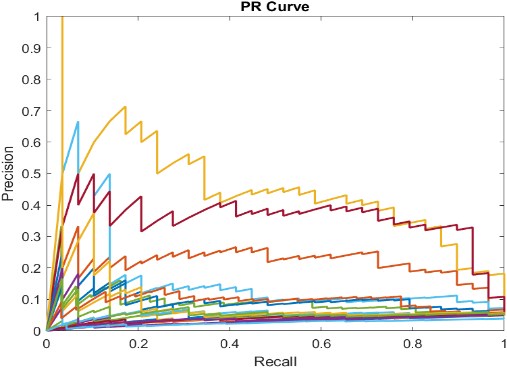
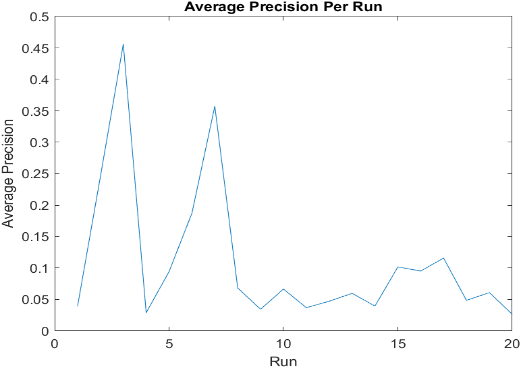
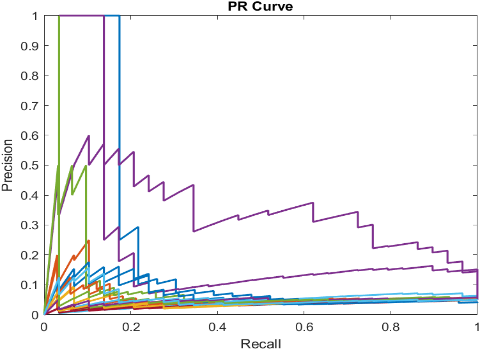
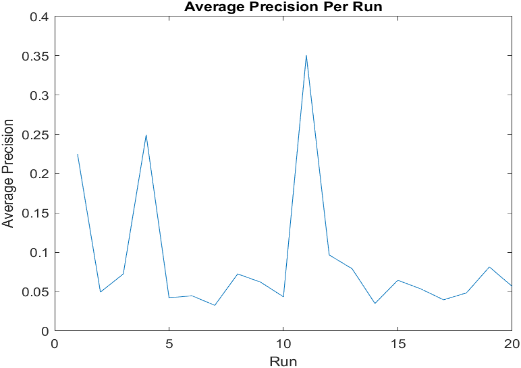
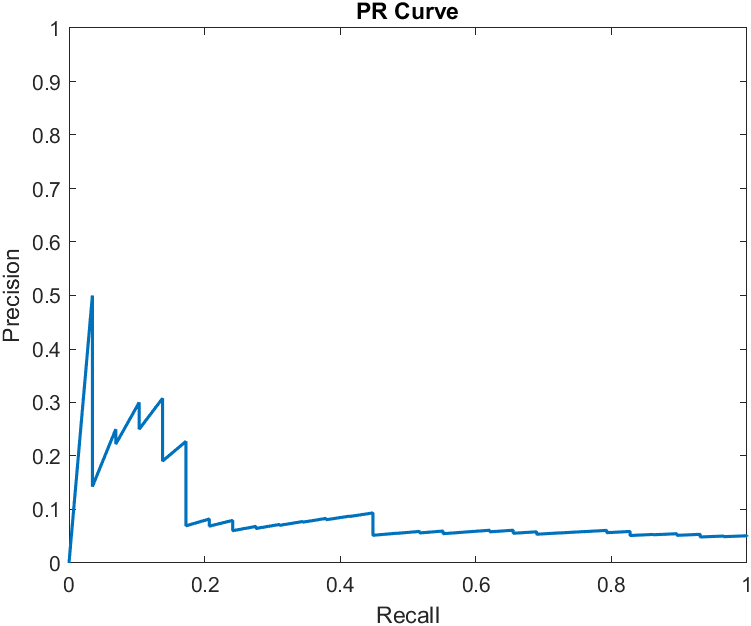
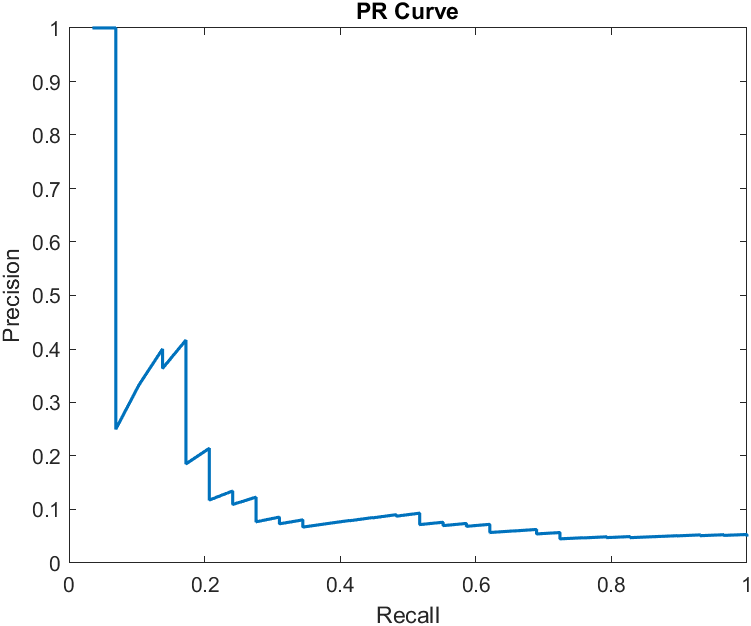
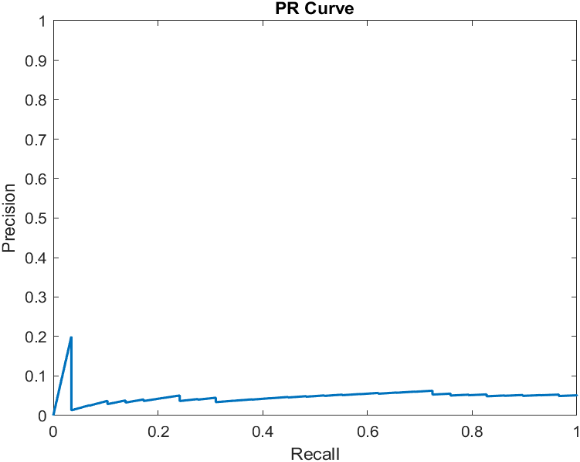
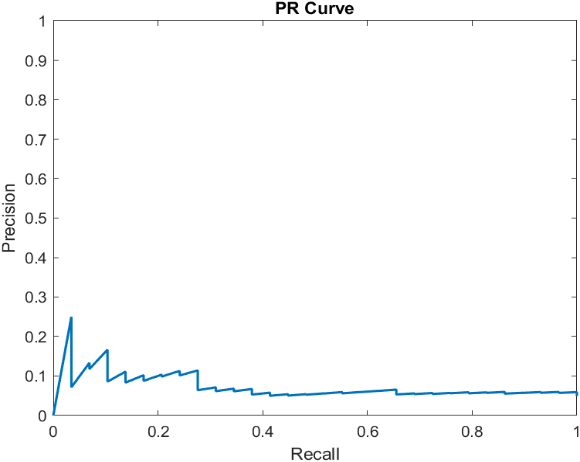
This section will present various experiments and comparisons conducted using the descriptors, distance measures and classification strategies discussed previously.

## L1 vs L2 for GCD

For this experiment, a GCD with 4 and 12 bins was used to compare the LS and L1 norms.

At 4 bins the L1 norm performs marginally better than the L2 norm. At 12 bins the difference in performance is greater. When run over 20 queries with bins at 12, the L1 norm reaches higher values of precision earlier while the L2 norm displays peak precision values after about 10 queries have been passed. Also, to be noted, from the cumulative PR curves it is observable that L1 reaches peak precision 3 times, while L2 reaches precision value of 1 just once. These results are observable in figure 1.

Figure 1: L1 vs L2 Norm

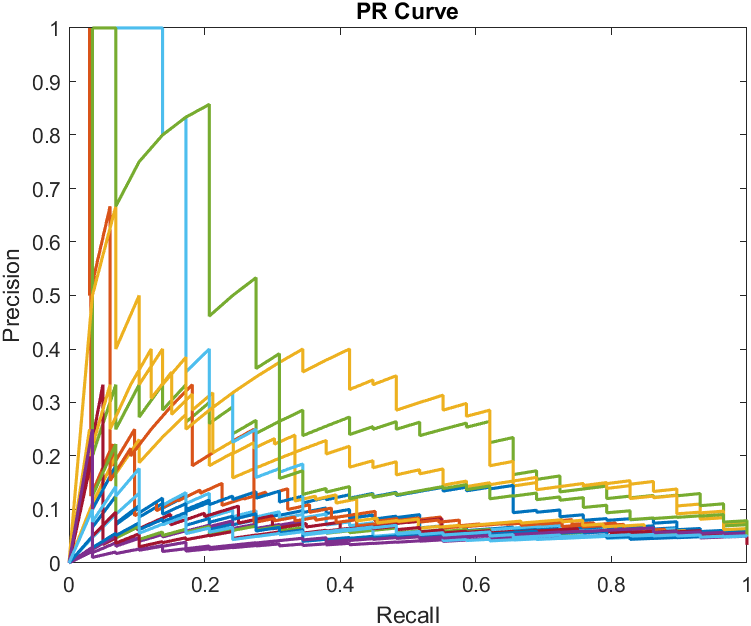
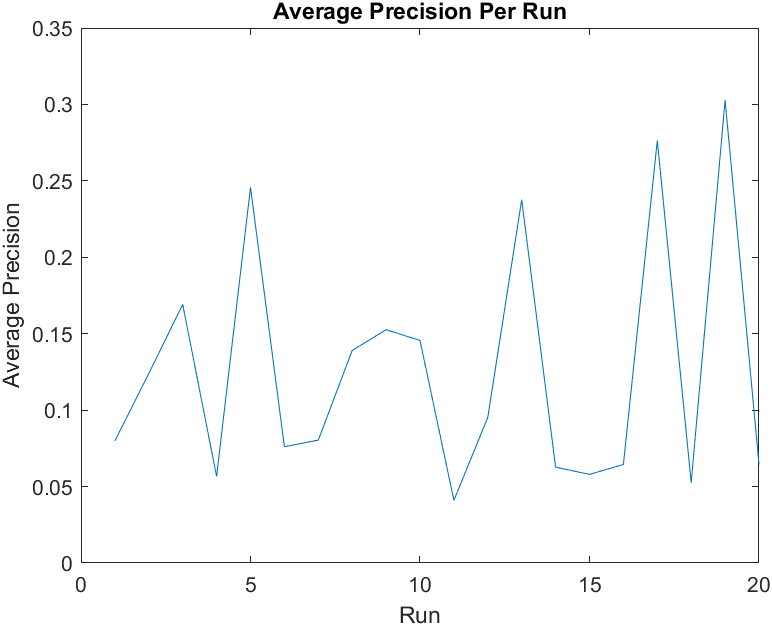
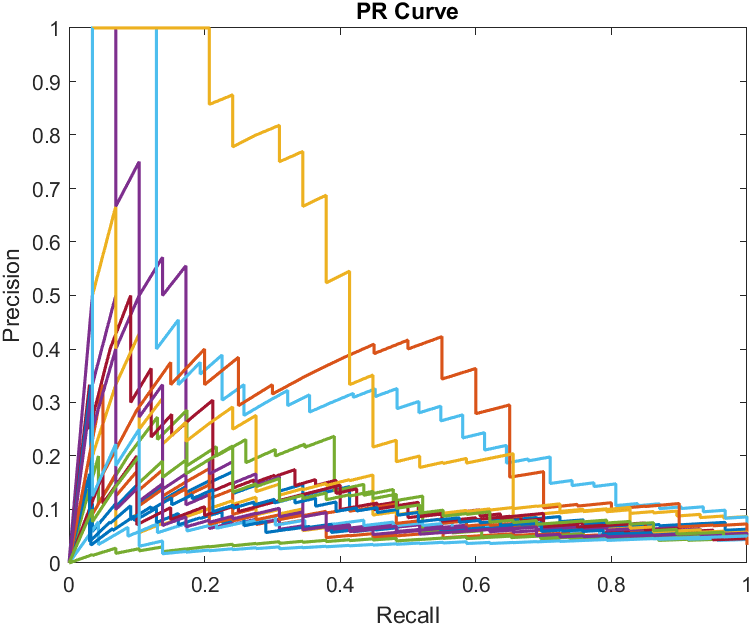
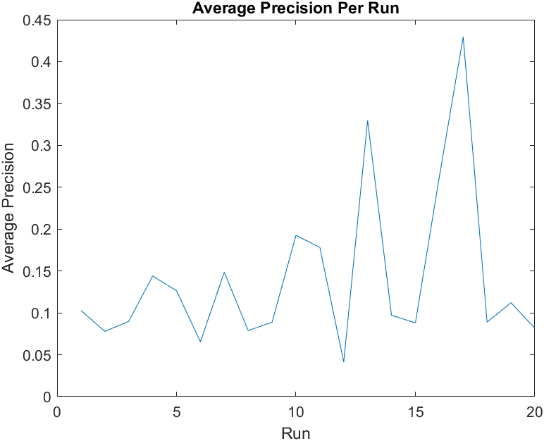


## Colour vs Texture Grid

The first observation that can be made for Colour vs Texture grids is the time it takes to develop descriptor. Texture grid takes considerably more processing time. This might be due to the multiple angle and intensity calculations required, Both the models were built using 4 as grids and L2 norm. Additionally, for Texture Grid, the angles were divided into 8 bins and threshold was set as 0.15. Both PR and MAP plots display similar properties for the colour and texture grids descriptors. The texture grid shows higher average precision early one but after about halfway both descriptors show similar properties. The PR curve also shows a similar trend.

These results are observable in figure 2.

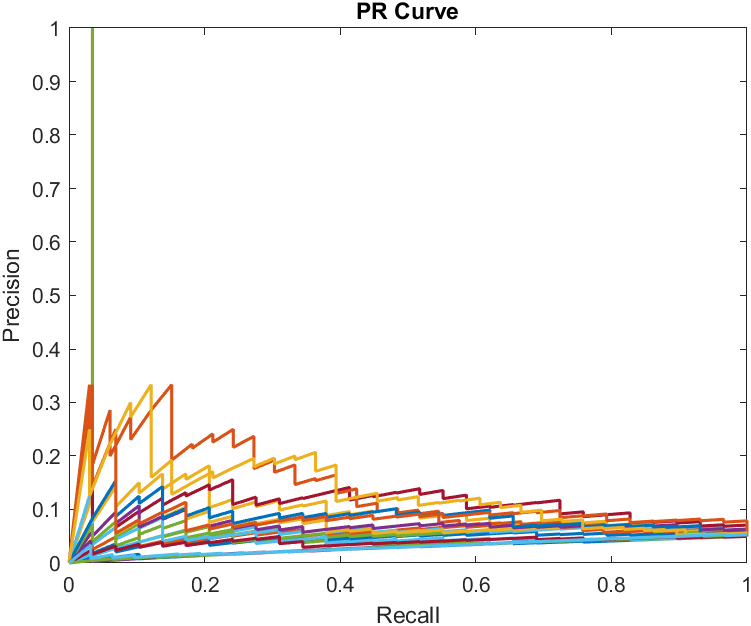
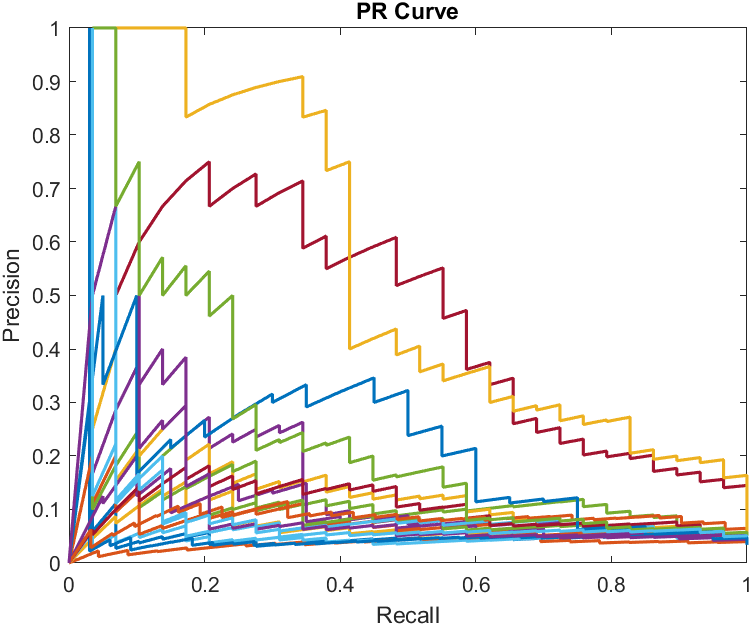
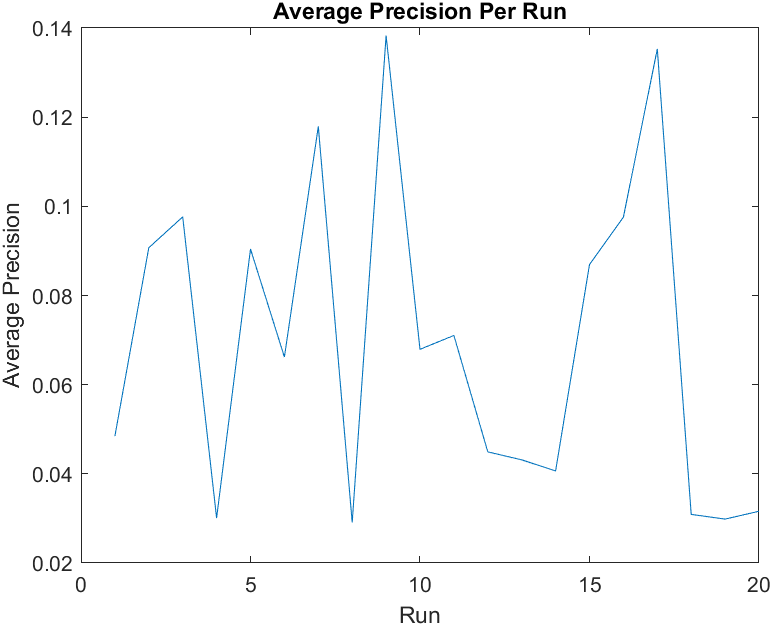
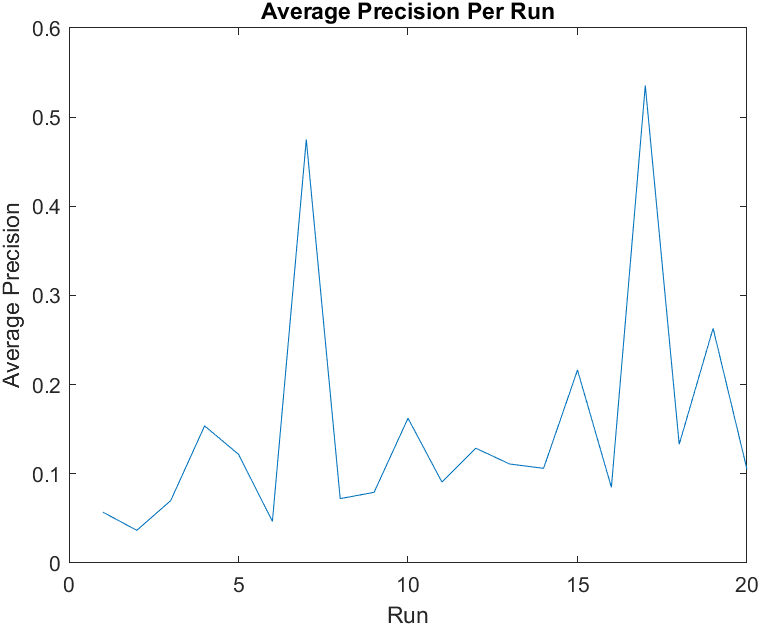
Figure 2: Colour vs Texture Grid



## PCA vs No PCA for Colour-Texture Grid

The colour-texture grid descriptor takes even longer to develop than the texture grid. This is expectable as the quantity of features has increased greatly. To counter this the PCA for this descriptor was evaluated to observe performance. With PCA the processing time is reduced. The non PCA graphs show greater queries reaching precision of 1 while the PCA graphs show how unstable the MAP is over the multiple queries. This is observable in figure 3.

Figure 3: No PCA vs PCA



## SVM

To compute SVM, the dataset had to be first altered. First, images displaying objects were shortlisted and placed into a new folder. In all 355 images were selected belonging to 12 categories. Depending on the descriptor used the SVM can take a good amount of processing time to develop a model. As the complexity of the descriptor is increased so is the processing time, but also the accuracy.

# 6.Conclusion

To summarise, each descriptor has varying efficacy for different applications and the context. Optimum internal parameters for each are data and query dependent but well performing values can be obtained experimentally. Spatial texture techniques were found to perform better than exclusively colour based methods such as a global colour histogram.