Visual Search of an Image Collection

EEE3032 – Computer Vision Pattern Recognition

ELENA MYLONA

6843123

University of Surrey

2023

Abstract

This report delves into a comprehensive analysis of a visual search system applied to the MSRCv2 dataset, focusing on a diverse array of image descriptors such as the Global Colour Histogram, Spatial Gridding with integrated colour and texture features, and their collective variants. Central to this study is the use of Principal Component Analysis (PCA) for condensing these image descriptors into a lower-dimensional space, thereby enhancing both interpretability and computational efficiency. The investigation thoroughly evaluates a range of distance metrics, including the Mahalanobis distance, Minkowski distance, and specifically the L2 norm (Euclidean distance) and L1 norm, to determine their effectiveness in measuring image similarities. This research also undertakes the development of a Bag of Visual Words model, employing a Harris detector, as a strategic approach. In addition, the system's performance is rigorously assessed using precision-recall curves, where similarity is defined in terms of object categories. This approach includes the computation of a confusion matrix, providing a detailed insight into the categorisation accuracy of the system. The research also highlights the importance of reducing the dimensional complexity of descriptors to improve the image retrieval system's performance. The findings underscore the significance of dimensionality reduction in descriptors to augment the efficiency of the image retrieval system.

Table of Contents

1. Methodology: Visual Search Techniques

1.1 Data Exploration

For this series of experiments, the Microsoft MSRC version 2 dataset was utilised. This dataset comprises 591 images distributed across 20 distinct categories. An important characteristic of this dataset is the presence of certain similarities and overlaps among its categories, which has a notable impact on the achievable results. For instance, the first category encompasses images featuring animals including horses, sheep, and cows on a greenery farmland with grass scenes despite sheep and cows having separate categories. Similarly, the classes of water elements and boats have predominantly similar feature scenes, despite being distinct categories.

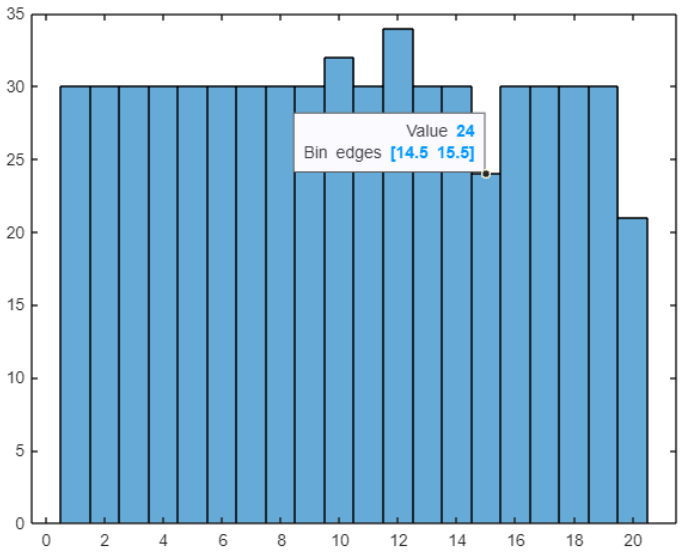


Figure 1: Total Images of Each Image Category

Index	Classification
1	Farm
2	Tree
3	Building
4	Aeroplane
5	Cow
6	Selfie
7	Car
8	Bike
9	Sheep
10	Flower
11	Sign
12	Bird
13	Book
14	Bench
15	Cat
16	Dog
17	Road
18	Water
19	People
20	Coast

Figure 2: Category Indices and their corresponding classes

1.2 Feature Descriptors

To begin with, a descriptor is represented by a column vector composed of numerical values derived from an image. This vector facilitates the representation of an image descriptor as a point within a feature space, which possesses the same number of dimensions as the vector itself. The proximity of images within this feature space suggests a similarity in their descriptors. The approach used to measure the distance between these points is key to how the images are ordered or ranked.

1.2.1 Global Colour Histogram

The global colour histogram (GCH) is a representation of an image's overall colour distribution. It functions by analysing the RGB values of the image's pixels and mapping them onto a three-dimensional plot. This plot is then segmented into uniformly sized sections, known as bins, in a process referred to as quantization, which involves categorizing points within these bins. After quantization, a histogram is generated, counting the quantity of points within each bin. To maintain consistency across images of different sizes, this histogram is subsequently normalised.

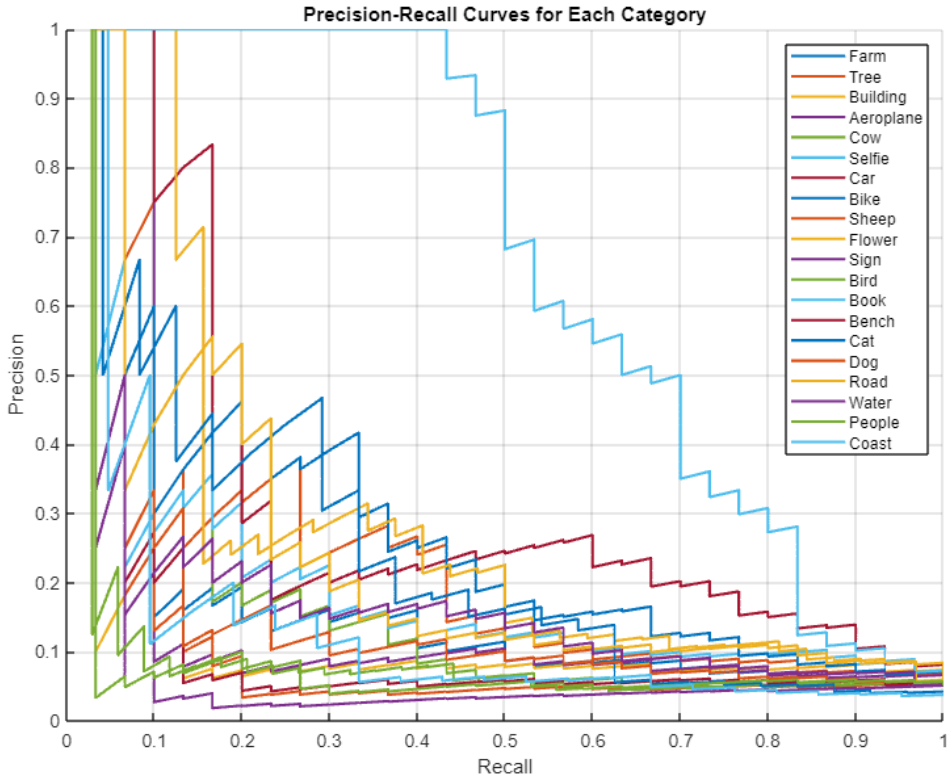


Figure 3: Precision-Recall Curves for Image Categories using Global Colour Histogram Descriptor

Observations:

The global colour histogram, while serving as a descriptor, is not always dependable because it solely captures the general colour distribution of an image without providing any insight into its spatial attributes. As a result, while being highly compact, it falls short in offering discriminative information as illustrated in Figure 3, limiting its effectiveness as a comprehensive descriptor *(See Appendix)*.

RGB Quantization Levels

As illustrated in Figure 4, for quantization level of 9, a significant portion of the test categories yielded enhanced performance compared to other quantization levels, as evidenced by the average precision (AP) of each category and the mean average precision (MAP) of the aggregate outcomes. At elevated quantization levels, greater than 9, pixel values are segmented into an extensive array, making it challenging for the image descriptor to precisely distinguish unique attributes. Conversely, at a lower quantization level such as 3, the RGB values are limited to merely two distinct options, leading to suboptimal results.

Table 1: Average Precision and MAP for Global Colour Histogram

Category	Q=3	Q=6	Q=9
Aeroplane	0.29822	0.34666	0.3173
Car	0.47354	0.41803	0.42392
Bike	0.28452	0.36552	0.42621
Tree	0.3319	0.41983	0.41987
MAP	**0.2692**	**0.3579**	**0.3721**

Figure 4: AP and MAP of different quantization levels on certain image categories

1.2.2 Spatial Gridding with Colour and Texture

Spatial Colour

Regarding the spatial colour analysis, evaluation trials were conducted on category responses as the dimensions of the grid were altered between 4x4, 8x8, 16x16 *(See Section 2.1)*. The array of outcomes was graphically represented as a surface to observe the impact of grid divisions adjustments on the mean average precision. The precision-recall curve representing the average was charted for the parameter settings that yielded optimal performance in these tests. These examinations utilised the L2 norm measure.

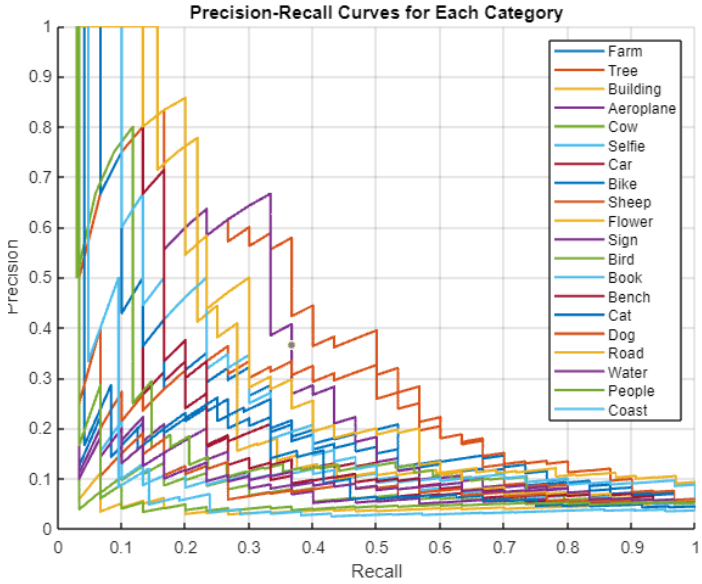


Figure 5: Colour Gridding curves indicate the variability in performance and descriptor' discriminability.

The Spatial Colour Grid descriptor segments an image into a predefined grid arrangement, proceeding to ascertain the mean RGB values for each segment. This process culminates in the formulation of the feature vector.

The categories represented by curves closer to the top-right corner of the graph (higher precision and recall) are performing better. In this case, categories such as "Aeroplane," "Bike," and "Car" appear to perform well, indicating that the spatial colour gridding descriptor is effective for these types of images. Conversely, categories with curves closer to the bottom-left corner, such as "Coast" or "People," have lower precision and recall, suggesting less effectiveness.

Spatial Texture

For the spatial texture analysis, the grid sizes selected were based on those that yielded high results from the spatial colour studies to assess the impact of changing the bin count and threshold value. While examining one parameter, the other was kept constant. After determining the optimal values for each, these were then applied to explore the influence of altering grid sizes. The range for grid size adjustment was limited compared to the spatial colour due to the extensive processing time required. These evaluations were meticulously executed utilizing the Euclidean metric.

Figure 6: Comparison of Original Image and Its Edge Angle Magnitudes with a Threshold Value of 0.02

The initial threshold value, which was kept constant during the bin count examination, was chosen after reviewing various threshold levels, with examples depicted in figure 6. This thresholding process results in a high-contrast image that isolates the edges of the flower, transforming them into prominent white lines against a black background, indicative of the edge angle magnitudes captured by the edge detection algorithm.

1.2.3 Merged Colour and Edge Orientation Histogram

The spatial texture parameters and grid dimension settings were consistently applied across both colour and texture descriptors when they were integrated, ensuring comparability of outcomes. Similarly, the variation in grid dimensions was maintained within the same predefined ranges. Every associated experiment employed the Euclidean Distance metric as the standard for assessment. A threshold of 0.07 was ultimately selected, with its visualisation presented in figure 7.

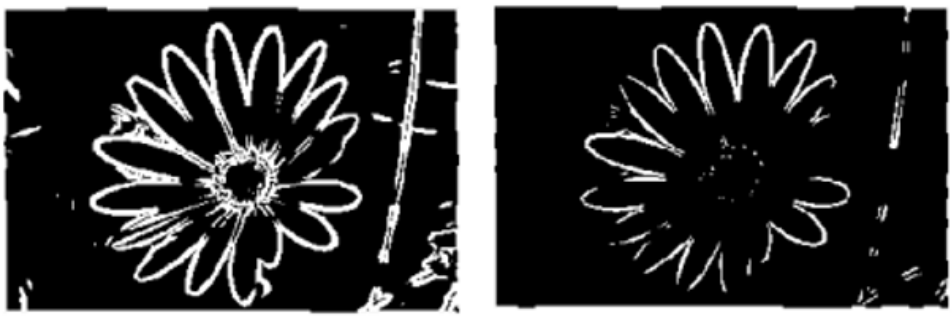


Figure 7: Edge Detection on a Flower Image at Two Different Levels, 0.07 and 0.2

1.3 Principal Component Analysis (PCA)

In pursuit of establishing a representative model for the dataset, the covariance across all descriptors was computed. This process facilitates the recognition of dimensions exhibiting minimal variance, which can then be eliminated by organising the eigenvectors in accordance with the eigenvalue magnitudes. Subsequently, the Mahalanobis distance was determined for each image pair.

The percentage energy reduction refers to the fraction of eigenvalues that are excluded from the model, and this reduction parameter can be modulated. To elucidate the impact of alterations in dimensionality reduction, descriptors employing previously established parameter configurations were subjected to a spectrum of energy reduction levels. This approach was undertaken to scrutinise its influence on the attainable mean average precision. The results correlating Mean Average Precision (MAP) with Energy Reduction are depicted in the following figures 8 and 9.

Mean Average Precision	Percentage Energy Reduction
0.1885	0.3
0.2017	0.4
0.2067	0.5
0.2217	0.6
0.2222	0.7
0.2329	0.8
0.2414	0.9
0.2377	1.0

Figure 9 Impact of Dimensionality Reduction on MAP, starting from 0.3% Energy Reduction Onwards

Percentage Energy Reduction	Mean Average Precision
2.5	0.2353
2.6	0.2274
2.7	0.229
2.8	0.2271
2.9	0.2261
3.0	0.2173
3.1	0.2172
3.2	0.218
3.3	0.2114
3.4	0.2234

Figure 8 Analysis of MAP Across Increasing Percentage Energy Reduction Levels, slightly increased and then decreased.

The efficiency of PCA is assessed for various descriptor configurations: a Global Colour Histogram at quantization level of 9, a spatial colour grid cell of 4x4, a spatial texture patch of 4x4, and an 8x8 cell that integrates both spatial colour and texture. The comparative performance metrics, detailed in figure 10, showcase the descriptors' values prior to PCA and subsequent to the application of PCA within the visual search framework.

Table 5: AP and MAP for Visual Search Before and After PCA for Different Image Descriptors

Category	GCH		Colour		Texture		Colour and Texture	
	No PCA	PCA	No PCA	PCA	No PCA	PCA	No PCA	PCA
Aeroplane	0.3173	0.28035	0.61782	0.56411	0.32911	0.46211	0.12036	0.45751
Car	0.42392	0.45632	0.20543	0.18695	0.45192	0.54849	0.12834	0.72267
Bike	0.42621	0.32489	0.35524	0.28102	0.12322	0.15408	0.21091	0.32798
Tree	0.41987	0.37378	0.54884	0.50291	0.56543	0.64652	0.20989	0.23101
MAP	**0.3721**	**0.3197**	**0.3610**	**0.31869**	**0.3744**	**0.3498**	**0.1368**	**0.4281**

Figure 10: Comparative Average Precision Before and After PCA Across Image Descriptors

The above figure indicates that the outcomes with PCA are generally consistent with those obtained without it. Nonetheless, a notable distinction is observed with the highly dimensional descriptor, specifically the 8x8 spatial colour and texture, where there is a marked enhancement in the results.

1.4 Distance Metrics

After identifying spatial parameters that deliver robust performance, mean average precision (MAP) scores were recorded using L1, L2 norms, and Mahalanobis distances. These scores were then utilised to evaluate and contrast the effectiveness of each distance metric alongside various descriptors, as well as to provide a conclusive overview of each descriptor's relative performance *(See 2.2)*.

1.4.1 Mahalanobis distance

The Mahalanobis distance refines the concept of the Euclidean *(L2 norm)* by taking into consideration the characteristics of a data model as established through principal component analysis. Rather than merely computing the distance between two points within a feature space, the model delineates a specific zone in that space. A more pertinent measure is the distance from a point to the nearest boundary of that zone.

This distance is quantified by the Mahalanobis metric by determining how many standard deviations a point is from the mean of the model, thereby adjusting for the model's

configuration within the feature space. When transitioning from the original reference frame to the one outlined by the PCA-based model, the Mahalanobis distance is analogous to the Euclidean distance. Essentially, it measures how far a point is from the model's perspective within its own reference frame.

1.4.2 Euclidean (L2 norm)

The L2 norm, commonly known as the Euclidean distance, represents the most direct path between any two points in a given space and can also be described as the length of a vector. For instance, in a three-dimensional Euclidean space, the length of a vector, x = (i, j, k), is calculated by the formula √ (i^2 + j^2 + k^2). This straightforward method of measuring distance, often likened to the linear path a bird might take in flight, is the most widely adopted norm for distances in Euclidean spaces.

1.4.3 Manhattan (L1 Norm)

The L1 norm, also known as the Manhattan distance, is determined by summing the absolute values of a vector's components. For a two-dimensional vector x = (i, j), the L1 norm is computed as the total of the absolute values of 'i' and 'j', which is represented by |i| + |j|. This norm reflects the total distance traveled along each axis to move from the starting point to the endpoint in a grid-like path, similar to navigating the streets of a city laid out in a grid pattern.

1.4.4 Minkowski (3rd high order)

Minkowski space with p=3 dimensions consists of a 3D vector space equipped with a metric tensor that gives rise to one timelike dimension and two spacelike dimensions. This spacetime represents the geometric structure of 2+1-dimensional special relativity, where events are specified by three coordinates (t, x, y). It represents a measure of distance that emphasizes differences in the higher-order dimensions more than the lower-order dimensions. In other words, it gives more weight to larger differences in individual components of the vectors.

It possesses homogeneity and isotropy, with no preferred origins. The light cone structure divides intervals into timelike, spacelike or lightlike. Proper time along worldlines represents the elapsed time for an observer following that path. Minkowski space obeys Lorentz transformations between inertial frames and exhibits a maximum speed limit c. The metric can be written with explicit indices as ημν = diag(1,-1,-1).

2. Further Analysis of Experimental Results

In the ongoing quest to refine image retrieval systems, a detailed analysis of experimental results is essential. The following section delves into the comparative performance of various image descriptors and their associated distance measures. By scrutinising mean average

precision (MAP) scores across a range of experimental conditions, we aim to discern the most effective combinations for accurate and efficient image retrieval.

2.1 Gridding and Angular Parameters

Gridding refers to the method of dividing an image into smaller, manageable regions or cells, which can then be individually analysed and compared. The Average Precision (AP) for different grid sizes suggests how well the gridding technique can capture and utilise spatial and textural information for image retrieval.

Angular parameters usually pertain to the orientation and rotation aspects within image descriptors. They can play a significant role in how textures, edges, and other spatial features are represented and recognised by the retrieval system.

A more refined grid, like a 4x4, offers a balanced trade-off between detail and computational efficiency, capturing enough spatial detail to accurately describe and match images without being overly sensitive to variations as a finer grid, like a 2x2, might be. Conversely, a coarse grid such as 16x16 may overlook important details that are essential for distinguishing between similar images, leading to less precise search results.

2.1.1 Merged Colour and Texture Angular Descriptor

Spatial Textural Grid Cells:

This method begins by processing coloured images to extract texture features, which entails specifying the grid size for segmentation. Initially, the image is transformed into a grayscale version to simplify the extraction process. Sobel filters are then applied to this grayscale image for edge detection. Following this, the algorithm computes the gradient magnitudes and orientations. It proceeds to process the information within each cell of the grid, collecting data on the magnitude and direction of the gradients and compiling this data into a histogram. Weak

Category	2x2	4x4	8x8	16x16
Aeroplane	0.17736	0.46211	0.2198	0.08084
Car	0.42460	0.54849	0.28386	0.11651
Bike	0.14188	0.15408	0.15791	0.14503
Tree	0.80643	0.64652	0.22192	0.15037
MAP	**0.3417**	**0.3744**	**0.2046**	**0.1065**

Figure 11: AP and MAP for Granularity of Textures within Grid Cells

edges are subsequently filtered out to maintain the integrity of the data. The orientations are adjusted to fall within a [0, 2π] range and categorized into 8 distinct bins for further analysis. This corresponding Average Precision and Mean Average Precision are illustrated in Figure 11 of the report.

Employing a 4x4 grid for segmenting an image and subsequently computing the Edge Orientation Histogram on each segment significantly enhances the precision of image analysis, surpassing that of alternative grid dimensions. This enhancement may stem from the fact that a 4x4 grid captures edge details more effectively, thus more accurately delineating the defining features of the image. Conversely, a grid measuring 16x16 fails to yield comparably sharp edge details. Within the context of spatial texture, the category of trees

exhibited the highest precision. This is attributed to the homogeneity of leaf patterns among the trees in the dataset, which resulted in a consistent textural representation.

Category	2x2	4x4	8x8	16x16
Aeroplane	0.50343	0.55802	0.36544	0.12036
Car	0.49141	0.49041	0.29863	0.12834
Bike	0.29453	0.30714	0.29652	0.21091
Tree	0.86013	0.7642	0.25947	0.20989
MAP	**0.4599**	**0.4379**	**0.2622**	**0.1368**

Figure 12: Grid Configurations to Capture Coloured Spatial and Textural Information of Images

The investigation deduces that utilising a 4x4 grid configuration yields superior search precision across the assessed spectrum of image categories, as indicated by their respective AP metrics. Notably, in figure 12, the 2x2 grid's MAP demonstrates a deceptively higher accuracy, an effect predominantly influenced by the disproportionately high AP value associated with the 'Trees', a similar trend observed within the spatial texture descriptor's performance for this category. Conversely, the 16x16 grid exemplifies the least effective grid size for visual search tasks, corroborating the findings related to the inefficacy of this descriptor.

2.2 Distance Metrics Evaluation

The following table of figure 13, delves into the core of several distance metrics, evaluating their performance in discerning the proximity between multidimensional data points. By comparing metrics such as Euclidean, Manhattan, and Mahalanobis distances.

Category	GCH		Colour		Texture		Colour Texture	
	L1	L2	L1	L2	L1	L2	L1	L2
Plane	0.37181	0.3173	0.65528	0.61782	0.5079	0.46211	0.62927	0.55802
Car	0.48345	0.42392	0.20991	0.20543	0.5740	0.54849	0.5100	0.49041
Bike	0.5494	0.42621	0.35941	0.35524	0.1463	0.15408	0.32686	0.30714
Tree	0.47084	0.41987	0.5613	0.54884	0.8458	0.64652	0.88149	0.76425
MAP	**0.4018**	**0.3721**	**0.3830**	**0.3610**	**0.4401**	**0.3744**	**0.49755**	**0.4379**

Figure 13: Average Precision and Mean Average Precision Across Various Distance Metrics

A side-by-side comparison of Average Precisions and Mean Average Precisions for six specific query images within the dataset is depicted on the above table. Notably, the MAP values are consistently superior when employing the L1 Norm in comparison to the L2 Norm. This finding implies that the L2 Norm might not be the optimal choice for visual search tasks.

2.3 Precision-Recall Curves and Confusion Matrix

The following figure 14 depicts a Precision-Recall Curve which serves as a graphical representation of a model's performance. Before applying Principal Component Analysis (PCA), this PR Curve illustrates the effectiveness of a Colour Texture Grid Descriptor. The various coloured lines on the graph likely represent different parameter settings or instances of the descriptor.

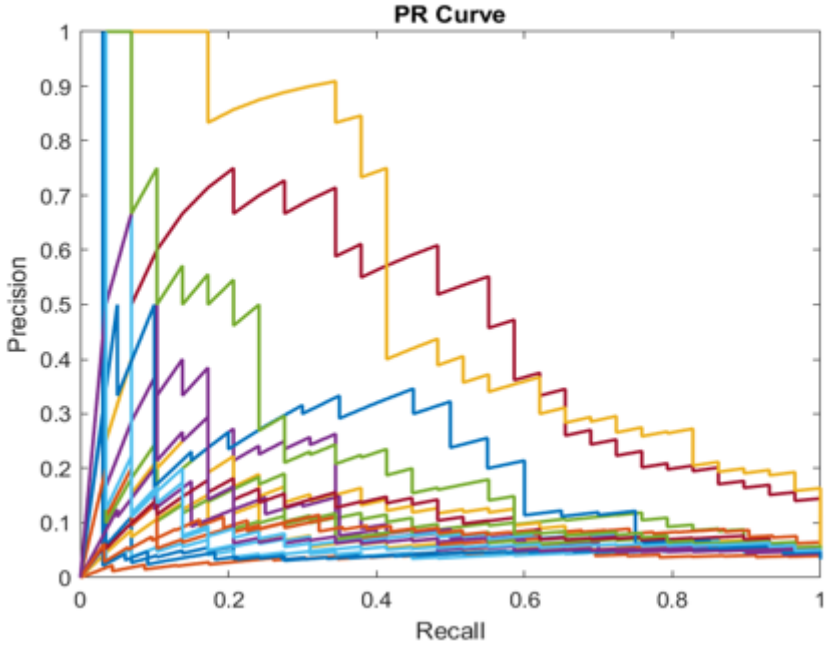


Figure 14: Observations of Image Descriptors in a High Dimensional Space without PCA

The graph shows a common trend where precision of Image Descriptors tends to decrease as recall increases, which is typical in retrieval systems. Descriptors with curves closer to the top-right corner indicate better performance, as they maintain high precision at higher levels of recall.

The confusion matrix, which is a tool used to measure the performance of classification models, indicates varying degrees of classification accuracy when PCA is not applied. Specifically, there is an increased misclassification of signs as bookshelves, rising from 32% to 40%. This is despite an improvement in the accurate identification of the correct category, which has risen from 44% to 48%. On a different note, there is a reduction in the mislabelling of sheep as birds, which has decreased from 44% to 32%. Conversely, the general misclassification of farm animals as birds has seen a notable increase, moving from 12% to 28%. These shifts in classification accuracy highlight the complex interplay of factors that affect the performance of classification without dimensionality reduction through PCA.

Following this, the incorporation of PCA into the processing workflow has been instrumental in reducing the time required for computation. Nonetheless, this increase in efficiency is accompanied by a trade-off in performance stability. Despite this improvement, there exists an uptick in the incorrect categorisation of farm animals as birds.

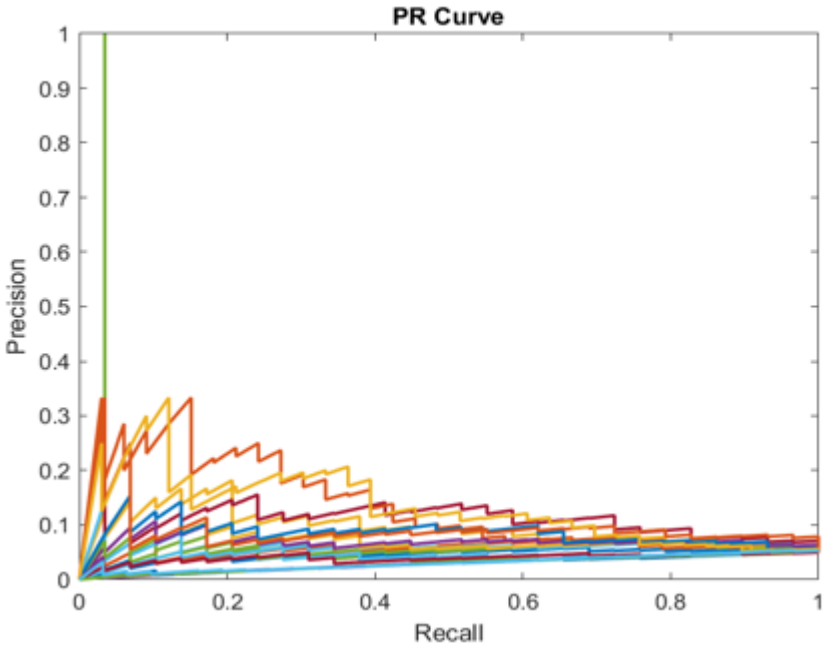


Figure 15: Observations of Image Descriptors in a Reduced Dimensional Feature Space with PCA

The application of PCA, while expediting the process, results in a more erratic Mean Average Precision (MAP) over various queries, demonstrating that the speedier processing may adversely affect the consistent accuracy of the classification outcomes.

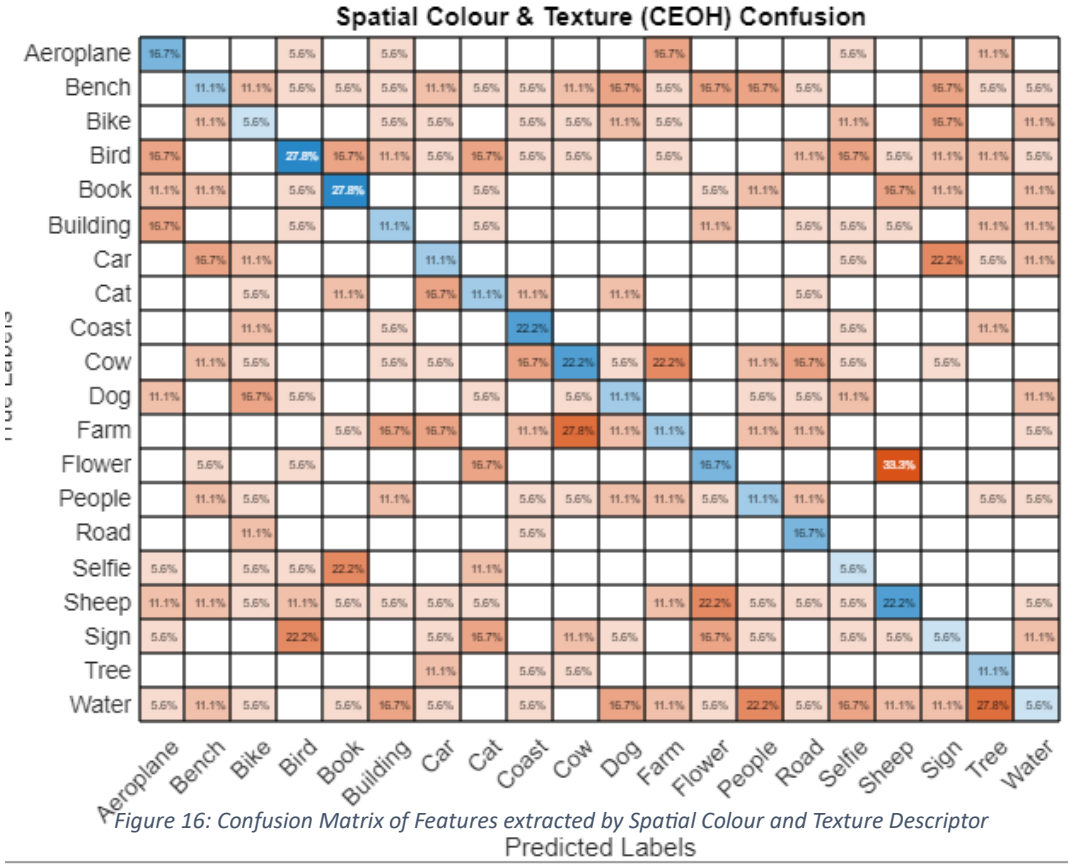


Figure 16: Confusion Matrix of Features extracted by Spatial Colour and Texture Descriptor

The provided confusion matrix visualises the performance of an image retrieval system using a Spatial Colour & Texture descriptor. Each cell in the matrix represents the percentage of times a certain category (true label) was predicted as another category (predicted label). High percentages on the diagonal from the top-left to the bottom-right would typically indicate correct classifications, as these cells represent instances where the predicted category matches the true category.

From the confusion matrix of figure 16, several interpretations can be conducted:

1. Certain categories like "Aeroplane", "Bird", "Car", and "Flower" have relatively high self-recognition, as indicated by higher percentages along the diagonal, which means the system is fairly accurate in retrieving these categories.
2. There are notable confusions between some categories. For example, "Sign" is frequently misclassified as "Book", and "Selfie" is often confused with "Flower" and "Road". These misclassifications could be due to similarities in colour and texture features between the confused categories.
3. Some categories like "Cow" and "Farm" are also commonly misclassified as "Bird", which suggests that the system may struggle to distinguish between certain types of images, possibly due to overlapping feature descriptors.
4. "Tree" is another category with a high rate of correct classification, but it also has a notable rate of being confused with "Building".

Comparative Analysis of Descriptor Performance in Image Retrieval:

In this finding, the highest mean average precision values for various descriptors were compared. For spatial descriptors, a standard 4x4 grid was used, and for spatial texture descriptors, a 7-bin count and a 0.10 edge magnitude threshold were implemented for uniform comparison. Adding spatial texture typically resulted in a boost in mean average precision compared to using only colour-based descriptors, with the notable exception being the Mahalanobis distance for spatial texture, which exhibited the lowest value. The Manhattan distance measure was the most effective overall, outperforming others in every category except spatial colour. The PCA and Mahalanobis distance ranked as the second-best distance measure in all but the spatial texture category. In contrast, Euclidean distance was less effective in all categories, except for spatial texture. Notably, the use of Manhattan with a merged spatial colour and texture descriptor yielded an approximately 50% higher mean average precision than a global colour histogram with Manhattan distance.

Corollary

3.1 Derivations from Experimental Outcomes

The study found that when these descriptors are subjected to a slight dimensionality reduction through Principal Component Analysis (PCA), there is a notable improvement in performance. PCA is a statistical technique used to simplify the complexity of data, focusing on the most significant elements by transforming the original data into a new set of principal components. This reduction, although minor, is significant in enhancing the system's efficiency without substantially losing essential information.

Comparatively, the use of the L2 norm, which is the standard approach in many image retrieval systems was overshadowed by the performance of the L1 norm during the experiments of combined colour and texture. It revealed that spatial texture techniques outshine methods based solely on colour, like global colour histograms. This indicates that texture features play a significant role in image characterisation and retrieval, potentially offering more detailed and nuanced information than colour alone. Spatial texture methods, which analyse patterns of intensity or colour variations across an image, provide insights into the structure and surface properties of objects within the image. These insights are often more distinctive and informative compared to colour histogram that merely quantify colour distribution without spatial context. While both colour and texture are valuable features, their combined effect does not always produce a linear or additive improvement in retrieval accuracy.

This study's findings suggest that the L1 norm was more effective across various methods tested. This could be due to its characteristic of being less influenced by outliers or significant variances in a few dimensions, as it sums the absolute values of the differences in each dimension. Unlike the L2 norm, which can disproportionately weight larger differences (due to squaring), the L1 norm considers the absolute differences, thus preserving a balanced representation of both colour and texture variations.

3.2 Limitations

A significant obstacle encountered in this project stemmed from the characteristics of the provided MSRCv2 dataset. The way images were categorised posed a challenge, especially considering instances of image repetition and the inclusion of certain objects or features in multiple categories. Consequently, these factors sometimes led to results that did not accurately reflect the true capabilities of a visual search.

To conclude, the efficacy and accuracy of a visual search system are influenced by a multitude of factors. These include the application of various descriptor techniques, such as quantization and gridding, the dimensionality of the feature space, and the selection of distance measures to determine similarity.

3.3 Approximate Completion of Bag of (Dictionaries) Visual Words

In the Bag of Visual Words (BoVW) model implemented in MATLAB, the process initiates by specifying the output directories for the storage of descriptors and the subsequent 'bag' of words. The function progresses by aggregating keypoint data and feature descriptors from the image dataset. It methodically analyses each image, utilising Harris keypoints for the detection of corners, and compiles these keypoints into a comprehensive collection of feature descriptors.

Upon completing the image processing, the function applies k-means clustering to these descriptors to discern a predefined number of cluster centroids, which represent the visual vocabulary. Subsequently, the function evaluates the proximity of each image's descriptors to these cluster centroids, generating a histogram of visual word occurrences for each image. This histogram is normalised, effectively translating the raw visual information into a feature vector per image.

Finally, the resultant feature vectors, indicative of the images' BoVW representation, are systematically saved. This representation enables the numerical analysis of images, facilitating tasks such as classification and retrieval within the computer vision domain.

Bibliography

IEEE/CVF. (2021). Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

IEEE/CVF. (2022). Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.

Gonzalez, R. C., & Woods, R. E. (2017). Digital Image Processing. Pearson.

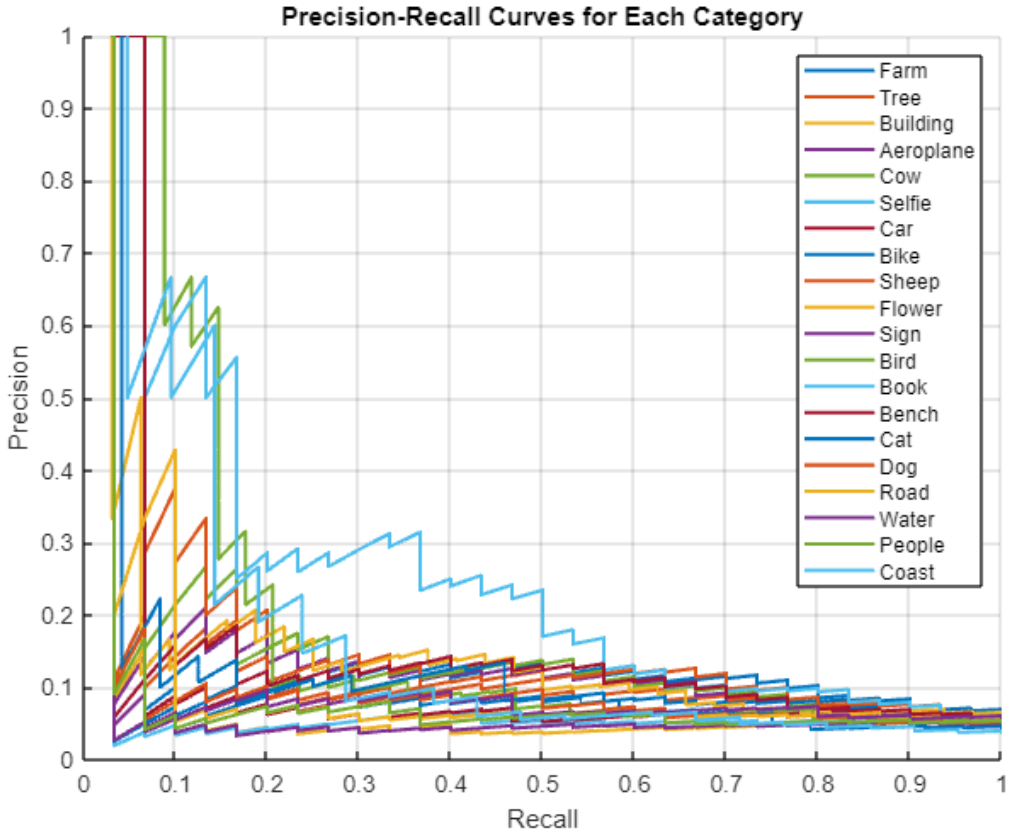
Duda, R. O., Hart, P. E., & Stork, D. G. (2012). Pattern Classification. Wiley.

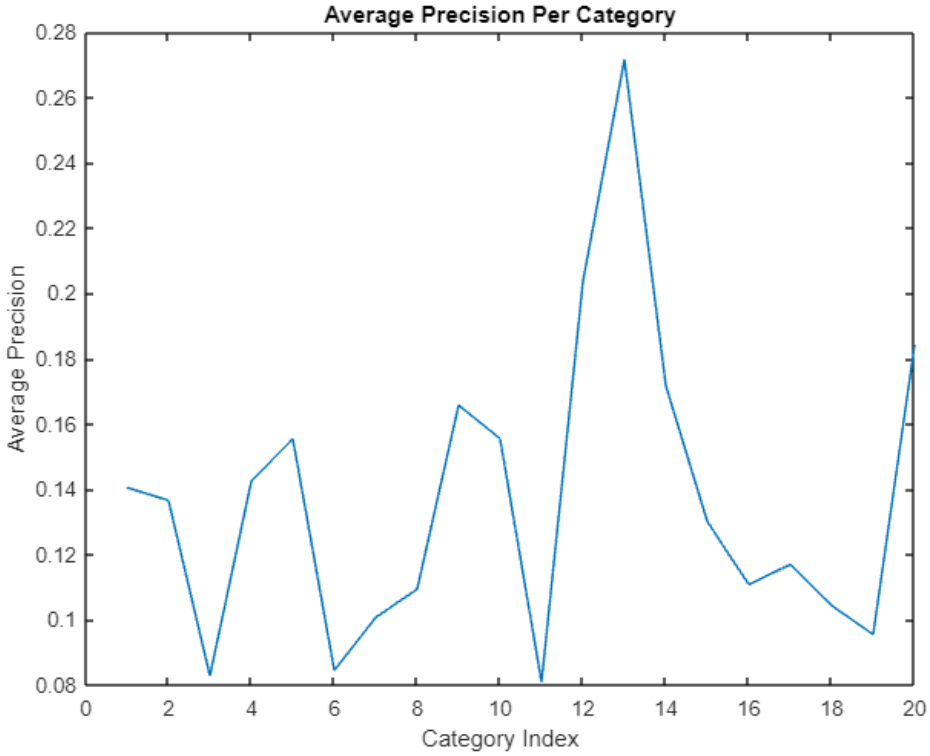
Szeliski, R. (2010). Computer Vision: Algorithms and Applications. Springer.

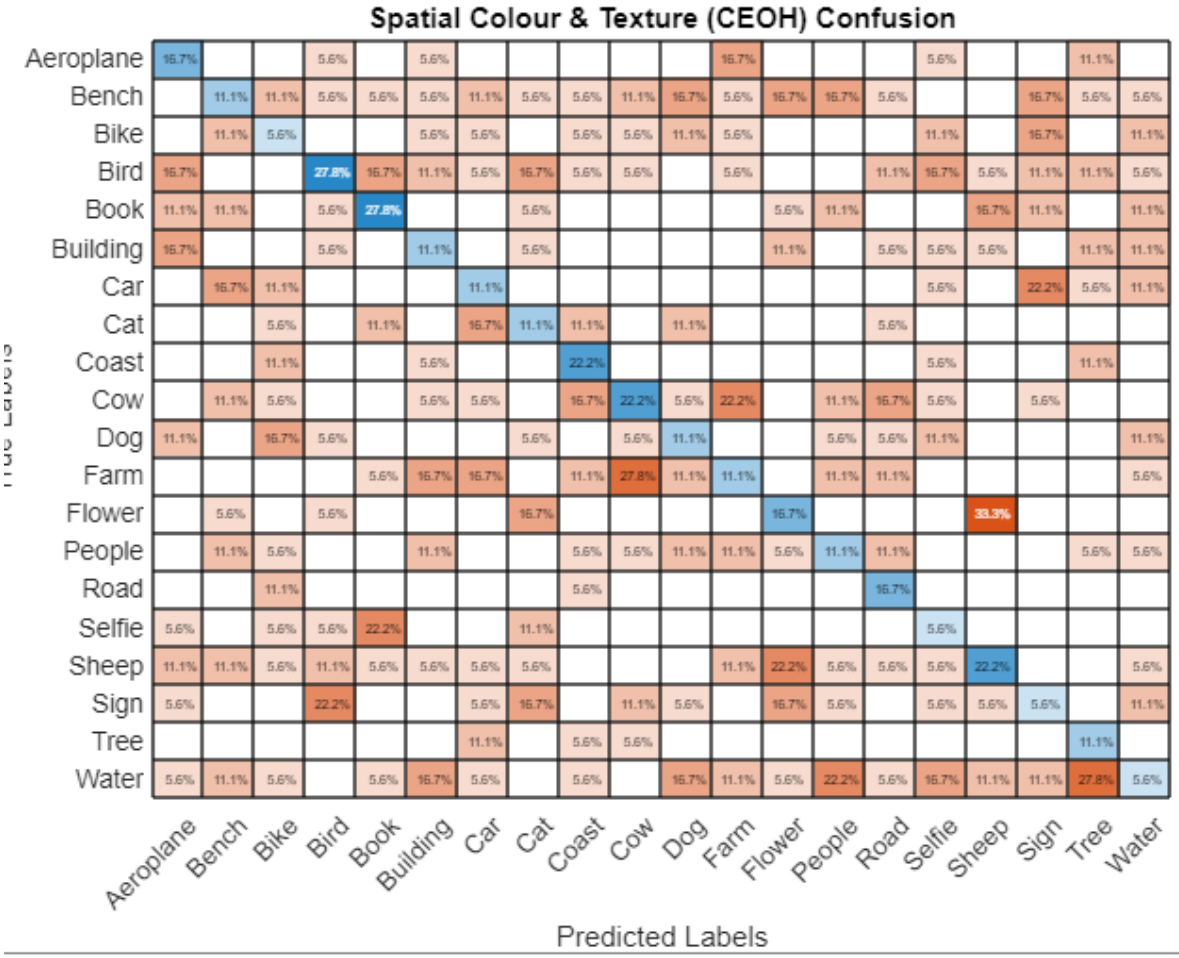
Sonka, M., Hlavac, V., & Boyle, R. (2014). Image Processing, Analysis, and Machine Vision. Cengage Learning.

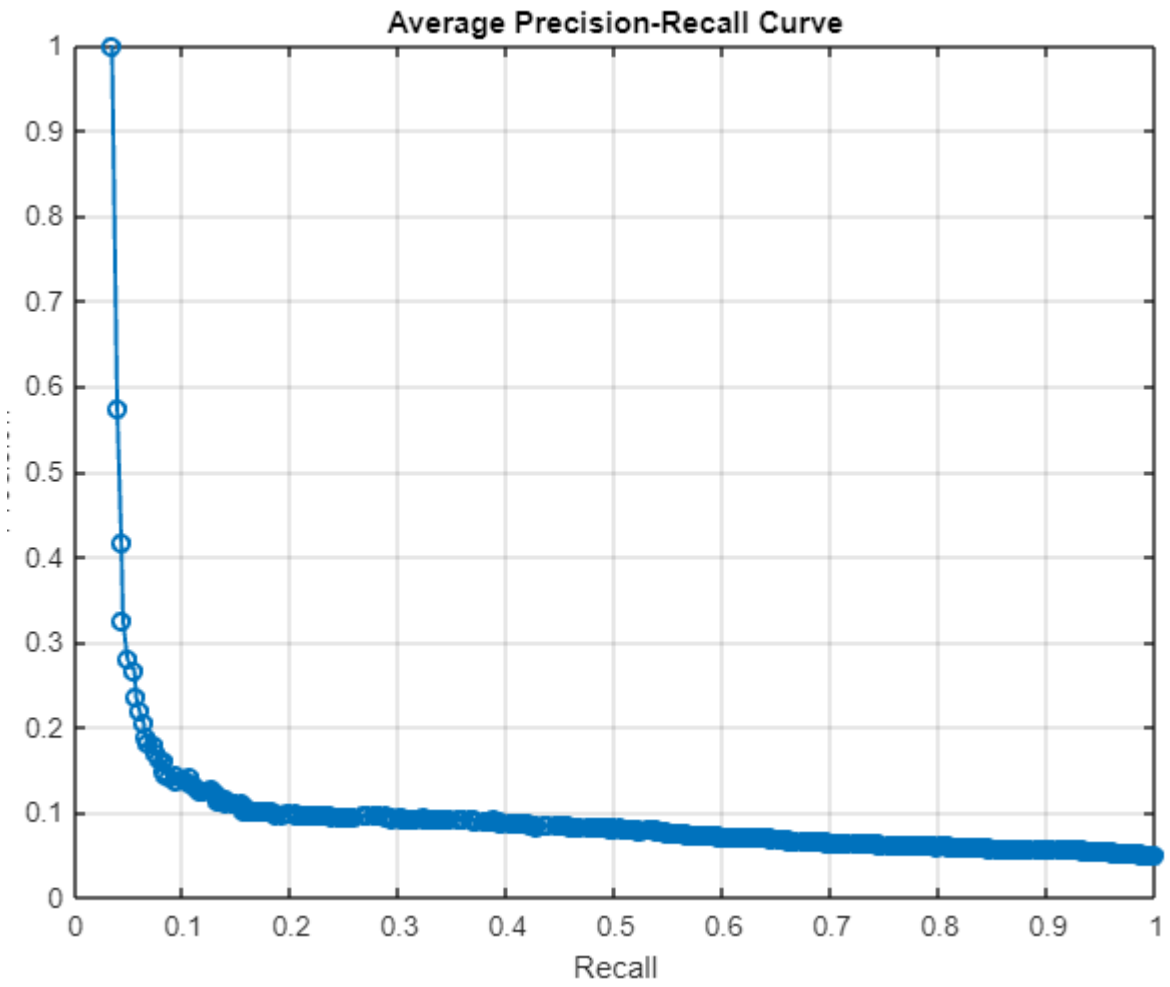
6. Appendix

Merged Spatial Colour and Angular Texture Descriptors

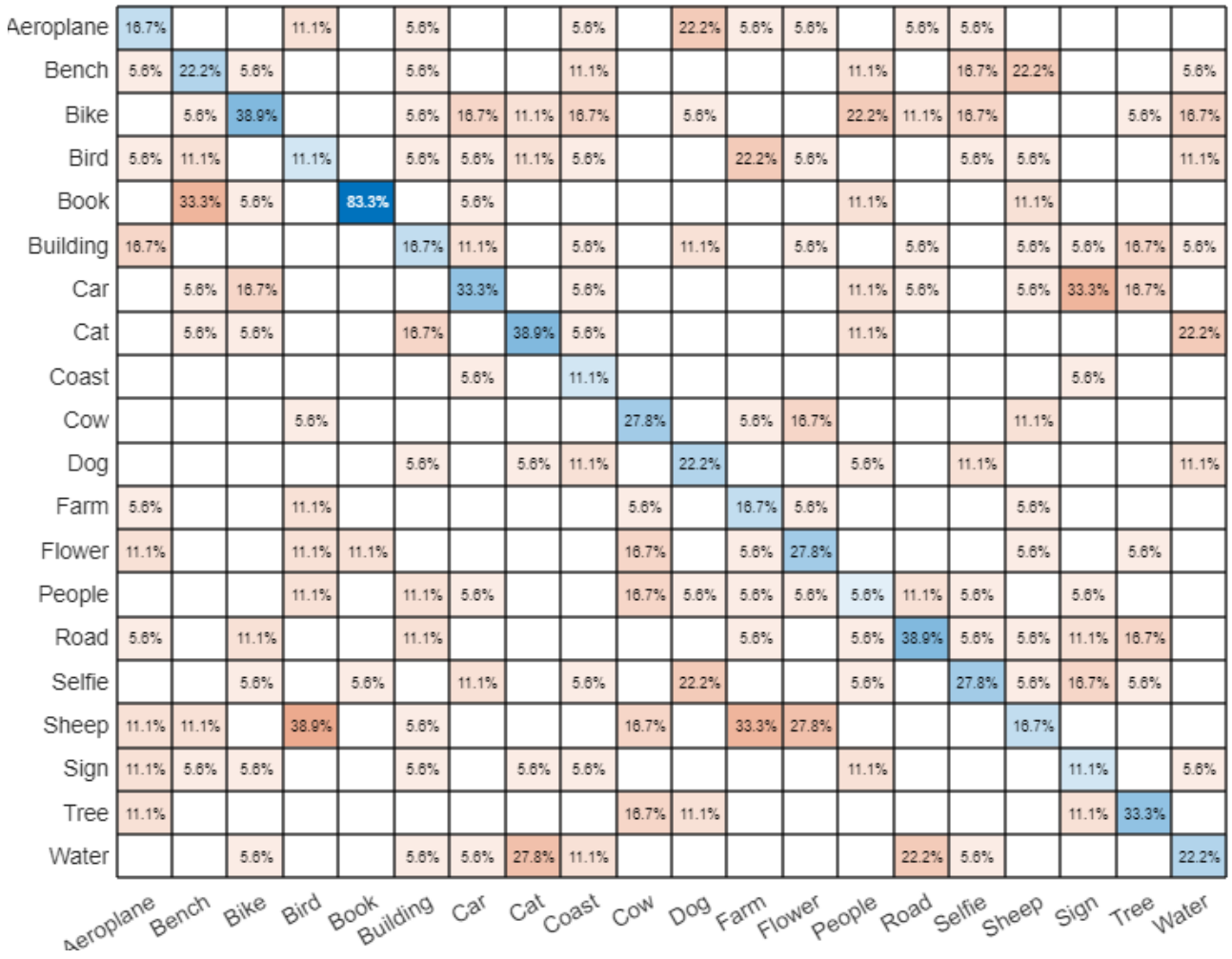



Average Precision Per Category
Average Precision
Category Index
0.28
0.26
0.24
0.22
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0
2
4
6
8
10
12
14
16
18
20


Spatial Colour & Texture (CEOH) Confusion
True Labels
Predicted Labels
Aeroplane
Bench
Bike
Bird
Book
Building
Car
Cat
Coast
Cow
Dog
Farm
Flower
People
Road
Selfie
Sheep
Sign
Tree
Water


Average Precision-Recall Curve
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Recall

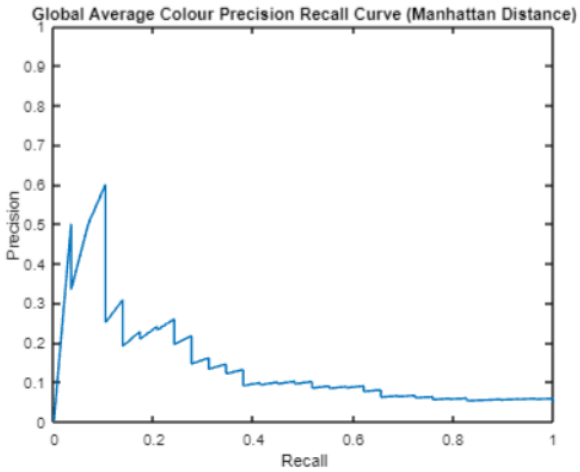
GCH with Book Category as Target Object



Average RGB with Minkowski:

Minkowski Distance

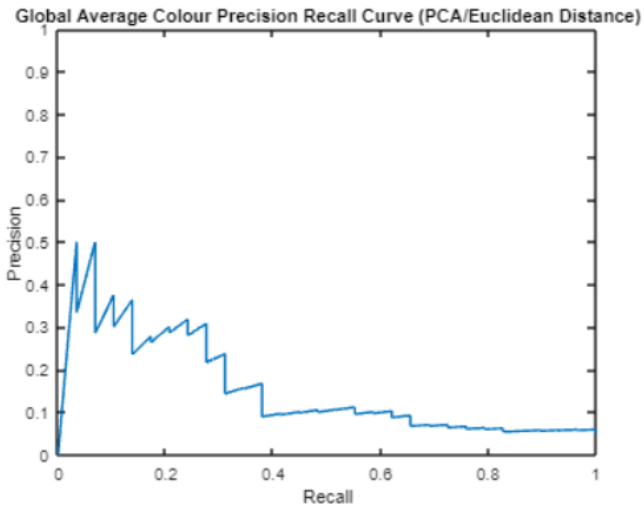

```
Precision: 0.285714
Recall: 0.137931
Average Precision 0.159375
```



Principal Component Analysis using L2 norm:

PCA with Euclidean Distance

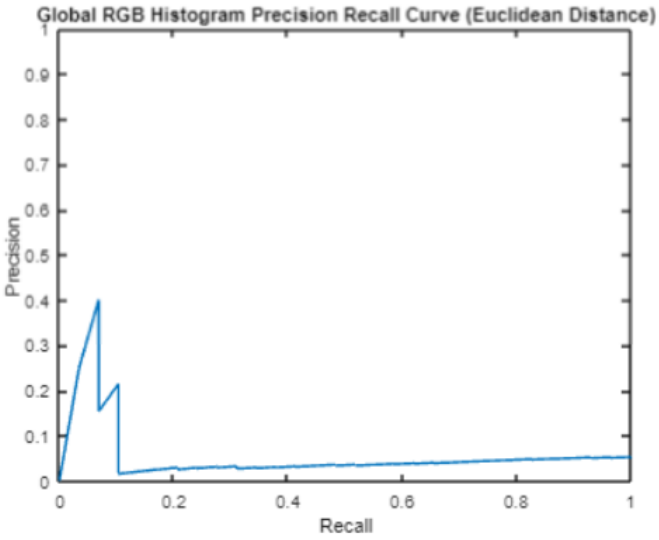

```
Precision: 0.285714
Recall: 0.137931
Average Precision 0.169900
```

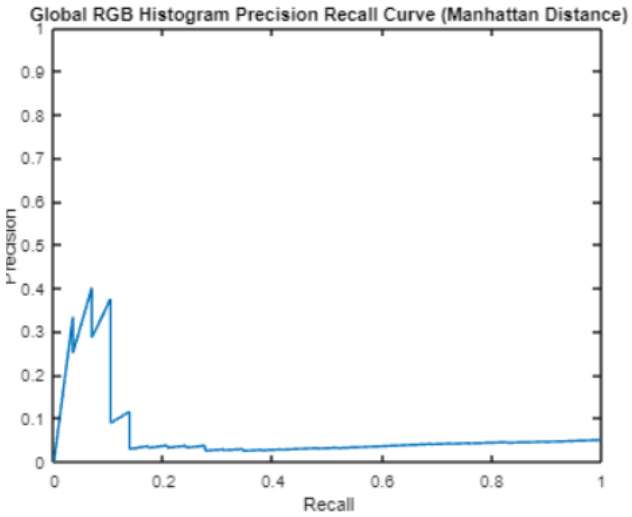


GCH with L2 norm:

Euclidean Distance

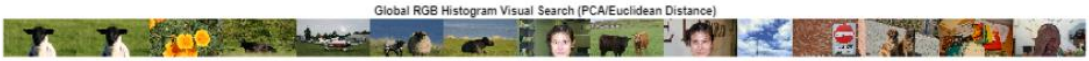
Manhattan Distance



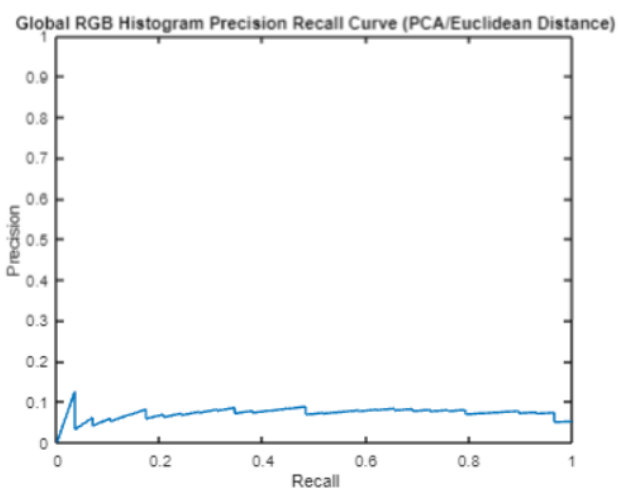


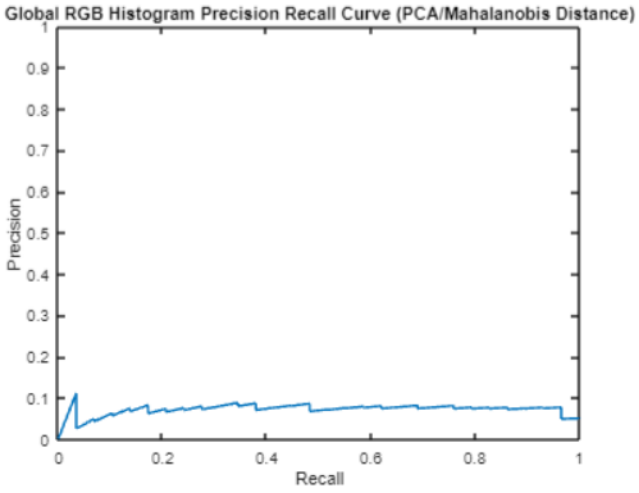
GCH with L2 vs Mahalanobis in a Reduced Dimensional Space:

PCA/Euclidean Distance

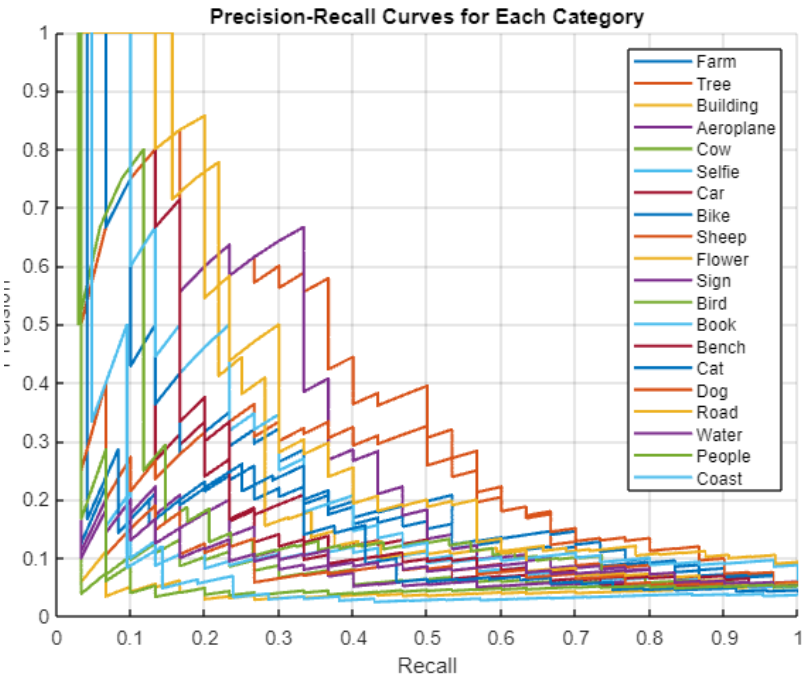


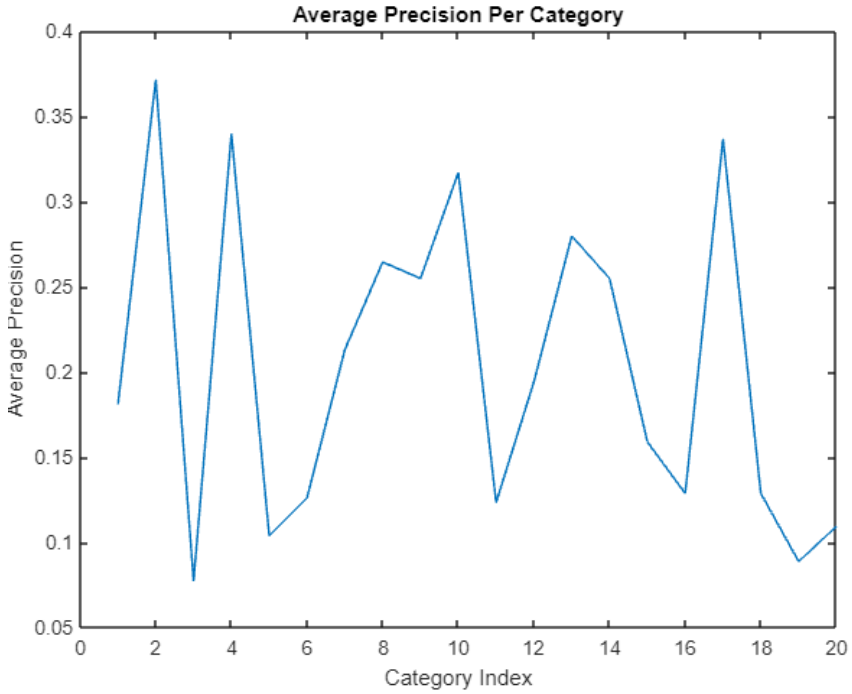
PCA/Mahalanobis Distance





Spatial Colour with L2





Colour Spatial Grid Experimenting Multiple Grid Divisions

Table 2: AP and MAP for Colour Spatial Grid Visual Search for Different Grid Cells

Category	4×4	8×8	16×16
Aeroplane	0.61782	0.66713	0.66434
Car	0.20543	0.20018	0.15461
Bike	0.35524	0.38222	0.38892
Tree	0.54884	0.58179	0.57429
MAP	**0.3610**	**0.3620**	**0.3445**

Spatial MAP of Grid Cells for Merged Spatial Colour and Texture

Grid Size	4	6	8
4	0.2149	0.205	0.1977
6	0.2035	0.1919	0.1828
8	0.2049	0.1939	0.1836