EEE3032- Computer Vision and Pattern Recognition

Coursework Assignment

Visual Search for Image Collection

Saizalpreet Kaur (6915210 | zs00774)

Abstract

This report details the implementation and evaluation of a Content-Based Image Retrieval (CBIR) system on the MSRC-v2 dataset. Several feature descriptors were systematically compared, starting with a baseline Global Colour Histogram and progressing to more complex Spatial Grid methods combining local colour and texture (Edge Orientation Histogram). The impact of different distance metrics (L1, L2, Chi-Squared) and the effectiveness of PCA for dimensionality reduction were also investigated. Performance was measured using Mean Average Precision (mAP). The baseline Global Colour Histogram achieved a peak mAP of 0.2131. The optimal system combined a spatial grid (8x8) with a local colour histogram (Q=24) and an Edge Orientation Histogram (T=16), achieving the highest mAP of 0.2442 using the L1 metric. This demonstrates the superior performance of combining local colour and texture features. This finding was further validated in a classification task, where the same descriptor achieved a 51.26% accuracy with an SVM, significantly outperforming the baseline descriptor.

Keywords: VisualSearch, Histograms, SVM, PCA

Introduction

Content-Based Image Retrieval (CBIR) enables searching for images based on their visual content, such as colour, texture, and shapes. This project implements and evaluates a visual search to retrieve visually similar images from a collection for a given query image. The system is developed using the Microsoft Research (MSRC-v2) dataset, which consists of 591 images across 20 object categories. Starting from a baseline Global Colour Histogram descriptor and progressing to more complex methods, the visual search system is on-line for improvement.

A range of techniques is explored, including spatial grid-based descriptors that combine local colour and texture (Edge Orientation Histograms). The impact of different distance metrics (Euclidean, L1, Chi-Squared) and dimensionality reduction using PCA is systematically investigated. Finally, the discriminative power of the best descriptors is tested on a classification task using a Support Vector Machine (SVM). The performance of all retrieval methods is quantitatively evaluated using standard information retrieval metrics, primarily Mean Average Precision (mAP) and Precision-Recall (PR) curves, to identify the most effective approach for this dataset.

Techniques Implemented

Throughout this report, different methods have been covered to improve the visual retrieval. The different descriptors, distance measures and metrics used are described in this section. Different python scripts were used to complete each requirement and the core structure including cvpr_compare, cvpr_evaluate, cvpr_computedescriptors was maintained.

1. Global Colour Histogram

The baseline descriptor was implemented using Global Colour Histogram method. It captures only the colour profile of the image, completely discarding the information about where those colours are. This method worked by first normalising the RGB values, then quantising the 3-D RGB space into 'Q' bins for each dimension. Then, it stacks the Q bins from all three dimensions, giving us a histogram with Q^3 bins, which captures the number of pixels falling into each bin for each colour. The pixels are assigned to bins using the formula: $R_{bin}Q^2 + G_{bin}Q + B_{bin}$. A normalised histogram is the feature vector for each image in the dataset with Q^3 dimensions. The value of Q was experimented with to find the best performing feature descriptor. The values of Q used are: 2, 4, 6, 8, 10, 12, 16, 24, 32, 40, 55, 64. The process is visualised in the figure below:

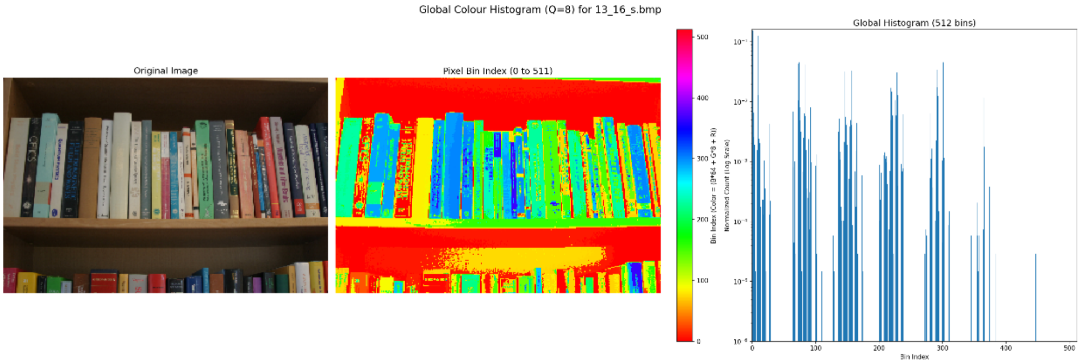


Fig 1: The division of image pixels into quantised bins for (Q=8)

The descriptors of all 591 images are saved once computed and then are called when a query image is used to compare the distance between the query features and the features of all other images. The distance metric used to compare features in this experiment was Euclidean or L2 Norm.

2. Evaluation Methodology

To compare the techniques used throughout the experimentation, all 591 images were used as query one by one against the remaining 590 images. Precision, defined as the number of true positives over the number of true positives and false positives and Recall, which is defined as the number of true positives over the number of true positives plus number of false negatives [1] were used as the fundamental metrics.

- **Precision-Recall curves** are plotted for every method used. A perfect system would have a precision of 1.0 for all recall values.
- **Confusion Matrix**: Used to evaluate the Top-1 accuracy, meaning for one query, what was the top retrieved result. Plots a matrix for every query being correctly or incorrectly classified.
- **Mean Average Precision** (mAP): For each query, Average Precision (AP) is calculated by evaluating both the precision and the ranking of the relevant images in the retrieved images, it measures the retrieval quality for every query. Mean Average Precision averages the AP scores of all queries, evaluated individually and gives a single value between 0 to 1 (1 being the best) capturing the performance of the system.

3. Spatial Grid Descriptors (Colour and Texture)

The limitation of using Global Colour Histogram as a descriptor was the loss of spatial information of pixels. To solve this problem, spatial grid method was used where the image is divided into a uniform grid of say 8x8. Each cell is then considered independent and the colour features for each cell are computed. The final descriptor is formed by concatenating the feature of each grid cell in a fixed order. This helps maintain the colour/texture image while also preserving the spatial information of pixels.

Three cell level features were implemented for colour and texture information.

1. **Average Colour**- For each grid cell, a 3-dimensional vector containing the values of R, G and B, averaged for that cell was used as the local feature, eventually concatenated with features from other grid cells as stated.
2. **Colour Histogram-** The method used for baseline was instead used here locally, per grid cell to maintain both spatial and colour information for pixels. This gives a Q^3 dimensional vector for every cell.
3. **Edge Orientation Histogram-** This texture descriptor captures the structural patterns- lines and edges within each cell by using a Sobel filter to calculate the image gradients in both x and y directions. The gradient magnitude represents the strength of the edges, and the orientation gives the direction of the edge.
   Weaker edges are ignored and for the stronger edges, based on their orientation, they are sub-categorised into angularly quantised sections Q, which make the bins for the edge orientation histogram.

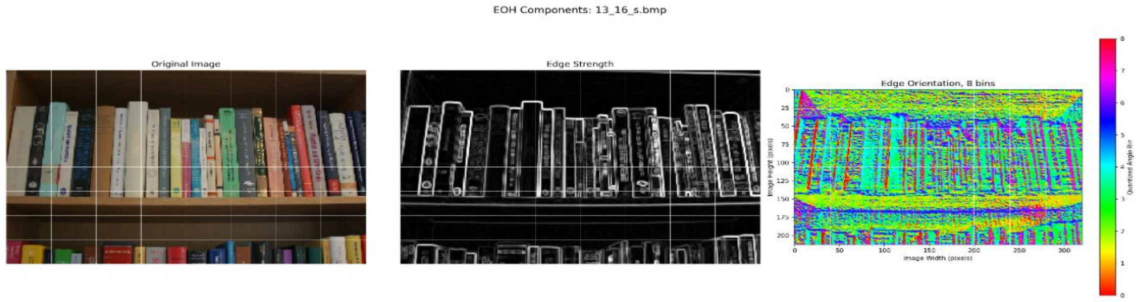


Fig 2: The Gradient Magnitude and Orientation (showing the edges in quantised angular bins using colour bar)

The script compute_descriptors_3.py uses these 3 methods and a combination of these to get both colour and edge information. The image is divided into a grid for all these instances. The experiments covered are using just colour-based features, just edge-based features, or combining the edge-based features with the two colour-based features. Different values of Quantisation and Angular Quantisation are explored.

4. PCA and Mahalanobis Distance

The descriptor generated by combining both colour and texture features, while using spatial grid reach up to 885760 dimensions for each image which is enormous. This is computationally draining and can suffer from the curse of dimensionality (the dataset is comparatively very small), which causes many of the feature dimensions to be noise.

Principal Component Analysis (PCA) is a dimension reduction technique that analyses the 591 images of the dataset and finds new principal components (uncorrelated) in the high-dimensional space, which capture the most amount of variance captured in the data. So, using PCA, we decrease the size of the descriptors massively.

Mahalanobis Distance is a weighted distance that accounts for the covariance and scale of the data. It is theoretically ideal distance measure used after applying PCA. It scales the distance calculation along each new principal component by its variance. This prevents components with naturally large variance from dominating the distance calculation, which is a common problem when using standard Euclidean distance on PCA transformed data.

5. Different Distance Measures and Descriptors

Hue Saturation Value (HSV)- Another global colour histogram descriptor was used. But instead of using RGB values of pixels, HSV was used. This model is more aligned with how human perceive images. It separates the colour information from brightness and focuses on both colour and brightness features. Hue represents the colour; Saturation represents the intensity of that colour and Value gives the brightness. Three separate -D histograms are used for H, S and V and then concatenated to form the feature descriptor. Different number of bins were used to experiment with the 1-D histograms. Using this method, we reduce the dimension of the feature vector substantially while also maintaining the good results.

Local Binary Patterns (LBP)- This is a texture descriptor**.** It works by converting the image to grayscale. It then scans each pixel's local neighbourhood, comparing the centre pixel's brightness to its surrounding neighbours (specified by radius parameter). This comparison creates a binary number where '1' means the neighbour is brighter and '0' means it's darker. These binary patterns, representing the micro-textures are stored to form histograms.

Once the features are calculated, we use mathematical distances between features of different images to measure the dissimilarity. The distance measures used throughout the experiment are described below:

1. **Euclidean Distance (L2 Norm):** Treats the feature vector as a point in high-dimensional space and calculates the straight-line distance between two features.

$$d = \sqrt{\sum_{i=1}^{n} f_i^2}$$

2. **Manhattan Distance (L1 Norm):** Measures the sum of absolute differences between each corresponding element in the feature vectors.

$$d = \sum_{i=1}^{n} |f_i|$$

3. **Chi-squared Distance:** It specifically measures the similarity or dissimilarity between two probability distributions (histograms). Unlike the Euclidean distance, which treats all differences the same, the chi-squared distance formula normalizes the squared difference between two bins by the sum of those bins. This gives more weight to differences that occur in bins with smaller values.

$$d(f,g) = \frac{1}{2}\sum_{i=1}^{n}\frac{(f_i - g_i)^2}{(f_i + g_i)}$$

4. **Cosine Distance:** This measures the cosine of the angle between two vectors.

$$d(f,g) = 1 - \frac{f.g}{\|f\|\|g\|}$$

6. Classification using Support Vector Machine (SVM)

A shift from Image Retrieval to classification was attempted using SVM. The dataset is divided into a training/testing split of 80:20. Stratification is employed to ensure a proportional class representation in both sets. The descriptors calculated in one of the above methods is used for this classification task. SVC from sklearn.svm is used to use one vs one strategy for the 20 classes. A standard scaler is used for fitting and transform the data.

Experimental Results

All images were used as query one by one, and the mAP score is calculated over the whole dataset. The query image used throughout for visualisation purposes is 13_14_s.bmp. Requirement 5 (different distance measures used is covered in each section). The results other than the best performing ones are given in the Appendix

1. Global Colour Histogram

Euclidean Distance measure was used for this experiment while varying the quantisation values.\ The table with mAP values for each Q and distance measure is given in the appendix. The plot showing the trend is given below:

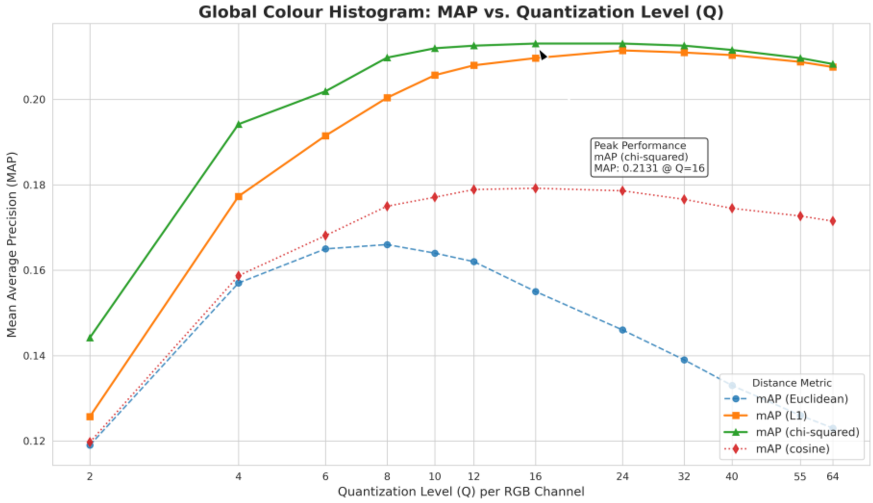


Fig 3: mAP vs Q for Global Colour Histogram

At low Q values (e.g., Q=2, 4), performance is poor due to under-quantization. The descriptor is too coarse, mapping visually distinct colours into the same bin and thus lacking discriminative power. At high Q values (e.g., Q > 32), performance degrades due to over-quantization. This introduces two problems: 1) The descriptor becomes highly sensitive to noise, as minor lighting variations can shift a pixel to a different, sparse bin. 2) The "Curse of Dimensionality" becomes significant. In the resulting high-dimensional, sparse space, all points tend to become equidistant. This second problem explains why the Euclidean (L2) distance performs worst. It is highly susceptible to the Curse of Dimensionality and peaks early at Q=8 (MAP ≈ 0.166). In contrast, the Chi-Squared and L1 metrics are inherently more robust for comparing sparse histograms, as they appropriately weight differences in low-count bins. These metrics are less affected by high dimensionality, allowing them to peak later (Q=16-24) and achieve the best overall MAP of ≈0.213. This peak represents the optimal balance between a discriminative and a robust descriptor.

The retrieved images are shown below for the query from class 13:

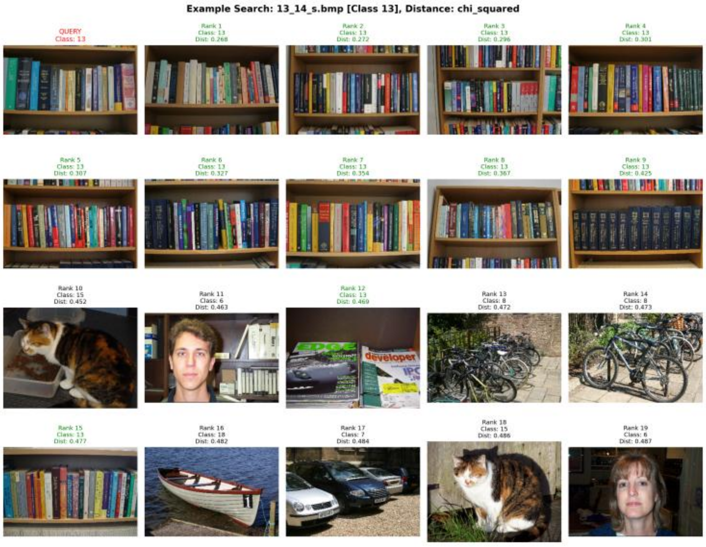

Example Search: 13_14_s.bmp [Class 13], Distance: chi_squared
QUERY Class: 13
Rank 1 Class: 13 Dist: 0.268
Rank 2 Class: 13 Dist: 0.272
Rank 3 Class: 13 Dist: 0.296
Rank 4 Class: 13 Dist: 0.301
Rank 5 Class: 13 Dist: 0.307
Rank 6 Class: 13 Dist: 0.327
Rank 7 Class: 13 Dist: 0.354
Rank 8 Class: 13 Dist: 0.367
Rank 9 Class: 13 Dist: 0.425
Rank 10 Class: 15 Dist: 0.452
Rank 11 Class: 6 Dist: 0.463
Rank 12 Class: 13 Dist: 0.469
Rank 13 Class: 8 Dist: 0.472
Rank 14 Class: 8 Dist: 0.473
Rank 15 Class: 13 Dist: 0.477
Rank 16 Class: 18 Dist: 0.482
Rank 17 Class: 7 Dist: 0.484
Rank 18 Class: 15 Dist: 0.486
Rank 19 Class: 6 Dist: 0.487

Fig 4: Top-19 retrieved images visualised for the selected query image, for Q=16 (chi-squared)

2. Evaluation Methodology

The PR curve for all the 591 queries over the remaining 590 retrieved images is generated. Also, an example PR curve for 13_14_s.bmp is computed over the remaining 590 images. The confusion matrix is shown for the same query.

Also, for Q=16, the average precision for each class is shown, telling that class 13 performs the best and class 11 performs the worst.

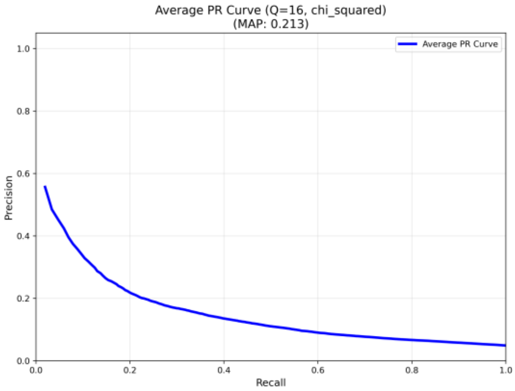


Fig 5a: Average PR Curve for all queries

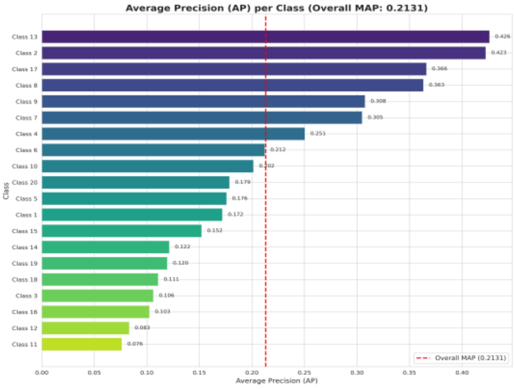


Fig 5b: Average Precision for each class

The confusion matrix and precision recall for the specific query are given in figure below:

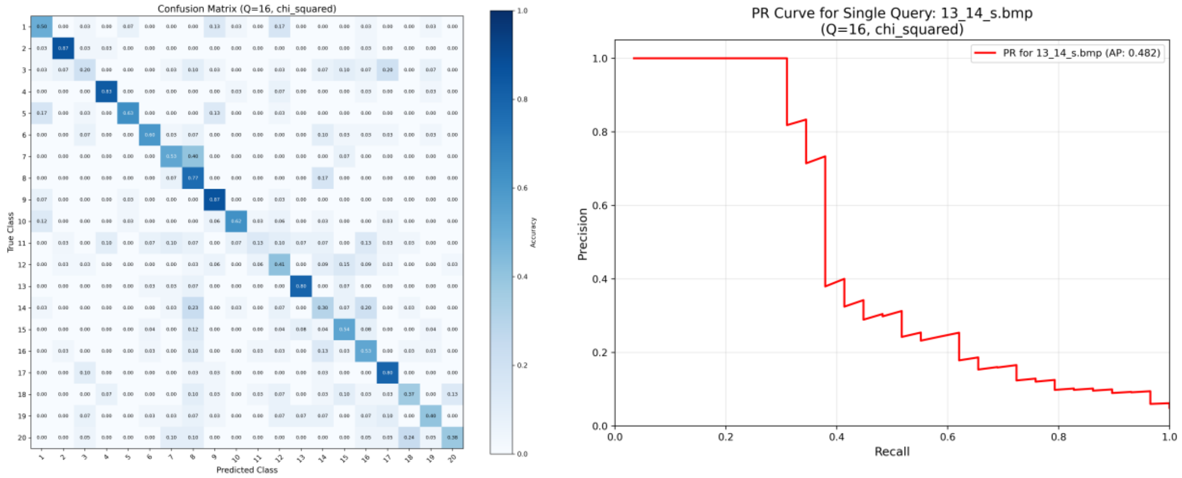


Fig 6: The Confusion Matrix and PR curve for 13_4_s.bmp for Q=16 (chi-squared)

3. Spatial Grid based Descriptors

The images were divided into a 8x8 grid for all these experiments. T= angular quantisation, Q= quantisation, And the distance measure being L2 norm. The values of Q are chosen from exp 1 which give the best mAP for each distance measure.

Group	Feature(s)	Parameters	MAP (Euclidean)	MAP (L1)	MAP (Chi-Squared)	MAP (Cosine)
EOH (Texture Only)	EOH	T=4	0.0803	0.0848	0.0816	0.0813
	EOH	T=8	0.1176	0.1358	0.1257	0.1262
	EOH	T=16	0.1265	0.1518	0.1400	0.1436
	EOH	T=32	0.1275	0.1527	0.1416	0.1509
Average Colour	Avg. Colour	-	0.1526	0.1581	0.1560	0.1614
	Avg. Colour + EOH	T=4	0.1463	0.1506	0.1438	0.1561
	Avg. Colour + EOH	T=8	0.1622	0.1776	0.1687	0.1766
	Avg. Colour + EOH	T=16	0.1646	0.1869	0.1774	0.1794
	Avg. Colour + EOH	T=32	0.1625	0.1868	0.1778	0.1765
Colour Histogram	Colour Hist	(Baseline)	0.1468 (Q=8)	0.2256 (Q=24)	0.2271 (Q=16)	0.1919 (Q=16)
	Colour Hist + EOH	T=4	0.1417 (Q=8)	0.2081 (Q=24)	0.2189 (Q=16)	0.1240 (Q=16)
	Colour Hist + EOH	T=8	0.1498 (Q=8)	0.2344 (Q=24)	0.2312 (Q=16)	0.1511 (Q=16)
	Colour Hist + EOH	T=16	0.1513 (Q=8)	0.2442 (Q=24)	0.2372 (Q=16)	0.1727 (Q=16)
	Colour Hist + EOH	T=32	0.1503 (Q=8)	0.2442 (Q=24)	0.2381 (Q=16)	0.1876 (Q=16)

Table 1: mAP scores for all experiments done with spatial grid, using both texture and colour features

The best performance comes from combining the colour histogram and EOH for Q=24 and T=16 and using the best compatible metric (L1). proving that feature synergy (combining local colour and texture) is critical and outperforms individual features like edge features or colour features individually. It is also apparent that the correct metric and feature combination affects the output.

The retrieved results for the Colour Histogram+ EOH with T=16 are:

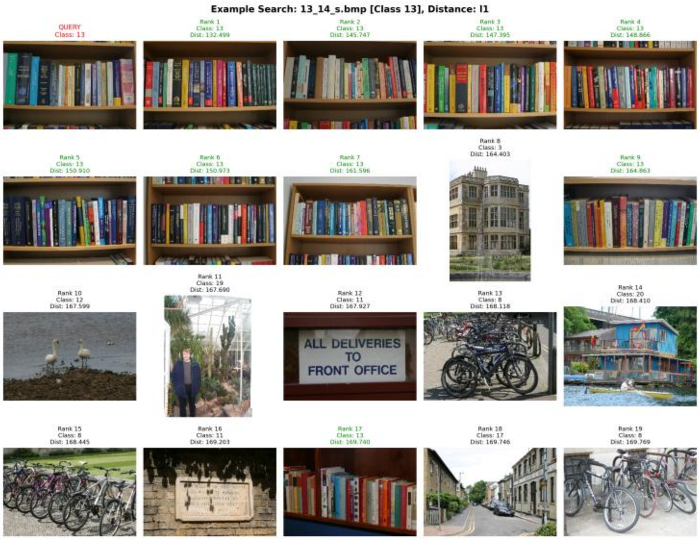
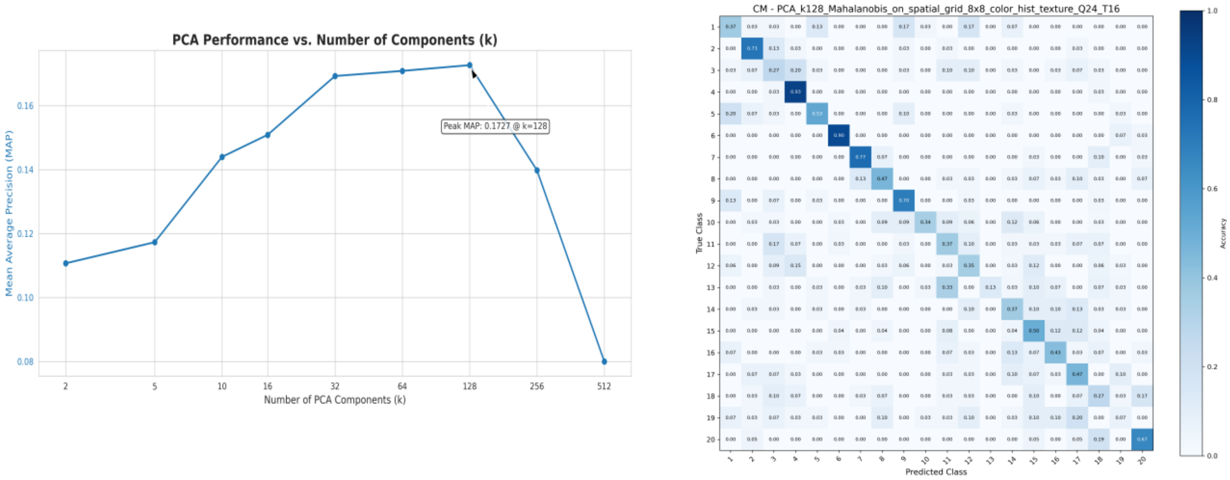


Fig 7: Retrieved images for the best performing feature descriptor

4. PCA and Mahalanobis Distance

The best result from experiment 3 is used for PCA dimension reduction. The feature dimension for colour histogram+ EOH is 885760. Using PCA, this is decreased to [2, 4, 8, 16, 32, 64, 128, 256, 512] one by one. Mahalanobis Distance is used as the distance measure for this experiment.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.37	0.03	0.03	0.00	0.13	0.00	0.00	0.00	0.17	0.03	0.00	0.17	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.73	0.13	0.03	0.00	0.00	0.00	0.00	0.03	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00
3	0.03	0.07	0.27	0.20	0.03	0.00	0.00	0.00	0.03	0.00	0.10	0.10	0.00	0.03	0.00	0.03	0.07	0.03	0.00	0.00
4	0.00	0.00	0.03	0.93	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00
5	0.20	0.07	0.03	0.00	0.53	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.03	0.00
6	0.00	0.00	0.00	0.00	0.00	0.90	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.03
7	0.00	0.00	0.00	0.00	0.00	0.00	0.77	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.10	0.00	0.03
8	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.47	0.00	0.00	0.03	0.03	0.00	0.03	0.07	0.03	0.10	0.03	0.00	0.07
9	0.13	0.00	0.07	0.00	0.03	0.00	0.00	0.00	0.70	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00
10	0.00	0.03	0.03	0.00	0.00	0.03	0.00	0.09	0.09	0.34	0.09	0.06	0.00	0.12	0.06	0.00	0.00	0.03	0.00	0.00
11	0.00	0.00	0.17	0.07	0.00	0.03	0.00	0.00	0.03	0.00	0.37	0.10	0.00	0.03	0.03	0.03	0.07	0.07	0.00	0.00
12	0.06	0.00	0.09	0.15	0.00	0.00	0.00	0.03	0.06	0.00	0.03	0.35	0.00	0.03	0.12	0.00	0.00	0.06	0.03	0.00
13	0.00	0.00	0.03	0.00	0.03	0.00	0.03	0.10	0.00	0.03	0.33	0.00	0.13	0.03	0.10	0.07	0.00	0.07	0.03	0.00
14	0.00	0.03	0.03	0.00	0.03	0.00	0.03	0.00	0.00	0.00	0.00	0.10	0.00	0.37	0.10	0.10	0.13	0.03	0.03	0.00
15	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.04	0.00	0.00	0.08	0.00	0.00	0.04	0.50	0.12	0.12	0.04	0.00	0.00
16	0.07	0.00	0.00	0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.07	0.03	0.00	0.13	0.07	0.43	0.03	0.07	0.00	0.03
17	0.00	0.07	0.07	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.00	0.10	0.07	0.03	0.47	0.00	0.10	0.00
18	0.00	0.03	0.10	0.07	0.00	0.00	0.03	0.07	0.00	0.00	0.03	0.03	0.00	0.00	0.10	0.00	0.07	0.27	0.03	0.17
19	0.07	0.03	0.07	0.03	0.03	0.00	0.00	0.10	0.00	0.03	0.03	0.10	0.00	0.03	0.10	0.10	0.20	0.00	0.07	0.00
20	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.05	0.19	0.00	0.67

Fig 8: mAP for the new reduced dimensions is shown in (a), while the confusion matrix for k-128 (best mAP) in (b)

The plot reveals the critical impact of dimensionality (k) on the performance of the PCA-transformed descriptor using Mahalanobis distance. Performance steadily increases from a low MAP of 0.1107 at k=2, as more discriminative variance is retained. The system reaches its optimal "performance at k=128, achieving a peak MAP of 0.1727. Beyond this peak, performance collapses dramatically, falling to a MAP of just 0.0800 by k=512. This severe drop indicates that the higher-order components (k > 128) are capturing noise rather than meaningful signals, and including them pollutes the descriptor, causing the Mahalanobis distance calculation to fail. Most importantly, the best PCA/Mahalanobis score (0.1727) is significantly lower than the MAP of 0.2442 achieved by the original, full-dimensional descriptor with an L1 metric, demonstrating that the PCA compression was lossy and discarded critical information.

5. Different Descriptors and Distance Measures

Two different descriptors were used- HSV histogram and LBP. These methods reduce the dimensionality. For HSV technique, H,S and V were used to form different 1-D histograms that were concatenated .

Method	Parameters	Distance	mAP
HSV	H=64,S=32,V=32	Chi_squared	0.2162
HSV	H=32, S=16, V=16	Chi_squared	0.2153
HSV	H=16, S=8, V=8	Chi_squared	0.2110
HSV	H=16, S=8, V=8	Euclidean	0.1918
HSV	H=32, S=16, V=16	Euclidean	0.1895
LBP	R=2	Chi_squared	0.1748
LBP	R=1	Chi_squared	0.1725
LBP	R=3	Chi_squared	0.1717
LBP	R=1	Euclidean	0.1653
LBP	R=2	Euclidean	0.1583

Table 2: The performance of the two new histogram methods using various distance measures

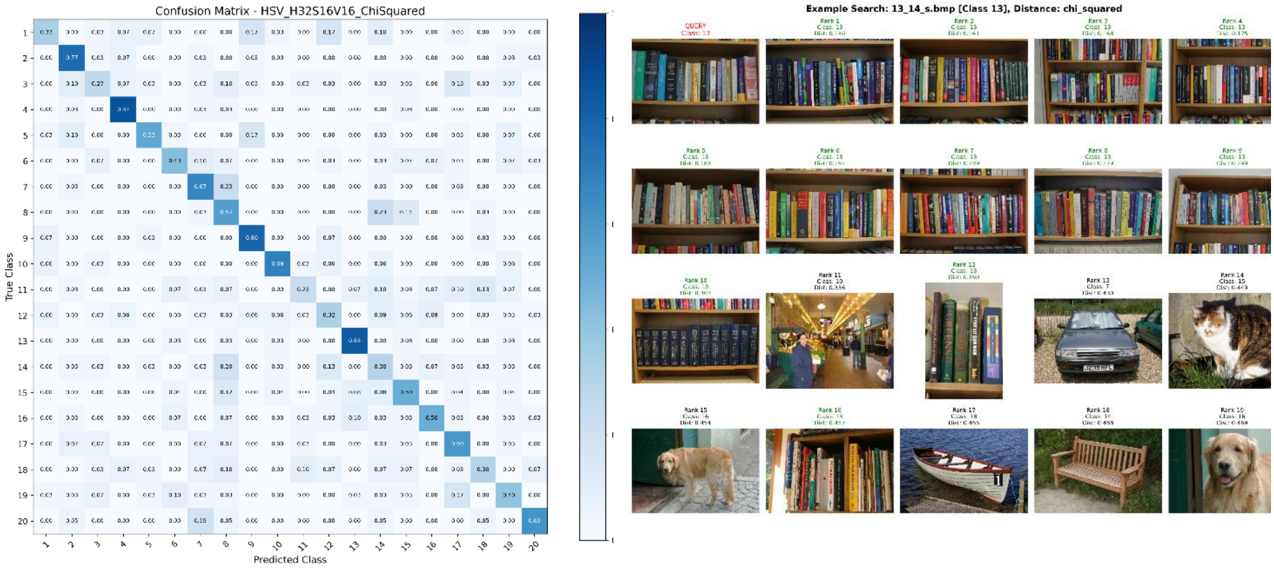


Fig 9: The results for H=64, S=32, V=32 compared using chi_squared distance measure

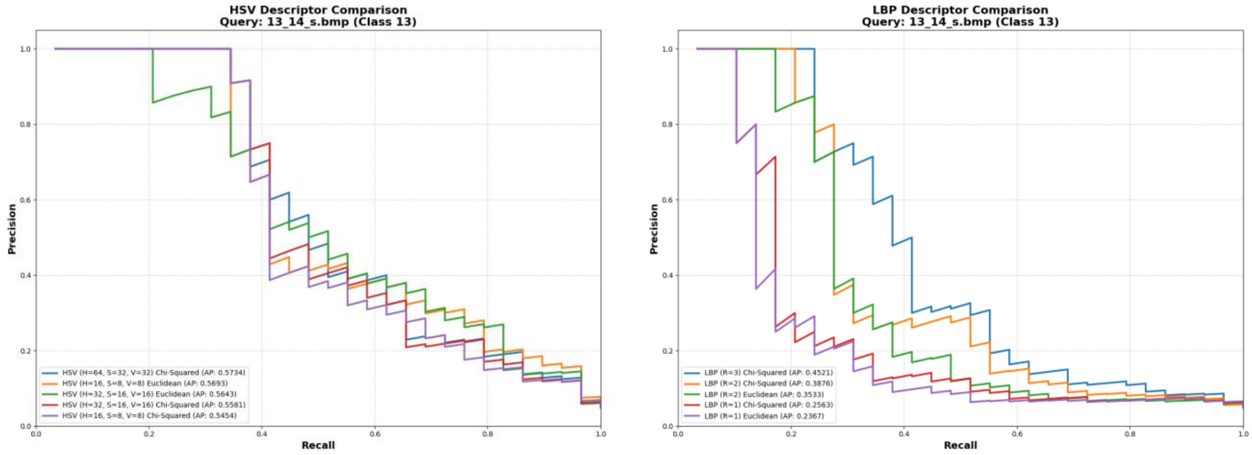


Fig 10: The P-R curves for both new descriptors for various configurations

6. SVM

SVM was used on two descriptors: globalcolourhistogram_24 and spatial_grid_colour_histo_Q24_T16. All histograms were normalized so that their values sum to one, ensuring scale invariance across images.

Before training, all feature vectors were standardized using z-score normalization, transforming each feature to have zero mean and unit variance. This step ensures that all dimensions contribute equally to the decision boundary. The dataset was split into 80% training and 20% testing using stratified sampling to preserve class balance.

Linear kernel was tried to test linear separability and Radial Basis Function (RBF) kernel was used to capture non-linear class boundaries.

The results are for global colour histogram are:

Experiment	Accuracy	Best Params
1_SVM_Linear_Fine_C	0.3950	{'C': 0.01, 'kernel': 'linear'}
2_SVM_Linear_Fine_C_Balanced	0.3950	{'C': 0.01, 'kernel': 'linear'}
3_SVM_RBF_Broad_Search	0.3697	{'C': 10, 'gamma': 'scale', 'kernel': 'rbf'}
5_SVM_RBF_Fine_Search	0.3697	{'C': 10, 'gamma': 0.0001, 'kernel': 'rbf'}
4_SVM_RBF_Broad_Balanced	0.3613	{'C': 10, 'gamma': 'scale', 'kernel': 'rbf'}
6_SVM_RBF_Fine_Balanced	0.3613	{'C': 10, 'gamma': 0.0001, 'kernel': 'rbf'}
7_SVM_Poly_Deg2_Search	0.2101	{'C': 50, 'degree': 2, 'kernel': 'poly'}
8_SVM_Poly_Deg2_Balanced	0.2101	{'C': 50, 'degree': 2, 'kernel': 'poly'}
9_SVM_Poly_Deg3_Search	0.1008	{'C': 50, 'degree': 3, 'kernel': 'poly'}
10_SVM_Poly_Deg3_Balanced	0.1008	{'C': 50, 'degree': 3, 'kernel': 'poly'}

Table 3: Performance of SVM performed on global colour histogram from exp 1

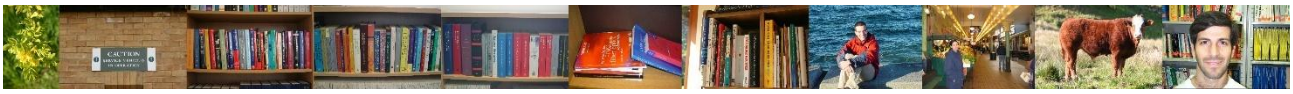
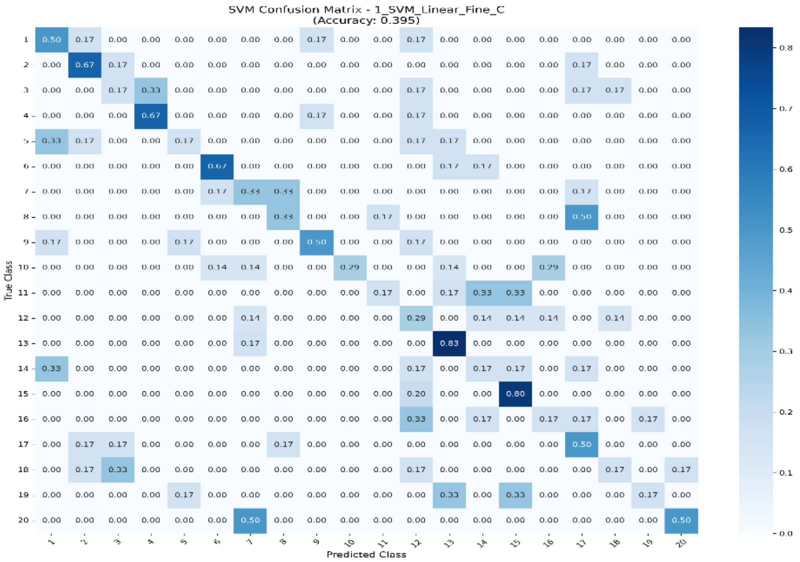


Fig 11: The images classified into class 13, from which the query has been used throughout



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.50	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.67	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.00
3	0.00	0.00	0.17	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.17	0.17	0.00	0.00
4	0.00	0.00	0.00	0.67	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.33	0.17	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.67	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.17	0.00	0.00	0.00	0.00	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.17	0.33	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.00
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.00
9	0.17	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.50	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00	0.00	0.14	0.14	0.00	0.00	0.29	0.00	0.00	0.14	0.00	0.00	0.29	0.00	0.00	0.00	0.00
11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.17	0.33	0.33	0.00	0.00	0.00	0.00	0.00
12	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.00	0.00	0.00	0.00	0.29	0.00	0.14	0.14	0.14	0.00	0.14	0.00	0.00
13	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.83	0.00	0.00	0.00	0.00	0.00	0.00	0.00
14	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.17	0.17	0.00	0.17	0.00	0.00	0.00
15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.80	0.00	0.00	0.00	0.00	0.00
16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.17	0.00	0.17	0.17	0.00	0.17	0.00
17	0.00	0.17	0.17	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.00
18	0.00	0.17	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.17
19	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.33	0.00	0.00	0.00	0.17	0.00
20	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50

Fig 12: The confusion matrix for SVM using a linear kernel and c=0.01 for global colour histogram descriptor

The results for spatial grid colour histogram (Q24)+EOH(T16):

Experiment	Accuracy	Best Params
1_LinearKernel	0.5126	{'C': 0.01, 'kernel': 'linear'}
2_RBF_HighDim	0.2941	{'C': 100, 'gamma': 1e-06, 'kernel': 'rbf'}

Table 4: SVM on spatial grid using both colour and texture from exp 3

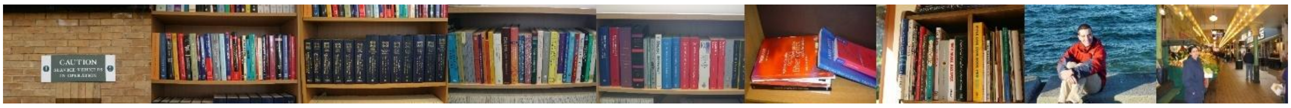


Fig 13: Images classified into class 13

SVM performs good but the factors to be considered are that the dataset is quite small for this classification task.

Discussion

The experiments conducted provide a clear understanding of the factors influencing visual search performance. The baseline Global Colour Histogram establishes an initial benchmark, achieving a peak mAP of 0.2131 with a Chi-Squared distance metric and Q=16. This method, though simple, was highly sensitive to the choice of distance metric, with L1 and Chi-Squared significantly outperforming Euclidean (L2) distance. This could likely be due to their robustness to the high-dimensional, sparse feature vectors. A major limitation of the global histogram is its disregard for spatial information. This was addressed by the Spatial Grid descriptors. Using colour features (Average Colour, Colour Histogram) in a grid provided a modest improvement. However, the most significant performance gain came from feature combination: combining the local Colour Histogram (Q=24) with an Edge Orientation Histogram (T=16). This descriptor, using an L1 metric, achieved the highest mAP of the entire project: **0.2442**. This result strongly indicates that combining local colour and texture information is critical for discriminating between categories in this dataset.

Further experiments with descriptors showed that an HSV histogram (mAP 0.2162) performed comparably to the baseline RGB histogram, while reducing the feature space dimension drastically. The LBP texture descriptor on its own (mAP 0.1748) was less effective, suggesting that colour is a more dominant feature than texture alone for this dataset, which can also be observed in Exp 3 where EOH was used on its own.

PCA and Mahalanobis distance was used on Colour Histogram + EOH. Performance peaked at mAP 0.1727 (k=128) before collapsing at higher dimensions. This was significantly worse than the original, full-dimensional descriptor's mAP of 0.2442. This suggests that while PCA reduces dimensionality, it discards critical, discriminative information for this specific task, and the Mahalanobis distance could not effectively compensate.

Finally, the SVM classification reinforced these findings. The baseline Global Colour Histogram descriptor yielded a modest 39.5% accuracy. In contrast, using the best retrieval descriptor (Spatial Grid Colour+EOH) dramatically improved classification accuracy to 51.26%. This confirms that the features that are effective for retrieval are also highly discriminative for classification.

It was noted that for the dataset used, many classes have similar/ intersecting visible features. So, there is a semantic overlap. [2] For example, the first class features mainly grass, but includes images of different animals or flowers on grass as well. But, in the dataset, the separate classes of cows, sheep and flowers exist. This causes

In conclusion, the most effective visual search system was one that used a spatial grid to capture local colour and texture information (Colour Histogram + EOH), demonstrating that simple, global methods are insufficient for this complex dataset.

Bibliography

[1] "Scikit-learn," [Online]. Available: https://scikit-learn.org/stable/auto_examples/model_selection/plot_precision_recall.html.

[2] [Online]. Available: https://mldta.com/dataset/msrc-v2/.

[3] R. Szeliski, in *Computer Vision: Algorithms and Applications*, 2010.

[4] *EEE3032 Week 3 Lecture Slides, University of Surrey.*

Appendix

Experiment 1: Global Colour Histogram

Q	mAP (Euclidean)	mAP (L1)	mAP (chi-squared)	mAP (cosine)
2	0.119	0.1257	0.1442	0.1197
4	0.157	0.1773	0.1942	0.1587
6	0.165	0.1915	0.2019	0.1681
8	0.166	0.2004	0.2098	0.175
10	0.164	0.2057	0.212	0.1771
12	0.162	0.208	0.2126	0.1789
16	0.155	0.2097	0.2131	0.1792
24	0.146	0.2115	0.2131	0.1786
32	0.139	0.211	0.2126	0.1766
40	0.133	0.2104	0.2116	0.1745
55	0.126	0.2088	0.2097	0.1727
64	0.123	0.2076	0.2083	0.1715

Table 5: Performance of different quantisation values for different distance measures

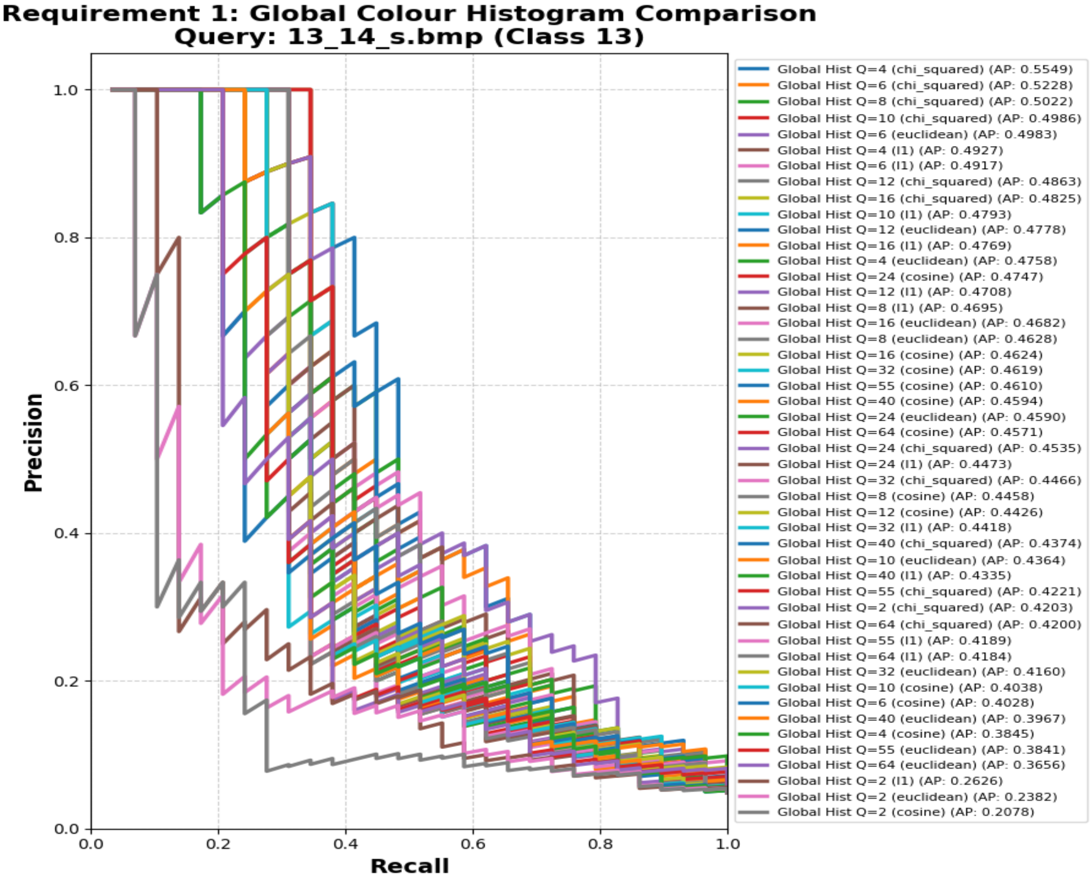
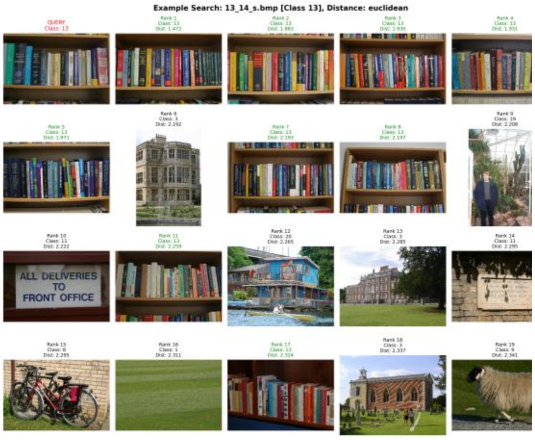
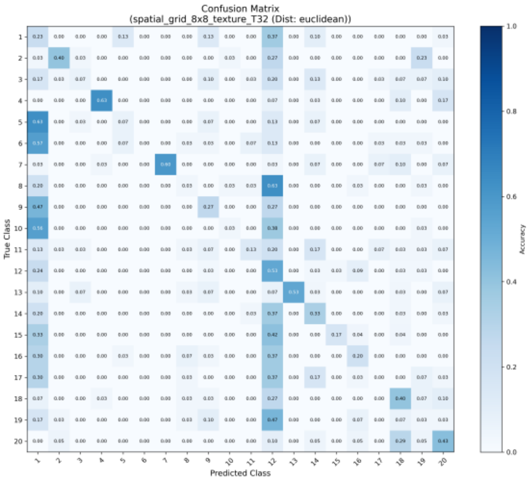


Fig 14: P-R curve for the query 13_14_s.bmp for all the attempted methods

Experiment 3: Spatial Grid

1. EOH only

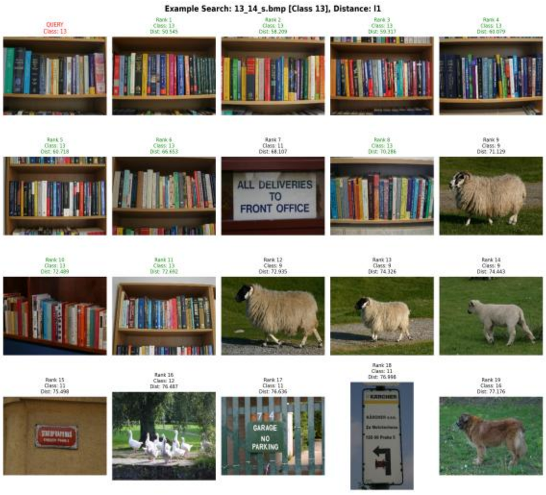


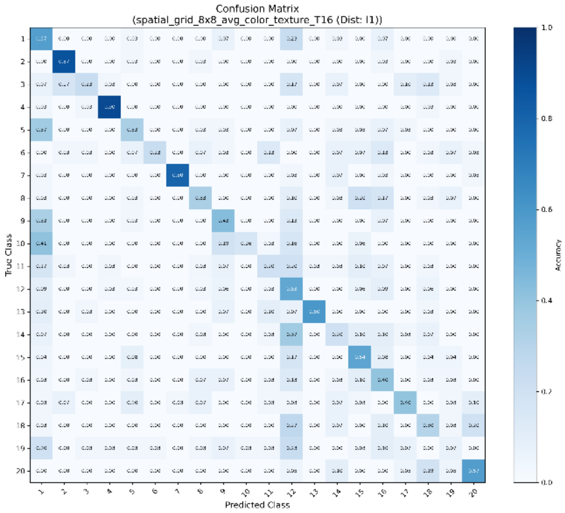

Confusion Matrix
(spatial_grid_8x8_texture_T32 (Dist: euclidean))

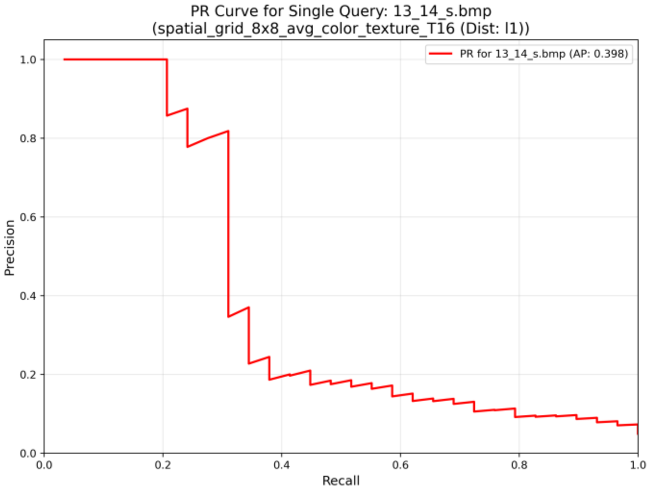
True Class \ Predicted Class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.23	0.00	0.00	0.00	0.13	0.00	0.00	0.00	0.13	0.00	0.00	0.37	0.00	0.10	0.00	0.00	0.00	0.00	0.00	0.03
2	0.03	0.40	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.27	0.00	0.00	0.00	0.00	0.00	0.00	0.23	0.00
3	0.17	0.03	0.07	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.03	0.20	0.00	0.13	0.00	0.00	0.03	0.07	0.07	0.10
4	0.00	0.00	0.00	0.63	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.03	0.00	0.00	0.00	0.10	0.00	0.17
5	0.63	0.00	0.03	0.00	0.07	0.00	0.00	0.00	0.07	0.00	0.00	0.13	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00
6	0.57	0.00	0.00	0.00	0.07	0.00	0.00	0.03	0.03	0.00	0.07	0.13	0.00	0.00	0.00	0.00	0.03	0.03	0.03	0.00
7	0.03	0.00	0.00	0.03	0.00	0.00	0.60	0.00	0.00	0.00	0.00	0.03	0.00	0.07	0.00	0.00	0.07	0.10	0.00	0.07
8	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.03	0.03	0.63	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00
9	0.47	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.27	0.00	0.00	0.27	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	0.56	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.38	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00
11	0.13	0.03	0.03	0.00	0.00	0.00	0.00	0.03	0.07	0.00	0.13	0.20	0.00	0.17	0.00	0.00	0.07	0.03	0.03	0.07
12	0.24	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.53	0.00	0.03	0.03	0.09	0.00	0.03	0.03	0.00
13	0.10	0.00	0.07	0.00	0.00	0.00	0.00	0.03	0.07	0.00	0.00	0.07	0.53	0.03	0.00	0.00	0.00	0.03	0.00	0.07
14	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.37	0.00	0.33	0.00	0.00	0.00	0.03	0.00	0.03
15	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.42	0.00	0.00	0.17	0.04	0.00	0.04	0.00	0.00
16	0.30	0.00	0.00	0.00	0.03	0.00	0.00	0.07	0.03	0.00	0.00	0.37	0.00	0.00	0.00	0.20	0.00	0.00	0.00	0.00
17	0.30	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.37	0.00	0.17	0.00	0.03	0.00	0.00	0.07	0.03
18	0.07	0.00	0.00	0.03	0.00	0.00	0.00	0.03	0.03	0.00	0.00	0.27	0.00	0.00	0.00	0.00	0.00	0.40	0.07	0.10
19	0.17	0.03	0.00	0.00	0.00	0.00	0.00	0.03	0.10	0.00	0.00	0.47	0.00	0.00	0.00	0.07	0.00	0.07	0.03	0.03
20	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.05	0.00	0.05	0.00	0.29	0.05	0.43

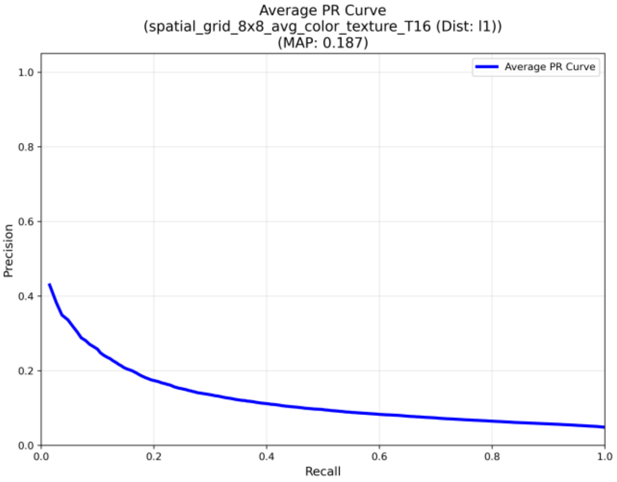
Fig : Results for only EOH using Euclidean distance for best performance

2. Average Colour
These figures show the results acheived for Average colour + EOH using 16 T and L1 as the distance measure

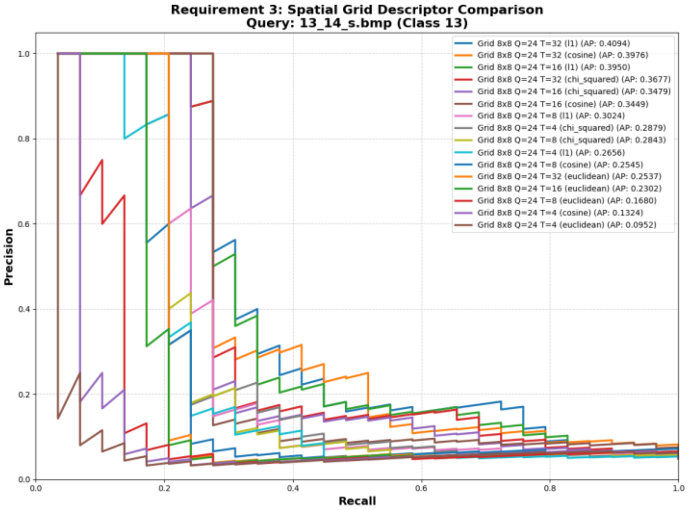

Example Search: 13_14_s.bmp [Class 13], Distance: l1
QUERY Class: 13
Rank 1 Class: 13 Dist: 50.545
Rank 2 Class: 13 Dist: 58.209
Rank 3 Class: 13 Dist: 59.317
Rank 4 Class: 13 Dist: 60.079
Rank 5 Class: 13 Dist: 60.718
Rank 6 Class: 13 Dist: 66.653
Rank 7 Class: 11 Dist: 68.107
ALL DELIVERIES TO FRONT OFFICE
Rank 8 Class: 13 Dist: 70.286
Rank 9 Class: 9 Dist: 71.129
Rank 10 Class: 13 Dist: 72.489
Rank 11 Class: 13 Dist: 72.692
Rank 12 Class: 9 Dist: 72.935
Rank 13 Class: 9 Dist: 74.326
Rank 14 Class: 9 Dist: 74.443
Rank 15 Class: 11 Dist: 75.498
Rank 16 Class: 12 Dist: 76.487
Rank 17 Class: 11 Dist: 76.636
GARAGE NO PARKING
Rank 18 Class: 11 Dist: 76.998
Rank 19 Class: 16 Dist: 77.176







3. Colour Histogrm + EOH



Precision recall for exp 3 for the query image 13_14_s.bmp

For experiment 3, the retreived

Chi_Squared:

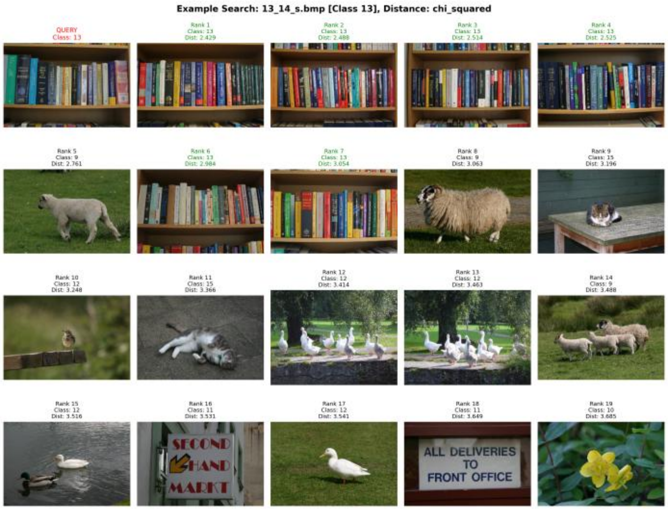

Example Search: 13_14_s.bmp [Class 13], Distance: chi_squared
QUERY Class: 13
Rank 1 Class: 13 Dist: 2.429
Rank 2 Class: 13 Dist: 2.488
Rank 3 Class: 13 Dist: 2.514
Rank 4 Class: 13 Dist: 2.525
Rank 5 Class: 9 Dist: 2.761
Rank 6 Class: 13 Dist: 2.984
Rank 7 Class: 13 Dist: 3.054
Rank 8 Class: 9 Dist: 3.063
Rank 9 Class: 15 Dist: 3.196
Rank 10 Class: 12 Dist: 3.248
Rank 11 Class: 15 Dist: 3.366
Rank 12 Class: 12 Dist: 3.414
Rank 13 Class: 12 Dist: 3.463
Rank 14 Class: 9 Dist: 3.488
Rank 15 Class: 12 Dist: 3.516
Rank 16 Class: 11 Dist: 3.531
SECOND HAND MARKT
Rank 17 Class: 12 Dist: 3.541
Rank 18 Class: 11 Dist: 3.649
ALL DELIVERIES TO FRONT OFFICE
Rank 19 Class: 10 Dist: 3.685

Fig: For average colour

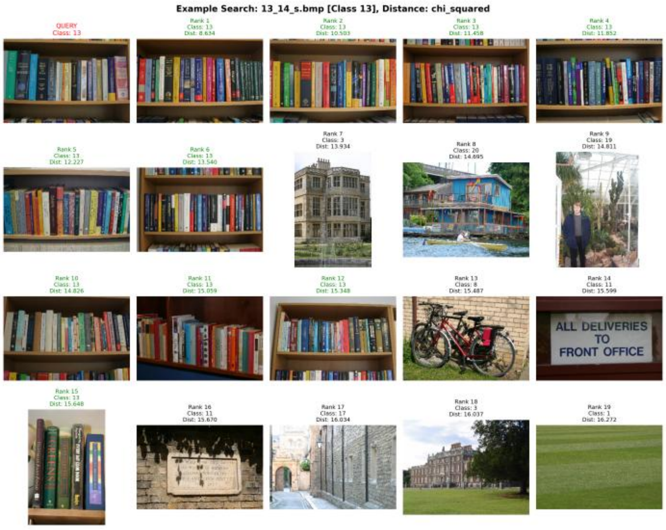

Example Search: 13_14_s.bmp [Class 13], Distance: chi_squared
QUERY Class: 13
Rank 1 Class: 13 Dist: 8.634
Rank 2 Class: 13 Dist: 10.503
Rank 3 Class: 13 Dist: 11.458
Rank 4 Class: 13 Dist: 11.852
Rank 5 Class: 13 Dist: 12.227
Rank 6 Class: 13 Dist: 13.540
Rank 7 Class: 3 Dist: 13.934
Rank 8 Class: 20 Dist: 14.695
Rank 9 Class: 19 Dist: 14.811
Rank 10 Class: 13 Dist: 14.826
Rank 11 Class: 13 Dist: 15.059
Rank 12 Class: 13 Dist: 15.348
Rank 13 Class: 8 Dist: 15.487
Rank 14 Class: 11 Dist: 15.599
ALL DELIVERIES TO FRONT OFFICE
Rank 15 Class: 13 Dist: 15.648
Rank 16 Class: 11 Dist: 15.670
Rank 17 Class: 17 Dist: 16.034
Rank 18 Class: 3 Dist: 16.037
Rank 19 Class: 1 Dist: 16.272

Fig: For texture only T=16

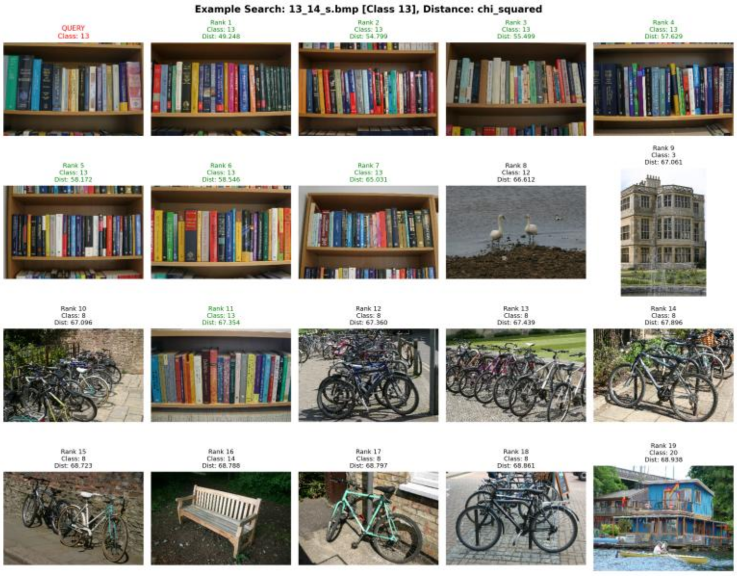

Example Search: 13_14_s.bmp [Class 13], Distance: chi_squared
QUERY Class: 13
Rank 1 Class: 13 Dist: 49.248
Rank 2 Class: 13 Dist: 54.799
Rank 3 Class: 13 Dist: 55.499
Rank 4 Class: 13 Dist: 57.629
Rank 5 Class: 13 Dist: 58.172
Rank 6 Class: 13 Dist: 58.546
Rank 7 Class: 13 Dist: 65.031
Rank 8 Class: 12 Dist: 66.612
Rank 9 Class: 3 Dist: 67.061
Rank 10 Class: 8 Dist: 67.096
Rank 11 Class: 13 Dist: 67.354
Rank 12 Class: 8 Dist: 67.360
Rank 13 Class: 8 Dist: 67.439
Rank 14 Class: 8 Dist: 67.896
Rank 15 Class: 8 Dist: 68.723
Rank 16 Class: 14 Dist: 68.788
Rank 17 Class: 8 Dist: 68.797
Rank 18 Class: 8 Dist: 68.861
Rank 19 Class: 20 Dist: 68.938

Fig: For colour hist+EOH Q=16, T=16

EUCLIDEAN

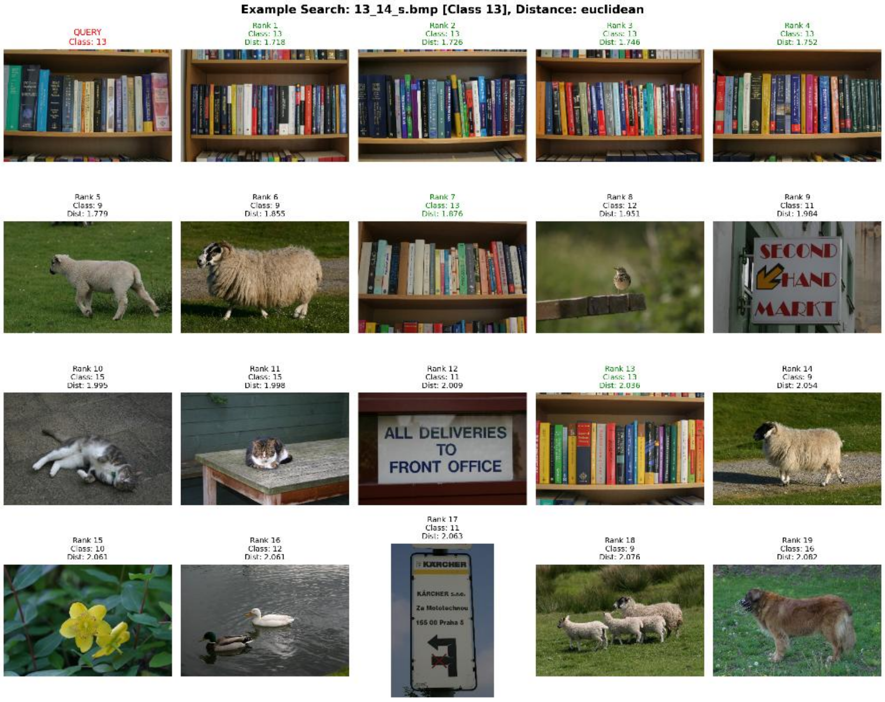


Fig: Average Colour

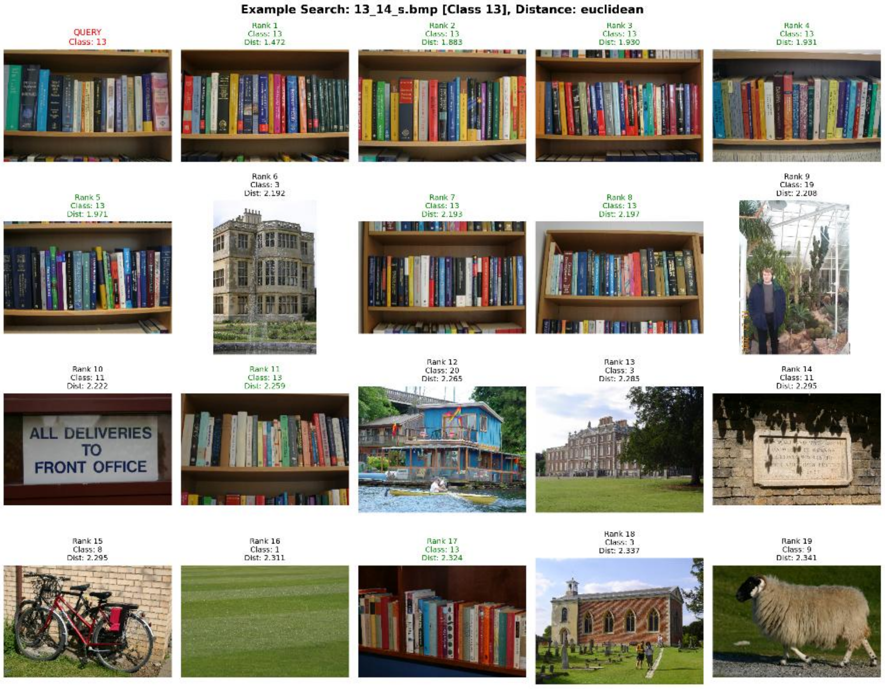


Fig: Texture alone for T=16

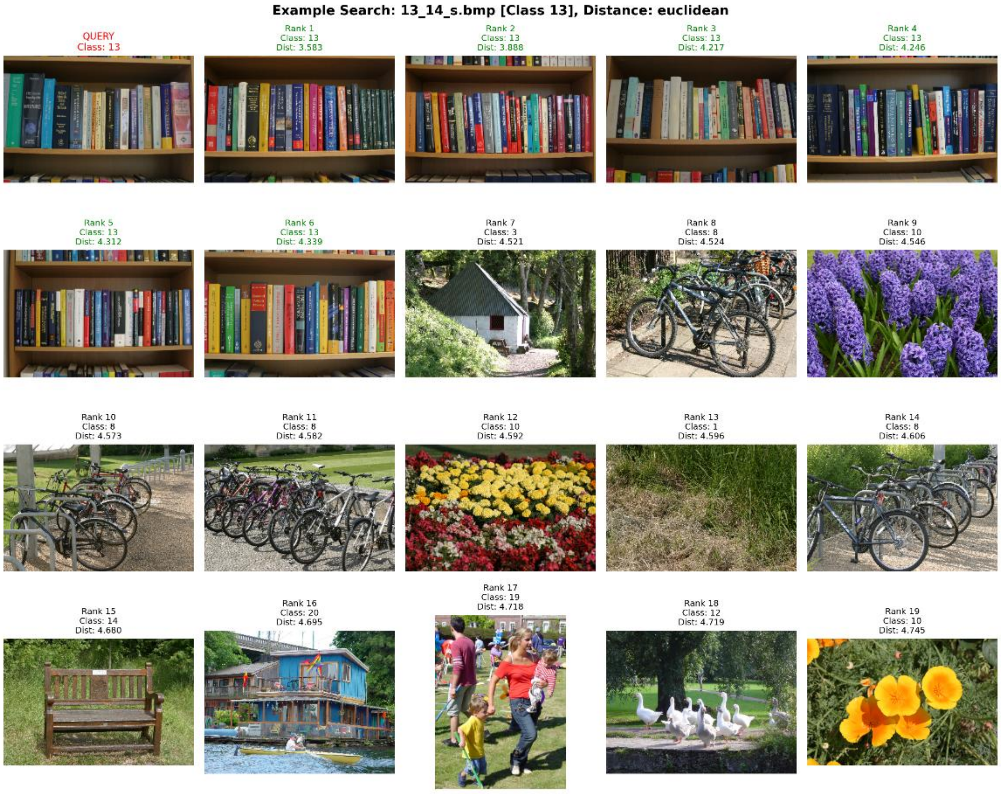

Example Search: 13_14_s.bmp [Class 13], Distance: euclidean
QUERY Class: 13
Rank 1 Class: 13 Dist: 3.583
Rank 2 Class: 13 Dist: 3.888
Rank 3 Class: 13 Dist: 4.217
Rank 4 Class: 13 Dist: 4.246
Rank 5 Class: 13 Dist: 4.312
Rank 6 Class: 13 Dist: 4.339
Rank 7 Class: 3 Dist: 4.521
Rank 8 Class: 8 Dist: 4.524
Rank 9 Class: 10 Dist: 4.546
Rank 10 Class: 8 Dist: 4.573
Rank 11 Class: 8 Dist: 4.582
Rank 12 Class: 10 Dist: 4.592
Rank 13 Class: 1 Dist: 4.596
Rank 14 Class: 8 Dist: 4.606
Rank 15 Class: 14 Dist: 4.680
Rank 16 Class: 20 Dist: 4.695
Rank 17 Class: 19 Dist: 4.718
Rank 18 Class: 12 Dist: 4.719
Rank 19 Class: 10 Dist: 4.745

Fig: Colour Hist+EOH Q=8, T=32

COSINE

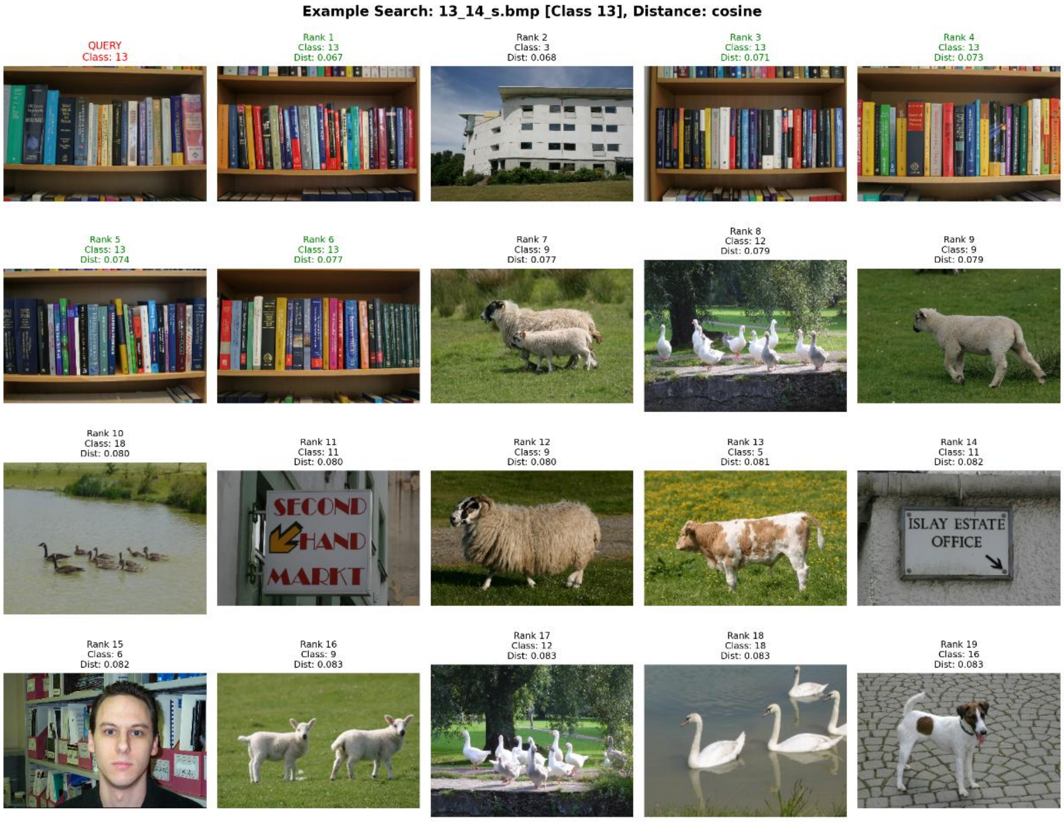

Example Search: 13_14_s.bmp [Class 13], Distance: cosine
QUERY Class: 13
Rank 1 Class: 13 Dist: 0.067
Rank 2 Class: 3 Dist: 0.068
Rank 3 Class: 13 Dist: 0.071
Rank 4 Class: 13 Dist: 0.073
Rank 5 Class: 13 Dist: 0.074
Rank 6 Class: 13 Dist: 0.077
Rank 7 Class: 9 Dist: 0.077
Rank 8 Class: 12 Dist: 0.079
Rank 9 Class: 9 Dist: 0.079
Rank 10 Class: 18 Dist: 0.080
Rank 11 Class: 11 Dist: 0.080
SECOND HAND MARKT
Rank 12 Class: 9 Dist: 0.080
Rank 13 Class: 5 Dist: 0.081
Rank 14 Class: 11 Dist: 0.082
ISLAY ESTATE OFFICE
Rank 15 Class: 6 Dist: 0.082
Rank 16 Class: 9 Dist: 0.083
Rank 17 Class: 12 Dist: 0.083
Rank 18 Class: 18 Dist: 0.083
Rank 19 Class: 16 Dist: 0.083

Fig: Average colour

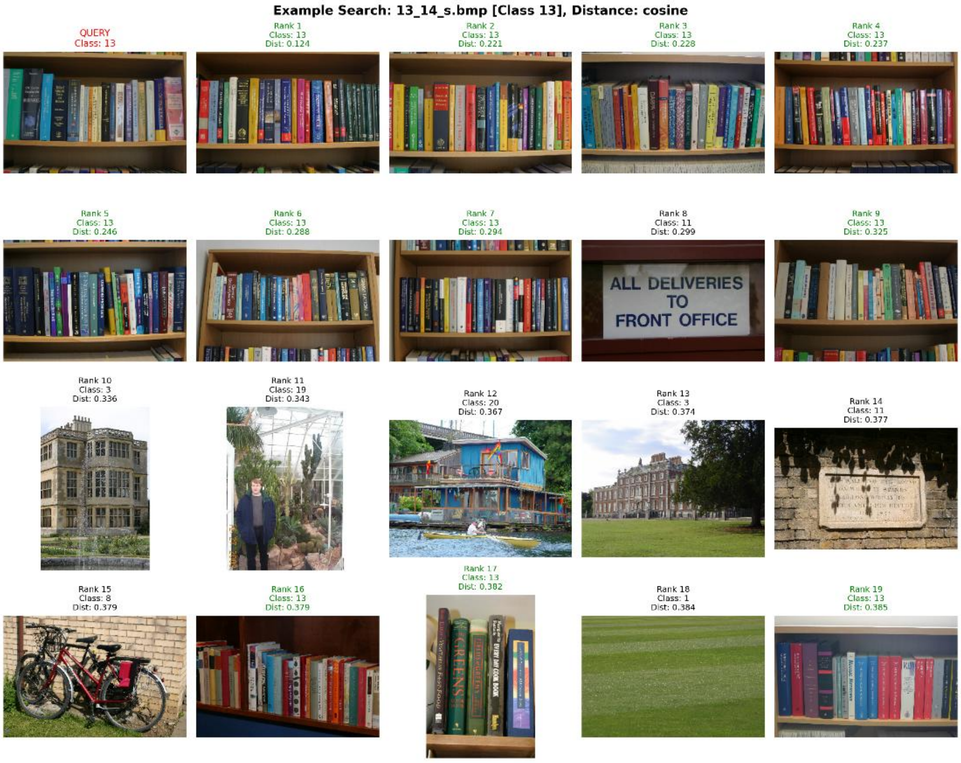

Example Search: 13_14_s.bmp [Class 13], Distance: cosine
QUERY Class: 13
Rank 1 Class: 13 Dist: 0.124
Rank 2 Class: 13 Dist: 0.221
Rank 3 Class: 13 Dist: 0.228
Rank 4 Class: 13 Dist: 0.237
Rank 5 Class: 13 Dist: 0.246
Rank 6 Class: 13 Dist: 0.288
Rank 7 Class: 13 Dist: 0.294
Rank 8 Class: 11 Dist: 0.299
ALL DELIVERIES TO FRONT OFFICE
Rank 9 Class: 13 Dist: 0.325
Rank 10 Class: 3 Dist: 0.336
Rank 11 Class: 19 Dist: 0.343
Rank 12 Class: 20 Dist: 0.367
Rank 13 Class: 3 Dist: 0.374
Rank 14 Class: 11 Dist: 0.377
Rank 15 Class: 8 Dist: 0.379
Rank 16 Class: 13 Dist: 0.379
Rank 17 Class: 13 Dist: 0.382
Rank 18 Class: 1 Dist: 0.384
Rank 19 Class: 13 Dist: 0.385

Fig: Texture only for T=32

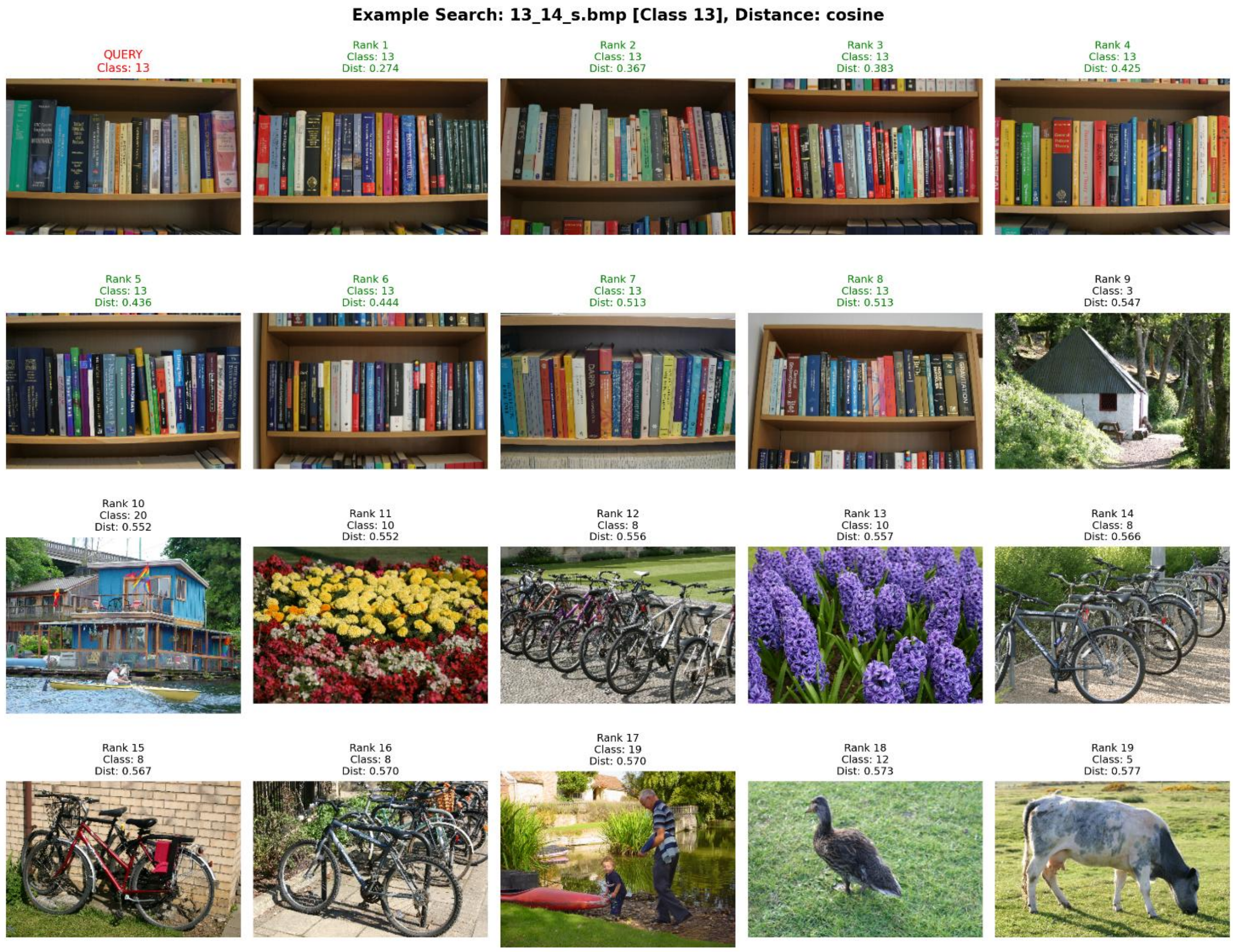


Fig: for colour hist+EOH for Q=16, T=16