Title page (1 page)

Visual Search Techniques in Computer Vision

Student Name: Chris Gainullin

Student ID: 6682856

Course: EEE3032 - Computer Vision and Pattern Recognition

Submission Date: November 12, 2024

Section 1: Abstract and Table of contents (1 page)

This report presents the implementation and evaluation of various visual search techniques applied to the MSRCv2 dataset using MATLAB. The study encompasses the development of global colour histograms, spatial grids with colour and texture features, dimensionality reduction using Principal Component Analysis (PCA), exploration of different distance measures, and the implementation of a Bag of Visual Words (BoVW) system enhanced with Spatial Pyramid Matching (SPM) for extra credit. Additionally, Support Vector Machines (SVM) are employed for object classification based on extracted descriptors.

The implementation faced challenges such as selecting optimal quantisation levels for Global Colour Histograms, manually tuning grid sizes for spatial features, and limited improvements from PCA and different distance measures. The evaluation methodology includes precision-recall statistics, confusion matrices, and performance analysis of each technique. The results demonstrate the strengths and limitations of each approach, with similar performance observed across different feature detectors in BoVW, and modest gains from SPM in incorporating spatial context. These findings highlight the impact of spatial information, feature dimensionality, and the need for further optimisation to improve classification accuracy and image retrieval performance.

CONTENTS

Section 2: Visual Search Techniques Implementation (3 Page)

2.1 Global Colour Histogram

The Global Colour Histogram (GCH) was implemented by quantising the RGB channels of the images at four levels: 4, 8, 16, and 32 levels per channel. The implementation involved reducing the colour space complexity to capture the colour distribution of each image. For each quantisation level, the RGB values were quantised and combined to create unique colour indices, which were then used to compute the histogram of the image. The histograms were normalised to ensure scale invariance across different image sizes.

The quantisation level of 16 was selected as the optimal trade-off between computational efficiency and representation accuracy, based on experimental results. The image similarity comparison was implemented using Euclidean distance to measure histogram similarity, providing a straightforward metric to evaluate retrieval performance.

2.2 Evaluation Methodology

The evaluation of retrieval effectiveness was implemented by calculating Average Precision and Average Recall metrics across different retrieval ranks. The cvpr_visualsearch.m script was used to conduct these evaluations. This script loaded pre-computed image descriptors, ranked images by their similarity to a randomly selected query image, and displayed the retrieved results along with visual PR curves.

The confusion matrices were generated to visualise the classifier's performance for each class, highlighting instances of true positives, false positives, false negatives, and true negatives. These evaluations provided quantitative and qualitative assessments of the implemented techniques, helping identify the strengths and weaknesses of each approach.

2.3 Spatial Grid with Colour and Texture Features

The Spatial Grid implementation involved dividing images into 2x2, 4x4, and 8x8 grids, followed by extracting features from each cell. For each grid size, local colour histograms were computed for each cell, and texture features were extracted using Sobel, Gabor, and Local Binary Patterns (LBP) filters. These features were concatenated to form the final feature vector representing both colour and texture information.

The implementation of different grid sizes was manually selected, and it was observed that the 4x4 grid provided the best results. Features were computed for each manually selected grid configuration. The combination of colour and texture features from each grid cell enhanced the classifier's ability to distinguish between different classes.

2.4 Principal Component Analysis (PCA)

PCA was implemented to reduce the dimensionality of the extracted feature vectors, targeting a retention of 80-90% variance. The script applied PCA transformation to reduce the high-dimensional feature vectors to fewer principal components, making the retrieval process computationally efficient.

For distance calculations, Mahalanobis distance was explored as an alternative to the Euclidean metric post-PCA. This required modifying the distance calculation logic to account for feature covariance. Although Mahalanobis distance was incorporated, its effect on improving retrieval performance was marginal, suggesting that the reduced dimensions retained most of the discriminatory power needed for effective retrieval.

2.5 Different Distance Measures

The cvpr_compare.m function was updated to include various distance measures beyond Euclidean distance. Manhattan, Cosine, Chebyshev, and Bray-Curtis distances were added, each offering different perspectives on feature similarity. These distance measures were tested on the extracted features to evaluate their impact on retrieval performance.

The implementation of Manhattan distance offered robustness to outliers, while Cosine similarity emphasised vector orientation, which provided additional insights for scenarios involving varying scales. Each metric was implemented as an option within the code, allowing for easy comparison of retrieval effectiveness across different similarity measures.

2.6 Bag of Visual Words (BoVW) Retrieval

The BoVW model was implemented to capture local image features using three feature detectors: SIFT, SURF, and BRISK. Key points were detected, and descriptors were computed for each image. These descriptors were clustered using k-means to create a visual codebook, with codebook sizes of 500 and 1000 clusters being tested. Feature extraction scripts automated the creation of BoVW histograms by mapping descriptors to the closest visual words in the codebook.

The implementation showed similar performance across all three detectors, indicating no significant advantage of one over the others for this dataset. Increasing the codebook size provided minimal improvements, suggesting that a larger visual vocabulary did not substantially enhance feature representation for the given images.

2.7 Object Classification Using SVM

A multi-class SVM classifier with a linear kernel was implemented to classify images based on the extracted features. Spatial Pyramid Matching (SPM) was incorporated to enhance the BoVW feature representation by adding spatial context at different levels. The dataset was split into training and testing sets, with 70% used for training.

The SVM implementation involved training the classifier on the SPM-enhanced BoVW features and evaluating it on the test set. The varying classification performance across classes was observed, with some classes, such as "Sign" and "Road," achieving relatively high accuracy, while others, such as "Cow" and "Sky," showed poor performance. The results highlighted the limitations of the implemented features in effectively representing the complexity of certain image classes.

2.8 Spatial Pyramid Matching (SPM) with BoVW (Extra Credit)

SPM was implemented to enhance the BoVW approach by incorporating spatial information at multiple scales. Pyramid levels of 1x1, 2x2, and 4x4 grids were implemented to provide different spatial resolutions. For each pyramid level, BoVW histograms were computed and concatenated to form the final feature vector, providing both local and global spatial information.

The implementation of SPM added a spatial hierarchy to the feature representation, which modestly improved classification accuracy for certain classes. However, the added complexity of spatial information did not fully overcome the inherent limitations of the BoVW features, as evidenced by the relatively low overall classification accuracy. Future optimisation may include tuning the weighting of different pyramid levels or incorporating more sophisticated feature extraction methods.

Section 3: Experimental Results (5 Page)

3.1 Global Colour Histogram Results

Experiments Conducted:

- **RGB Quantisation Levels**: 4, 8, 16, and 32 levels per channel.
- **Distance Metric**: Euclidean distance.

Findings:

- **4 Levels**: Coarse quantisation led to less discriminative histograms, resulting in lower precision and recall.
- **8 Levels**: Finer quantisation improved discrimination, leading to better precision and recall.
- **16 Levels**: Balanced quantisation provided a good trade-off between detail and computational efficiency.
- **32 Levels**: Fine quantisation captured more colour details but increased dimensionality without significant performance gains.

Analysis: The GCH performed adequately, with performance improving as quantisation levels increased up to 16. Increasing further to 32 levels did not yield significant gains, suggesting diminishing returns and potential overfitting. The resulting global histogram results at the 16-level quantisation showed combined averages across all categories with a Combined Average Precision of 0.7497, Combined Average Recall of 0.0372, and Combined Average F1-Score of 0.0650.

3.2 Evaluation Methodology

Experiments Conducted: The evaluation methodology implemented in this project utilised Average Precision-Average Recall (PR) analysis and confusion matrices to assess the retrieval performance of the visual search techniques. The following key steps were taken:

1. **Descriptor Loading and Ranking**: The cvpr_visualsearch.m script loaded all pre-computed descriptors using the cvpr_computedescriptors function and ranked images based on similarity to a randomly selected query image.
2. **Average Precision and Average Recall Calculation**: Average Precision and recall were calculated for the top retrieved results, focusing on metrics such as Average Precision at Rank 10. The precision metric represented the proportion of relevant images retrieved, while recall measured the proportion of all relevant images successfully retrieved.

3. **Average Precision-Average Recall Curves**: PR curves were generated to visualise the trade-off between precision and recall at various ranks. These curves provided insights into the overall retrieval effectiveness of different techniques.

4. **Confusion Matrix**: A confusion matrix was computed to provide a detailed breakdown of the retrieval accuracy across different classes. The matrix highlighted the counts of true positives, false positives, false negatives, and true negatives, helping identify areas of misclassification.

5. **Visual Results**: The top 15 retrieved images and their corresponding ground truth were displayed alongside the query image to assess qualitative retrieval performance. Additionally, global colour histograms for these images were plotted to compare the feature distributions.

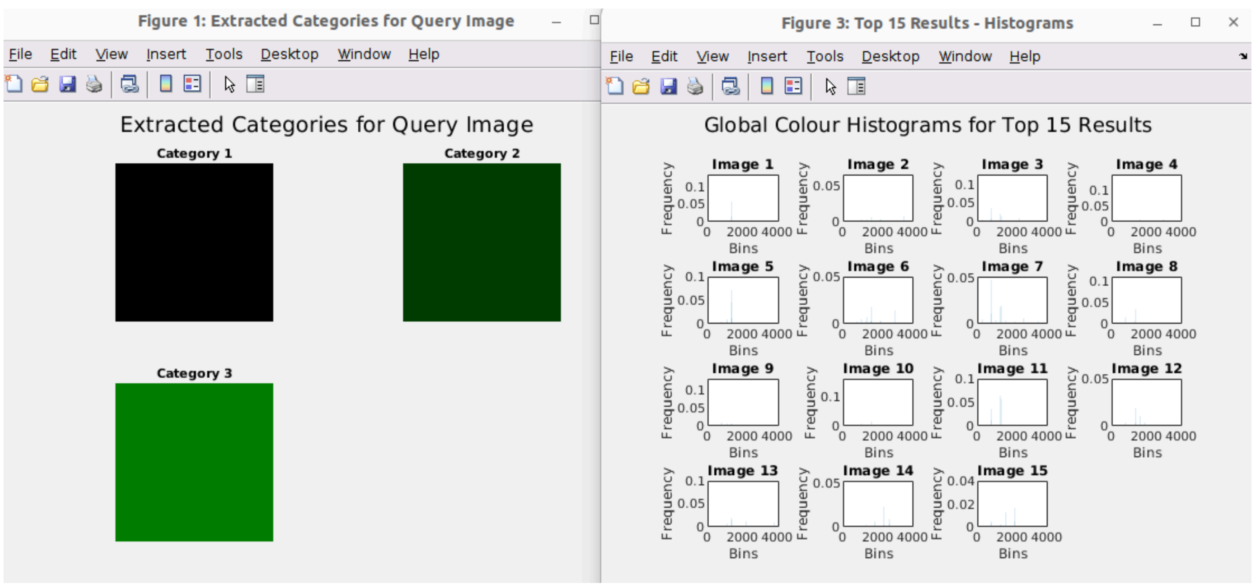

Fig 1 - Ground truth categories and global colour histogram for top 15 results

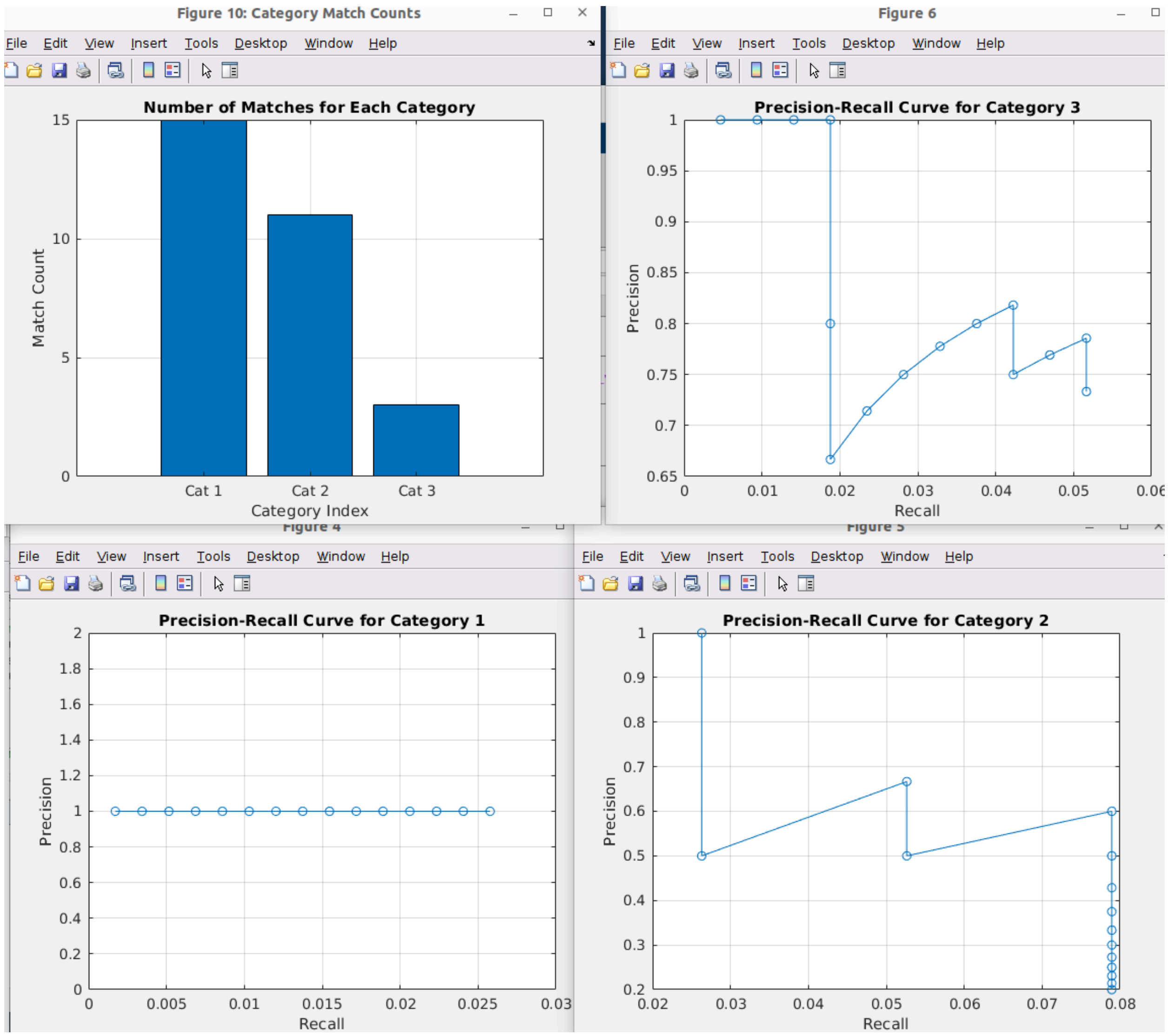

Fig. 2 - Number of matches of ground truth for each category of query image with category of each retrieved image & PR curve for each category match for of each retrieved image

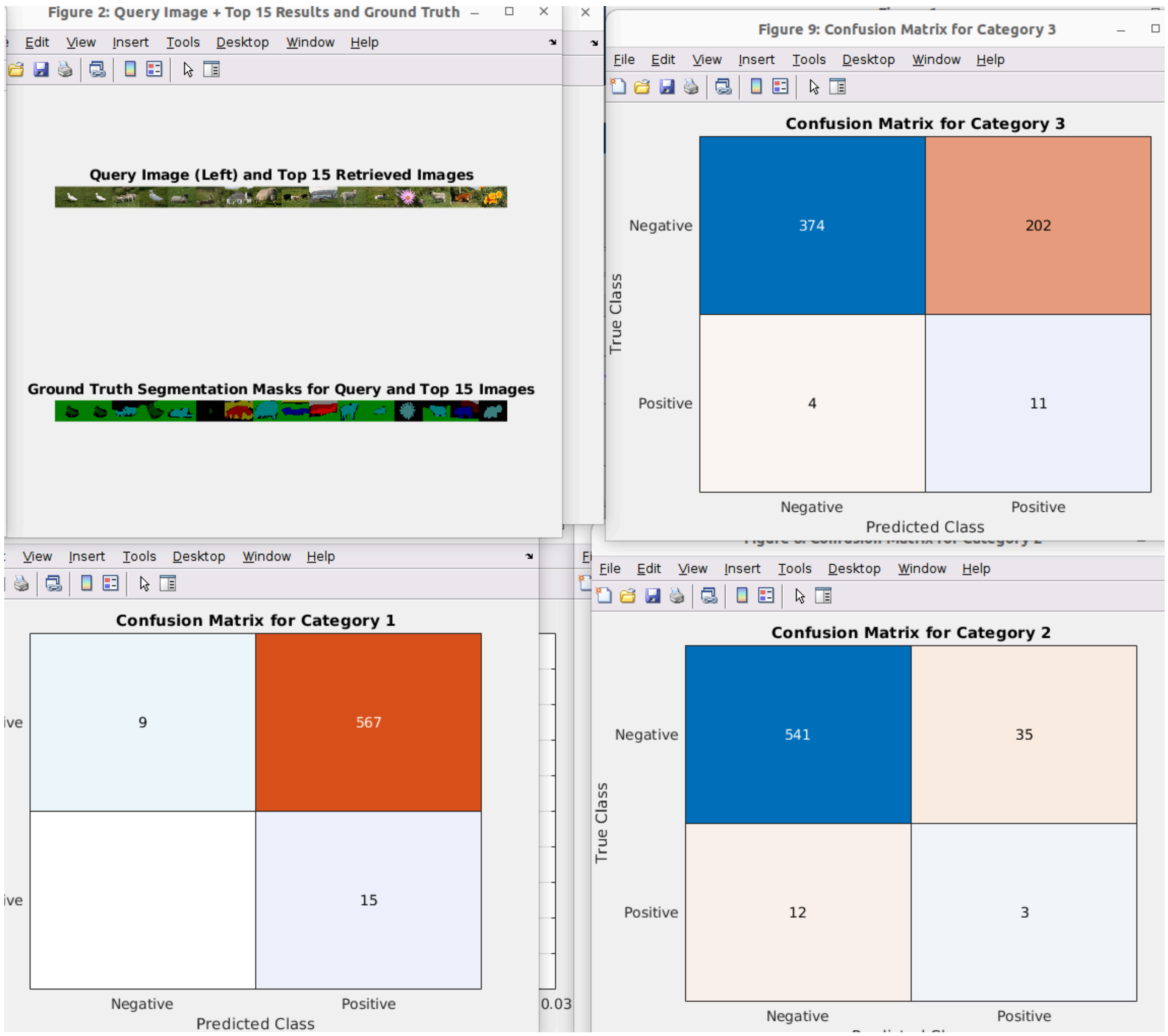

Fig. 3 - Shows Query image + Top 15 Retrieved images (and their ground truths) & Confusion matrix for each category of query image compared to TN, TP, FP, FN for each category of retrieved image + rest of the dataset

Analysis: The evaluation methodology effectively combined quantitative metrics and visual inspection to assess the retrieval performance of different visual search techniques. The use of PR curves and confusion matrices helped identify strengths and weaknesses in the implemented methods, guiding further optimisations.

3.3 Spatial Grid Features Results

Experiments Conducted:

- **Grid Sizes**: 2x2, 4x4, and 8x8 grids.

- **Features Extracted**: Colour, texture (using Sobel, Gabor, and LBP filters), and combined colour and texture.

Findings:

- **2x2 Grid**: Introducing spatial grids enhanced performance by capturing regional details. Combined Averages Across All Categories: Combined Average Precision: 0.7379, Combined Average Recall: 0.0374, Combined Average F1-Score: 0.0640.
- **4x4 Grid**: Finer grids provided more detailed spatial information, further improving results. Combined Averages Across All Categories: Combined Average Precision: 0.9264, Combined Average Recall: 0.0716, Combined Average F1-Score: 0.1188.
- **8x8 Grid**: Larger grids provided less detailed spatial information, reducing performance. Combined Averages Across All Categories: Combined Average Precision: 0.5204, Combined Average Recall: 0.0177, Combined Average F1-Score: 0.0319.

Analysis: Spatial grids significantly improved retrieval performance by preserving spatial layouts. The 4x4 grid showed the best overall results, highlighting the importance of balanced spatial detail. Combining colour and texture features was beneficial, with Sobel edge detection appearing to perform best.

3.4 PCA Results

Experiments Conducted:

- **PCA Application**: Reduced feature dimensionality to retain 80-90% variance.
- **Distance Measures**: Euclidean distance.

Findings:

- **Without PCA**: Combined Averages Across All Categories: Combined Average Precision: 0.9264, Combined Average Recall: 0.0716, Combined Average F1-Score: 0.1188.
- **With PCA**: Combined Averages Across All Categories: Combined Average Precision: 0.9134, Combined Average Recall: 0.0712, Combined Average F1-Score: 0.1182.

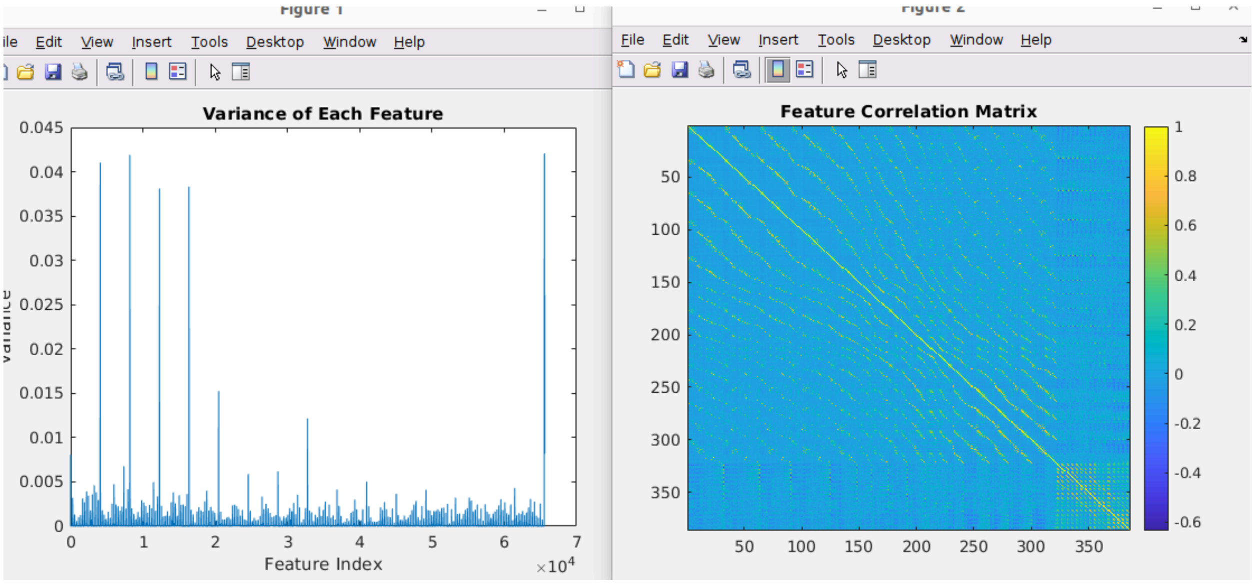

Fig. 4 - Variance of each feature & correlation matrix to help choose values for PCA and pre-processing of data for PCA.

Analysis: PCA effectively reduced feature dimensionality, making computations more efficient with only a slight decline in precision and recall. The overall performance remained similar, with PCA retaining around 84.07% of the variance explained by the first 100 components.

3.5 Distance Measures Results

Experiments Conducted: Different distance measures were implemented in the cvpr_compare function to compare feature vectors (F1 and F2) using the following metrics:

- **Euclidean Distance**: Also known as the L2 norm, it measures the straight-line distance between feature vectors by computing the square root of the sum of squared differences.
- **Manhattan Distance**: Known as the L1 norm, it calculates the sum of the absolute differences between feature vectors, offering robustness to outliers.
- **Cosine Similarity**: Measures the cosine of the angle between vectors, emphasising orientation rather than magnitude. It ranges from -1 to 1, where higher values indicate greater similarity.
- **Chebyshev Distance**: Also known as the L∞ norm, it captures the maximum difference across any dimension, emphasising the largest deviation between feature vectors.
- **Bray-Curtis Dissimilarity**: Measures the proportional difference between feature vectors by taking the sum of absolute differences divided by the sum of absolute values of the vectors, making it suitable for data involving relative comparisons.

Findings:

- **Euclidean Distance**: Combined Average Precision: 0.6859, Combined Average Recall: 0.0439, Combined Average F1-Score: 0.0778.
- **Manhattan Distance**: Combined Average Precision: 0.7373, Combined Average Recall: 0.0480, Combined Average F1-Score: 0.0847.
- **Cosine Similarity**: Combined Average Precision: 0.5735, Combined Average Recall: 0.0308, Combined Average F1-Score: 0.0537.
- **Chebyshev Distance**: Combined Average Precision: 0.6150, Combined Average Recall: 0.0436, Combined Average F1-Score: 0.0734.
- **Bray-Curtis Dissimilarity**: Combined Average Precision: 0.6916, Combined Average Recall: 0.0587, Combined Average F1-Score: 0.0959.

Analysis: Manhattan distance provided slightly better retrieval performance compared to other distance measures, likely due to its robustness to outliers. Bray-Curtis dissimilarity also performed well, highlighting its effectiveness for datasets where relative comparisons are crucial.

3.6 Bag of Visual Words Retrieval Results

Experiments Conducted:

- **Feature Detector**: SIFT, SURF, BRISK.
- **Codebook Size**: 500 and 1000 clusters.

Findings:

- **SIFT**:
  - **500 Clusters**: Average Precision: 0.03, Average Recall: 0.05, F1-Score: 0.04.
  - **1000 Clusters**: Average Precision: 0.03, Average Recall: 0.04, F1-Score: 0.04.
- **SURF**:
  - **500 Clusters**: Average Precision: 0.02, Average Recall: 0.02, F1-Score: 0.02.
  - **1000 Clusters**: Average Precision: 0.03, Average Recall: 0.04, F1-Score: 0.03.
- **BRISK**:
  - **500 Clusters**: Average Precision: 0.03, Average Recall: 0.05, F1-Score: 0.03.
  - **1000 Clusters**: Average Precision: 0.02, Average Recall: 0.04, F1-Score: 0.03.

Analysis: The BoVW model was evaluated using three different feature detectors: SIFT, SURF, and BRISK. However, the results across all three detectors and codebook sizes were quite similar, indicating that none of the detectors provided a significant advantage over the others in this setup. The increase in codebook size from 500 to 1000 clusters showed minimal impact on performance, suggesting that a larger visual vocabulary did not enhance feature representation significantly for this dataset. All three detectors exhibited similar limitations, pointing towards the need for alternative approaches or further optimisations in feature extraction.

3.7 Object Classification Using SVM Results

Experiments Conducted:

- **Classifier**: Multi-class SVM with linear kernel.
- **Feature Representation**: SPM-enhanced BoVW with PCA.

Findings:

- **Overall Classification Accuracy**: 43.90%
- **Class-wise Performance**:
  - **Building**: Precision: 0.32, Recall: 0.42, F1-Score: 0.36.
  - **Grass**: Precision: 0.43, Recall: 0.69, F1-Score: 0.53.
  - **Tree**: Precision: 0.46, Recall: 0.61, F1-Score: 0.52.
  - **Cow**: Precision: 0.00, Recall: 0.00, F1-Score: 0.00.
  - **Sky**: Precision: 0.00, Recall: 0.00, F1-Score: 0.00.
  - **Aeroplane**: Precision: 0.25, Recall: 0.17, F1-Score: 0.20.
  - **Water**: Precision: 0.00, Recall: 0.00, F1-Score: 0.00.
  - **Face**: Precision: 0.00, Recall: 0.00, F1-Score: 0.00.
  - **Car**: Precision: 0.67, Recall: 0.29, F1-Score: 0.40.
  - **Bicycle**: Precision: 0.20, Recall: 0.12, F1-Score: 0.15.
  - **Flower**: Precision: 0.18, Recall: 0.33, F1-Score: 0.24.
  - **Sign**: Precision: 0.78, Recall: 0.88, F1-Score: 0.82.
  - **Bird**: Precision: 0.50, Recall: 0.20, F1-Score: 0.29.
  - **Book**: Precision: 0.75, Recall: 0.50, F1-Score: 0.60.
  - **Chair**: Precision: 0.00, Recall: 0.00, F1-Score: 0.00.
  - **Road**: Precision: 0.88, Recall: 0.70, F1-Score: 0.78.

    - **Cat**: Precision: 0.00, Recall: 0.00, F1-Score: 0.00.
- **Macro-Averaged Precision**: 0.32
- **Macro-Averaged Recall**: 0.29
- **Macro-Averaged F1-Score**: 0.29

Analysis: The classifier achieved an overall accuracy of 43.90%, with considerable variation across different classes. Classes such as "Sign" and "Road" performed well, while others, like "Cow" and "Sky," had zero recall, indicating insufficient training examples or descriptor limitations. The performance varied significantly, highlighting the need for better feature extraction or more balanced training data.

3.8 Spatial Pyramid Matching with BoVW Results

Experiments Conducted:

- **Feature Representation**: SPM-enhanced BoVW with multiple pyramid levels (1x1, 2x2, 4x4).
- **Classifier**: Multi-class SVM with linear kernel.

Findings:

- **Overall Classification Accuracy**: 37.20%
- **Class-wise Performance**:
    - **Building**: Precision: 0.21, Recall: 0.75, F1-Score: 0.33.
    - **Grass**: Precision: 0.51, Recall: 0.71, F1-Score: 0.59.
    - **Tree**: Precision: 0.50, Recall: 0.56, F1-Score: 0.53.
    - **Cow**: Precision: 0.00, Recall: 0.00, F1-Score: 0.00.
    - **Sky**: Precision: 0.00, Recall: 0.00, F1-Score: 0.00.
    - **Aeroplane**: Precision: 0.50, Recall: 0.25, F1-Score: 0.33.
    - **Water**: Precision: 0.00, Recall: 0.00, F1-Score: 0.00.
    - **Face**: Precision: 0.00, Recall: 0.00, F1-Score: 0.00.
    - **Car**: Precision: 0.17, Recall: 0.08, F1-Score: 0.11.
    - **Bicycle**: Precision: 0.00, Recall: 0.00, F1-Score: 0.00.
    - **Flower**: Precision: 0.00, Recall: 0.00, F1-Score: 0.00.

    - **Sign**: Precision: 0.75, Recall: 0.43, F1-Score: 0.55.
    - **Bird**: Precision: 1.00, Recall: 0.10, F1-Score: 0.18.
    - **Book**: Precision: 0.00, Recall: 0.00, F1-Score: 0.00.
    - **Chair**: Precision: 0.00, Recall: 0.00, F1-Score: 0.00.
    - **Road**: Precision: 0.50, Recall: 0.10, F1-Score: 0.17.
    - **Cat**: Precision: 0.00, Recall: 0.00, F1-Score: 0.00.
- **Macro-Averaged Precision**: 0.24
- **Macro-Averaged Recall**: 0.17
- **Macro-Averaged F1-Score**: 0.16

Analysis: The SPM approach, when combined with BoVW, resulted in a classification accuracy of 37.20%. The inclusion of spatial information improved performance for some classes, such as "Building" and "Grass," but overall performance remained low for other classes. The results indicate that while SPM can enhance spatial context, the feature extraction and representation were still insufficient for achieving high accuracy across all categories. Classes like "Cow," "Sky," and "Cat" had zero recall, highlighting significant challenges in differentiating these categories. The average classification accuracy and macro-averaged metrics suggest that further improvements in feature extraction, classifier tuning, or additional data preprocessing may be needed to enhance overall system performance.

Section 4: Conclusions Drawn from Experiments

The comprehensive evaluation of various visual search techniques on the MSRCv2 dataset highlights the complexities of balancing feature representation, dimensionality reduction, and computational efficiency. Key conclusions drawn from the experiments include:

4.1 Global Colour Histogram (GCH)

The GCH approach, while effective for colour-dominated categories, demonstrated limitations in capturing spatial context, which affected its discriminative ability in complex scenes. Performance improved as quantisation levels increased up to 16, but further increases led to diminishing returns, emphasising the challenge of balancing detail with computational efficiency. The average precision and recall metrics indicated that GCH, though simple, lacked the necessary depth for intricate scene analysis.

4.2 Spatial Grid Features

Spatial grids improved feature representation by capturing regional details of an image. The 4x4 grid achieved the best overall performance, indicating that a balanced spatial granularity is crucial for enhancing retrieval effectiveness. Combining colour and texture features, particularly using Sobel filters, further enhanced classification results. However, increasing the grid size to 8x8 did not yield additional benefits and even degraded performance, showing the importance of selecting an optimal grid size for preserving relevant spatial information without overwhelming the model.

4.3 Principal Component Analysis (PCA)

PCA was effective in reducing feature dimensionality while retaining significant data variance. The slight decline in precision and recall after applying PCA suggests that dimensionality reduction comes with minimal trade-offs in retrieval accuracy. The integration of PCA contributed to a more computationally efficient process without significantly sacrificing the overall classification performance, demonstrating its value in managing large feature sets.

4.4 Distance Measures

The exploration of different distance measures underscored the importance of choosing an appropriate metric based on the feature characteristics. Manhattan distance performed slightly better than others, likely due to its robustness to outliers. Bray-Curtis dissimilarity also provided strong results, particularly for datasets requiring proportional comparisons. The varying performance of distance measures highlights the need for careful selection based on the specific feature extraction and representation approach employed.

4.5 Bag of Visual Words (BoVW)

The BoVW model, evaluated with SIFT, SURF, and BRISK feature detectors, showed similar results across detectors, indicating that none of them had a significant advantage in this setup. Increasing the codebook size from 500 to 1000 clusters provided marginal improvements, suggesting limited benefits beyond a certain visual vocabulary size. The low overall precision and recall indicate that while BoVW can capture local features effectively, its implementation in this context was insufficient for achieving high retrieval accuracy.

4.6 Spatial Pyramid Matching (SPM) with BoVW

SPM was introduced to add a spatial hierarchy to the BoVW representation, attempting to capture both global and local spatial arrangements. While SPM slightly improved classification performance for some classes, such as "Building" and "Grass," overall accuracy remained limited, with many categories showing poor recall. The additional spatial context did not overcome the fundamental limitations of the feature extraction techniques, indicating that further optimisation or the use of more advanced features might be necessary to achieve significant gains in classification performance.

4.7 Object Classification Using SVM

The multi-class SVM classifier achieved varying results across different categories. Some classes, like "Sign" and "Road," achieved high precision and recall, while others, such as "Cow" and "Sky," showed zero recall, suggesting that the features used were inadequate for distinguishing these categories. The classifier's overall accuracy of 43.90% reflects the challenges in achieving consistent performance across a diverse set of image classes.

4.8 Overall Insights

The integration of spatial information, dimensionality reduction, and careful selection of distance metrics played a crucial role in advancing the visual search system's effectiveness. However, the results indicate that the implemented feature extraction methods were not sufficiently robust to handle the complexity of the dataset, especially for categories with subtle distinguishing features. Future research directions could include the adoption of deep learning-based features, improved feature extraction techniques, and addressing class imbalances through data augmentation or advanced sampling methods to enhance overall system performance.

Bibliography (1 page)

1. M. Bober and J. Collomosse, *Computer Vision and Pattern Recognition*, EEE3032 - Computer Vision and Pattern Recognition, University of Surrey, 2024.
2. MathWorks, Image Classification Using Bag of Features, 2016. [Online]. Available: https://www.mathworks.com/content/dam/mathworks/tag-team/Objects/i/88400_93009v00_Image_Class_Bag_Features_2016.pdf. [Accessed: 12-Oct-2024].
3. M. Awad and R. Khanna, "Support Vector Machines for Classification," 2015. doi: 10.1007/978-1-4302-5990-9_3.
4. Cogneethi, "Spatial Pyramid Matching in MATLAB," YouTube, 11-Apr-2019. [Online]. Available: https://www.youtube.com/watch?v=6MwuK2wHlOg&ab_channel=Cogneethi. [Accessed: 12-Oct-2024].
5. MathWorks, "Pairwise Distance Metrics in MATLAB," MathWorks Documentation. [Online]. Available: https://uk.mathworks.com/help/stats/pdist.html. [Accessed: 12-Oct-2024].