UNIVERSITY OF SURREY

EEE3032 – Coursework Assignment
Visual Search of an Image Collection

Rohit Krishnan, URN: 6839323, `r.00088@surrey.ac.uk`

November 14, 2023

Abstract

This paper investigates a comprehensive approach to improve the effectiveness of visual search within large image collections. The research focuses on evaluating the precision and recall of various image descriptors and also analysing its impact on performance after dimensionality reduction.

The evaluation is conducted on the Microsoft MSRC V2 dataset, comprising 591 images across 20 categories. The study delves into key feature extraction methods, including Global Color Histogram, Spacial Grid with Color and Texture Histogram, and Convolutional Neural Network (CNN). Evaluation metrics such as precision and recall are employed to assess the performance of these techniques. The experiments reveal insights into the effectiveness of different feature extraction methods for image retrieval.

Notable findings include how effective Convolutional Neural Networks are feature extraction, highlighting the strengths and limitations of each approach. The results contribute to the understanding of image retrieval systems' performance, emphasizing their significance in diverse domains such as medical imaging, multimedia content management, and e-commerce.

This research lays the groundwork for further exploration and optimization of content-based image retrieval techniques, aiming to enhance the capabilities of systems that play a pivotal role in managing the ever-expanding landscape of digital image data.

Contents

1 Introduction

The exponential growth of digital image data has made it necessary to develop efficient systems for retrieving relevant images from vast repositories. Image retrieval systems play a crucial role in various domains, including medical imaging, multimedia content management, and e-commerce. One of the most notable applications of visual search is Google Lens.

CBIR (Content-Based Image Retrieval) and TBIR (Text-Based Image Retrieval) are two approaches to retrieving images from a database based on their content or associated textual information. This paper explores Content-Based Image Retrieval.

Image retrieval systems need to be able to identify features of an image and generate descriptors. These descriptors are then used to compare the features of different images. Based on these comparisons, the image retrieval system ranks the images according to its similarity. Figure 1 gives an overview of how a simple image retrieval system works.

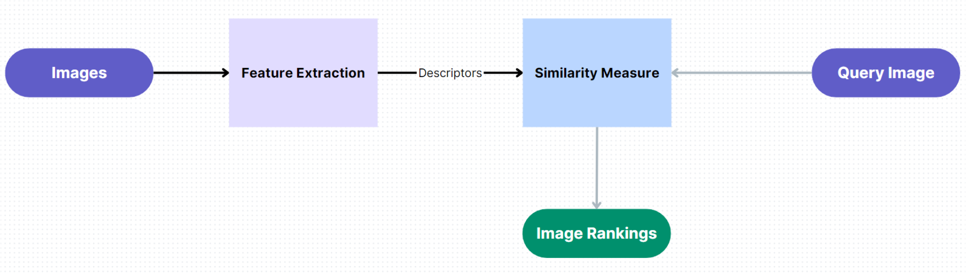

Figure 1: Flowchart of a simple image retrieval system

2 Visual Search Techniques

2.1 Evaluation Methods

It is crucial to assess an image retrieval system in order to fully understand its performance. A variety of metrics and approaches are used to assess an image retrieval system's performance. The experimental findings in this paper are evaluated using precision and recall metrics.

Precision of a system is determined by counting the number of retrieved items that are actually relevant. It is the proportion of all images retrieved that are relevant to all images retrieved. A high precision value means that a large proportion of the retrieved items are indeed relevant.

$$Precision = \frac{\text{Number of Relevant Items Retrieved}}{\text{Total Number of Items Retrieved}} \tag{1}$$

Recall measures how well the system can identify and retrieve all relevant images. It is the ratio of relevant images retrieved to the total number of relevant images in the database. A higher recall indicates that the system can find a larger proportion of relevant images.

$$Recall = \frac{\text{Number of Relevant Items Retrieved}}{\text{Total Number of Relevant Items}} \tag{2}$$

2.2 Feature Extraction

Images are 2-D arrays containing the pixel intensities for each color channel. However it is difficult to compare images based on their pixel intensities. This is because the pixel intensities are not aware of the subject in the image and fail to explain the subject's scale, shape, etc. Feature extractors are functions that can effectively analyse most of the image's features and then, output descriptors which are used to compare images.

2.2.1 Average Color

Average color is a very simple descriptor that outputs an array of three numbers representing the mean intensities of the red, green and blue channels of the image. However, this descriptor does not represent the subject of the image or its shape and edge information. The average color descriptor ignores a lot of information and as a result it is only used as a sub-descriptor (see section 2.2.4 for applications of the Average Color descriptor).

2.2.2 Global Color Histogram

The Average Color descriptor outputs the exact mean of the image's intensity values. This does not work well in real-world problems because the intensity values in an image can always vary depending on the lighting conditions of the scene. The Global Color Histogram is a descriptor that outputs a histogram of color distributions. The descriptor defines similarity as ”similar overall color”. This makes the descriptor tolerant to slight intensity variations between images.

The Global Color Histogram [3] descriptor outputs a descriptor that represents the overall color distribution of the image. The color histogram for a image is calculated by quantizing the color space into several bins. Equation 3 shows how the RGB intensities of the image are quantized, where q is the Quantization level. The pixels of the image are then categorized into these bins, which are then counted to output the color histogram descriptor.

$$r' = floor(r \times \frac{q}{256}) \qquad g' = floor(g \times \frac{q}{256}) \qquad b' = floor(b \times \frac{q}{256}) \tag{3}$$

2.2.3 Edge Histogram

Edge Histogram descriptor considers two images similar if they have similar edge orientations. The descriptor uses an edge detection algorithm to find significant changes in intensity (usually edges). The edges are then thresholded to include only the prominent edges in the image. A histogram is calculated based on the orientation of these edges. Edge Histogram can represent texture information in an image and is often used in combination with other descriptors. This paper uses Sobel [4] for edge detection. Figure 2 shows how the Sobel filter extracts edge information.

Figure 2: Image followed by a visualization of its edge magnitudes

2.2.4 Spacial Grid

All the descriptors mentioned above do not take the spacial information from the image into consideration. This means the intensity values on the top left of the image cannot be differed from similar intensities values on the bottom right of the image. Spacial Grid solves this problem by splitting the image into a 2-D grid and calculating descriptors for each cell.

Spacial grid can be used with many other descriptors to make the resulting descriptor sensitive to spacial information. A suitable descriptor is used on every cell in the image grid and the resulting grid of descriptors is flattened into a 1-D array. This paper explores the use of spacial grid with Average Color, Edge Histogram and a combination of both.

2.2.5 Convolutional Neural Network

Convolutional Neural Networks or CNNs are powerful architectures that can process images highly effectively for computer vision tasks such as image classification, object detection and feature extraction. CNNs automatically extract image features in the first few layers of its network. The outputs of these layers can be used as an image descriptor. For example, in this paper a CNN known as AlexNet [2] was used as a feature extractor (using outputs of AlexNet's "fc7" layer).

2.3 Similarity Measure

The descriptors from feature extractors can be compared to measure the similarity between two images. Image retrieval systems then sort images based on this measure. There are various similarity measures, however this paper only uses L1, Euclidean and Mahalanobis distance functions as similarity measures.

2.3.1 L1 Distance (Manhattan Distance)

The L1 distance, also known as the Manhattan distance or City Block distance, is a measure of the absolute differences between the coordinates of two points in a multidimensional space. The L1 distance between two points a and b is calculated with

$$L1\ Distance(a,b) = \sum_{i=1}^{n} |a_i - b_i| \tag{4}$$

2.3.2 Euclidean Distance (L2 Norm)

Euclidean distance is a measure of the straight-line distance between two points in Euclidean space. It is the most common metric used to quantify the dissimilarity or similarity between two vectors. The Euclidean distance between two points is calculated with

$$Euclidean\ Distance(a,b) = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2} \tag{5}$$

2.3.3 Mahalanobis Distance

Mahalanobis distance is a metric used to measure the distance between a point and a distribution, taking into account the correlations between variables. It is a generalized form of distance that considers the covariance structure of the data. Mahalanobis distance is particularly useful when dealing with multivariate data, where the variables are correlated. It is used often with PCA, a dimensionality reduction technique because it respects the structure of the data.

The Mahalanobis distance, D_M is calculated with

$$D_M(a,b) = \sqrt{\sum_{i=1}^{n} \frac{(a_i - b_i)^2}{v_i}} \tag{6}$$

where, v_i is the eigen value of the i^{th} space.

2.4 Principal Component Analysis

A popular dimensionality reduction technique in statistics and machine learning is principal component analysis, or PCA. It transforms a dataset into a lower dimensional coordinate system, emphasizing directions of maximum variance. The process for dimensionality reduction using PCA involves centering the data, calculating the covariance matrix, selecting the principal components, determining the eigenvectors and eigenvalues, and producing a projection matrix. PCA keeps important information while reducing dimensionality by projecting the data into this subspace. PCA is used for image compression, signal processing, and machine learning for feature reduction and analysis.

3 Experimental results

3.1 Dataset

Experiments from this paper use the Microsoft MSRC V2 dataset [1]. The dataset consists of 591 images classified into 20 different categories. However, there are some images that have incorrect or confusing categories. Because of this the precision values for a given feature extractor might be higher that its actual value.

3.2 Results

3.2.1 Global Color Histogram

One of the feature extraction methods used in the experiments was the Global Color Histogram. Changing the color histogram quantization can be observed in Figure 5 and Color Quantization 5 is noted to have the highest mean average precision. The best average PR curve for the Global Color Histogram descriptor was for quantization 5 as seen in Figure 3.

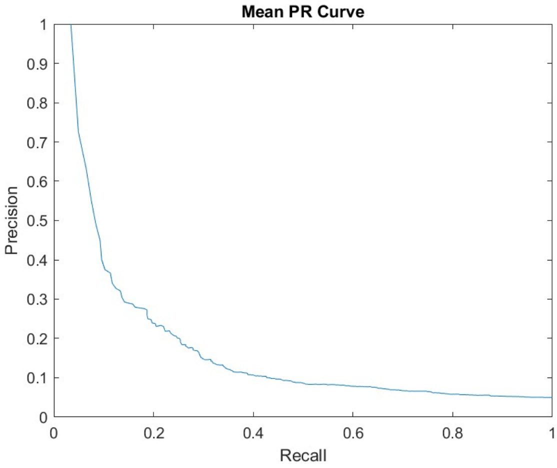

Figure 3: Mean Precision-Recall curve with Q=5

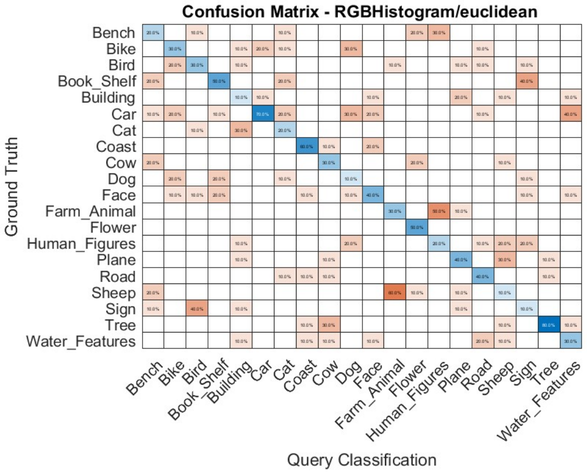

Figure 4: Confusion matrix for Global Color Histogram with Q=5

3.2.2 Spacial Grid with Color and Texture Histogram

Spacial Grid with Color and Texture Histogram was used to explore best grid sizes and color/texture quantizations for the Spacial Grid technique. As shown in Figure 8, a 3x3 grid has the best mean average precision.

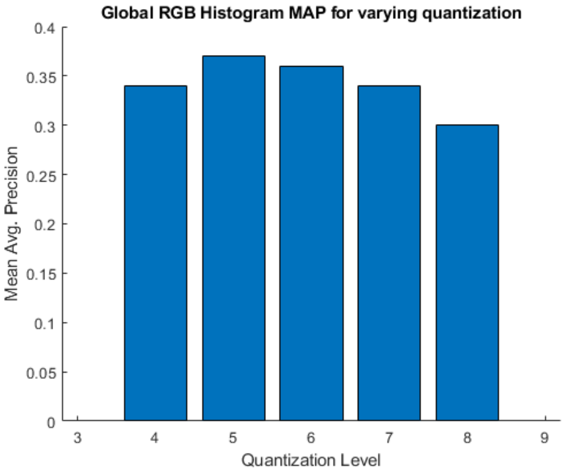

Figure 5: MAP values for varying Color Quantization

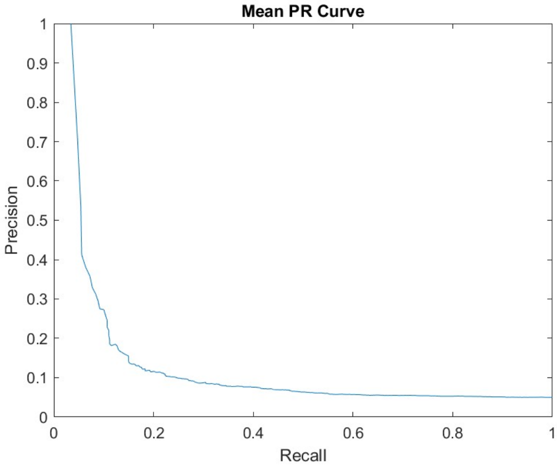

Figure 6: Mean Precision-Recall curve; 3x3 grid; Q=4

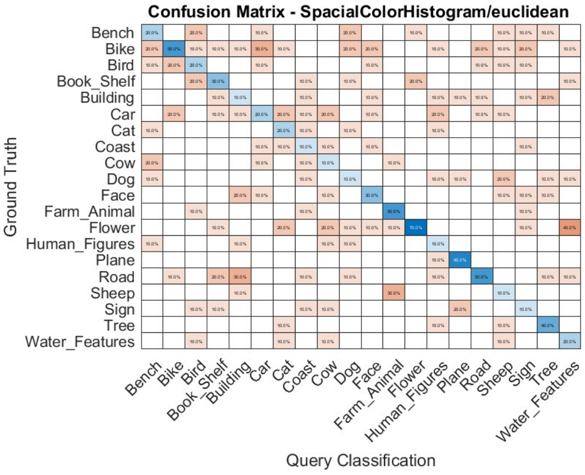

Figure 7: Confusion matrix for Spacial Grid with Color Histogram; 3x3 grid; Q=4

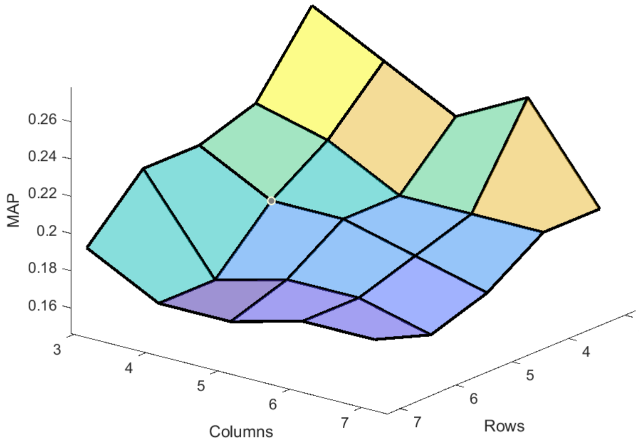

Figure 8: MAP values for varying grid rows and columns

3.2.3 Convolutional Neural Network

Convolutional Neural Networks extracted desirable features. This experiment used AlexNet's fc7 layer to extract feature information. Figure 9 shows an impressive result using CNNs.

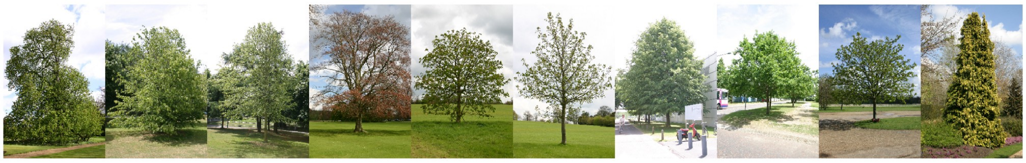

Figure 9: Search output using a CNN descriptor; AlexNet

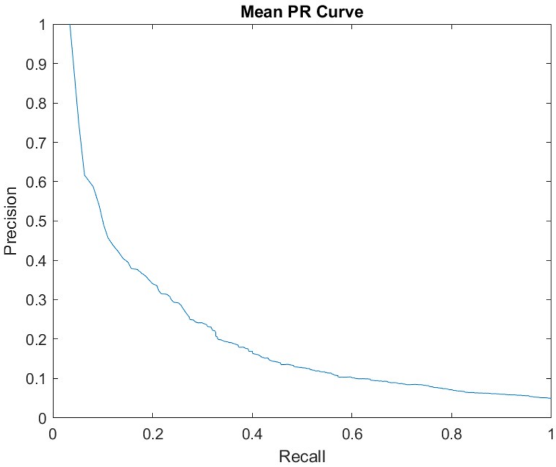

Figure 10: Mean Precision-Recall curve; AlexNet

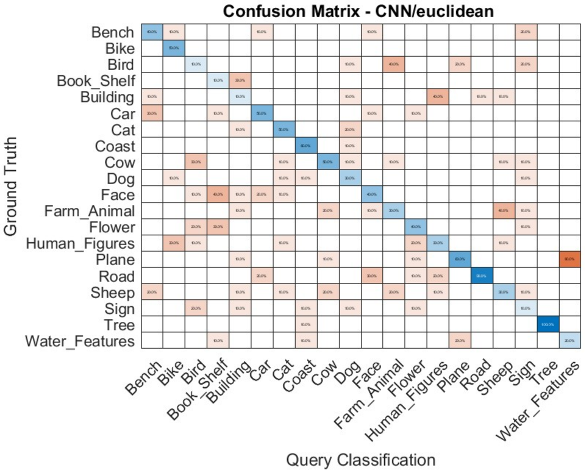

Figure 11: Confusion matrix for CNN feature extractor; AlexNet

4 Conclusion

In conclusion, this paper investigated several content-based image retrieval techniques such as Convolutional Neural Networks, Spacial Grid with Color and Texture Histogram and Global Color Histogram. The evaluation metrics such as precision-recall curves and confusion matrices, provided useful insights into the performance of various descriptors and distance functions.

The results suggests that Convolutional Neural Networks offer substantially good results compared the Global Color Histogram and Spacial Grid. However, there are many other approaches left to explore.

Because of their adaptability, content-based image retrieval systems are useful tools for a variety of industries, including e-commerce, multimedia content management, and medical imaging. They provide effective ways to handle large amounts of digital picture data.

References

[1] Image understanding, Jan 2000. 7

[2] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2012. 5

[3] Jyoti Narwade Manoorkar. Color histograms for content-based image retrieval. volume 1, pages 270–274, 01 2009. 4

[4] R. Woods R. Gonzalez. Digital image processing. pages 414–428. Addison Wesley, 1992. 4