Content Based Image Retrieval System

Srivashinie Dhamodharasamy
CS 510 - Foundations of Computer Vision, Spring 2024
Computer science Department
Portland State University
Email: sridhamo@pdx.edu

Abstract

With the exponential growth of digital image collections, the need for efficient image retrieval systems has become increasingly crucial. With the Content-Based Image Retrieval (CBIR), users can now search for images using characteristics of the visual content, such as color, texture, and shape. This paper explores the CBIR systems' architecture, functional components, and key techniques. In particular, the role of global and local features in CBIR techniques for image retrieval is discussed. The effectiveness of these methods is assessed through a comparative analysis using metrics like precision and response time. The experiments are carried out on a diverse dataset that includes various image types and content to determine the effectiveness of each technique in retrieving relevant images. The findings provide insight into the strengths and limitations of various CBIR approaches. This paper aims to give a comprehensive overview of CBIR systems and their applications.

1 Introduction

With the widespread use of computer networks, the collection of information data, particularly images, has increased rapidly and will continue to grow. The internet's expansion has made it possible to digitize and gather enormous volumes of image data. Massive image databases have been produced as a result, which spans a wide range of applications, including social media, e-commerce, satellite imagery, and medical imaging. The massive amount of image data necessitates the development of efficient indexing and querying techniques.Due to the subjectivity of tagging and the possibility of ambiguity, traditional text-based image retrieval techniques that rely on keywords or tags are frequently insufficient. This has increased demand for more advanced retrieval techniques that can directly analyze and interpret an image's visual content.

Content-Based Image Retrieval (CBIR) has emerged as a powerful solution to this challenge. Introduced in 1990, CBIR automates the retrieval process by analyzing the visual features of images, such as color, texture, shape, and objects, allowing for more accurate and intuitive searches. Unlike text-based methods, CBIR leverages the inherent properties of images to find relevant matches, making it particularly useful in scenarios where textual descriptions fall short.

Computer vision is essential to CBIR systems' operation because it offers the methods and algorithms required for feature extraction, feature representation, and similarity measurement. Convolutional neural networks (CNNs) and deep learning have made significant advances recently, which have improved the accuracy and efficiency of CBIR systems.

The paper is organized as follows: Section 2 discusses the architecture of the CBIR system, Section 3 covers the dataset, Section 4 explores CBIR techniques, Section 5 presents the results, Section 6 summarises the key ideas, and Section 7 provides references.

2 Architecture of CBIR System

This section explores the architecture of a Content-Based Image Retrieval (CBIR). The image retrieval system consists of several key components, each essential for ensuring efficient and accurate image retrieval.

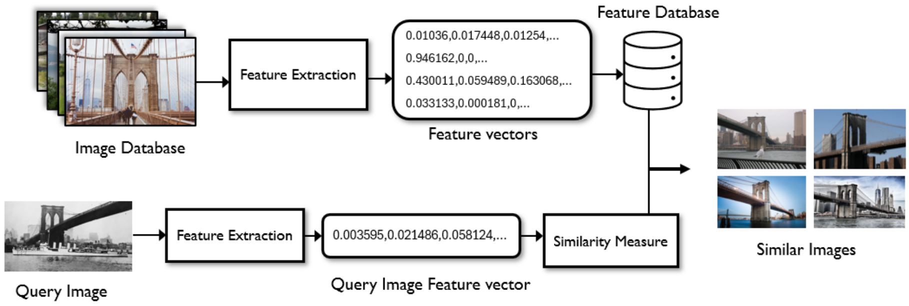

Figure 1: Architecture of Content Based Image Retrieval System

Image Dataset and Preprocessing: The process begins with the collection of an image dataset that needs to be indexed and made searchable. This dataset forms the foundation of the CBIR system, providing the raw data from which features are extracted and compared. After the image dataset is collected, the next step is preprocessing the images before feature extraction. The input image is first resized to a standard size of 256x256 pixels. Then, CLAHE(Contrast Limited Adaptive Histogram Equalization) is applied to enhance the contrast of the image. The image is converted to grayscale before applying CLAHE. Finally, the preprocessed image is returned for further processing. This preprocessing step ensures that the images are standardized and enhanced, making them suitable for feature extraction and comparison in the CBIR system.

Feature Extraction: Feature extraction is a critical process in image retrieval, involving the analysis of an image's visual content to identify distinctive features such as color, texture, shape, and objects. These features are essential for accurately representing the content of images, enabling effective retrieval based on visual similarity.In image retrieval, two main types of features are used: global and local features. Global features provide an overall description of an image, including its color distribution, texture patterns, shape properties, and spatial layout. These features capture the holistic characteristics of an image, making them useful for high-level image retrieval tasks.On the other hand, local features focus on specific regions or keypoints within an image, such as corners, edges, and blobs. These features are more detailed and descriptive, capturing the fine-grained information necessary for matching images based on their local structures. In recent years, Convolutional Neural Networks (CNNs) have emerged as powerful tools for feature extraction in image retrieval. CNNs are capable of learning complex hierarchical features from raw image data, leading to more discriminative and robust representations.

Feature Vectors and Feature Database: After feature extraction, the extracted features are organized into feature vectors, which are mathematical representations of the visual characteristics of an image. These feature vectors encode the extracted features in a structured format that can be easily compared and used for image retrieval.Each feature vector represents a unique image and contains information about its color, texture, shape, and other visual attributes. By organizing features into vectors, we can effectively compare the visual content of different images and retrieve those that are most similar to a query image.These feature vectors are then stored in a feature database, which acts as a repository for the image features. These features are then stored in a suitable format, such as a CSV file, a relational database (RDBMS), or a key-value store like Redis. The feature database is indexed to facilitate quick and accurate similarity measurements between query images and the images in the database.This comparison helps rank the images in the database based on their similarity to the query image, enabling the system to retrieve and present the most relevant images to the user.

Query Image Processing: When a user submits a query image to the CBIR system, it undergoes preprocessing to standardize its representation with dataset images, ensuring accurate comparison. The query image is resized, undergoes color space conversions if needed, and has its contrast and brightness adjusted for improved feature extraction. These steps align the query image with the dataset images. Subsequently, features are extracted from the query image using the same method as the dataset images, ensuring compatibility with the feature database. This consistent preprocessing and feature extraction process enables the system to accurately compare the query image with dataset images.

Similarity Measurement: In image retrieval, the feature vectors of a query image are compared with those stored in the feature database using various distance metrics and similarity measures to compute similarity scores. These scores are essential for ranking images based on their similarity to the query image.Distance measures, such as Euclidean distance, Manhattan distance, and Cosine similarity, calculate the dissimilarity between two feature vectors. A smaller distance value indicates a higher similarity between images, making these measures suitable for many image retrieval tasks. On the other hand, similarity metrics directly measure the similarity between feature vectors, providing a more nuanced understanding of similarity. These metrics often involve advanced mathematical and statistical methods to compare specific characteristics extracted from images, making them particularly useful for complex image retrieval tasks.In our project, we have utilized the Chi-square metric and the BFMatcher algorithm to enhance our image retrieval system. The Chi-square metric quantifies dissimilarity between histogram functions, allowing for accurate comparison of image features. The BFMatcher algorithm, or Brute Force Matcher, is specifically designed for matching ORB descriptors between a query image and dataset images, ensuring comprehensive matching through exhaustive comparisons.By integrating these techniques, we have significantly improved the efficiency and accuracy of our image retrieval system, enabling us to retrieve the most relevant images based on visual similarity.

Image Retrieval and Presentation: The system retrieves and presents the most similar images to the user, enabling an effective search experience. This structured approach ensures that the system can handle large volumes of image data efficiently and provide accurate results based on visual content analysis.

3 Dataset

The GPR1200 (General Purpose Retrieval) dataset is a curated collection of 12,000 images spanning 1,200 classes across six distinct image domains, created to serve as a comprehensive benchmark for assessing the general retrieval capabilities of deep learning models. This dataset aims to serve as an accessible benchmark for the research field of general-purpose content-based image retrieval (CBIR).
Dataset sources:

- Google Landmarks V2: Focused on natural and architectural landmarks.
- ImageNet Sketch: Black and white sketches of animals and objects.
- iNat: Images of plants, animals, insects, and fungi.
- INSTRE: Planar images and photographs of logos and toys.
- SOP: Products and objects, partly isolated.
- IMDB Faces: Images of human faces.

For each domain, 200 classes were selected, each with 10 images, ensuring a uniform class distribution and solvability of the retrieval tasks.The GPR1200 dataset is designed to provide a diverse and challenging benchmark for evaluating the generalization power of CBIR models across various image domains

4 CBIR Techniques

Content-Based Image Retrieval (CBIR) techniques are essential for efficiently retrieving images based on their visual content. These techniques use visual features such as color, texture, and shape to identify similar images in a dataset. This section explores the different CBIR techniques used for image retrieval and their effectiveness in retrieving relevant images based on visual content analysis.

4.1 CBIR using Color Features

Color is a fundamental and widely used low-level visual feature in Content-Based Image Retrieval (CBIR) systems. It is valued for its robustness against changes in image size and orientation, as well as its high expressiveness in image retrieval tasks. The system focuses on utilizing color features to describe and compare images. The system employs the color histogram intersection method to capture the color distribution of images. This method involves extracting Hue, Saturation, and Value (HSV) color features from segmented regions of the image. Representing colors in the HSV color space allows us to capture both the chromatic and brightness information of colors, providing a more comprehensive description of image content.To compare color histograms efficiently and accurately for image retrieval, we use the chi-square distance metric. This metric quantifies the dissimilarity between histograms by considering the frequency distribution of color values. By utilizing the chi-square distance metric, we can effectively measure the similarity between images based on their color content.The system searches for similar images in a dataset based on the color features extracted using the color descriptor. The similarity index is calculated using the chi-squared distance metric between the query image and the indexed image, which measures the difference between two histograms. A lower similarity index signifies that the feature distributions of the two images are more similar, implying that the images are visually more alike. Conversely, a higher similarity index would indicate greater differences between the images, suggesting that they are less visually similar. Thus, the retrieval system prioritizes images with the lowest similarity indices as the most relevant matches to the query image.The system returns the top 5 most similar images from the dataset.

Figure 2: Query Image

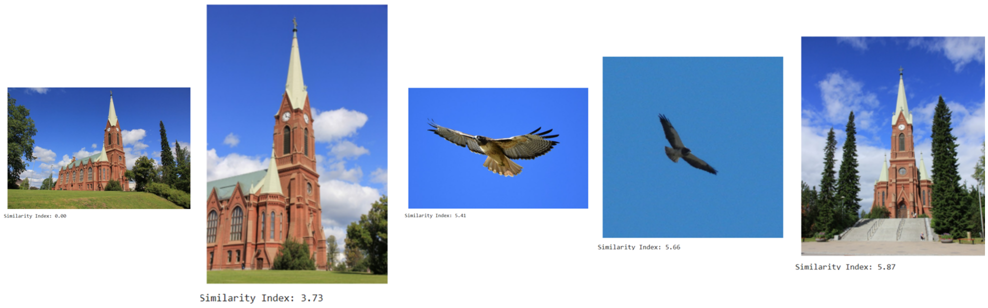

Figure 3: Images Retrived using Color Features

4.2 CBIR using Combined Features

In the combined feature method, ORB descriptors and color histograms are merged to form a hybrid feature representation for image retrieval. First, the ORB detector is used to extract local keypoint descriptors. ORB, or Oriented FAST and Rotated BRIEF, focuses on identifying distinctive features within an image such as corners, edges, and blobs. The similarity score for ORB is based on the number of good matches, with a higher number indicating more similarity. These local features are crucial for capturing the detailed structure and unique characteristics of the image.Next, color histograms are incorporated to represent the overall color composition. The color histogram captures the distribution of colors within an image, providing a global perspective on the image's color characteristics.Color histograms are extracted and compared using the chi-squared distance. Then, the similarity index is calculated by combining both the ORB and color histogram values. This combined similarity measure is then normalized, where a value closer to 1 indicates images that are more similar. By combining local and global features, the method is more robust to various transformations such as rotation, scaling, and illumination changes. The ORB descriptors capture local features that are invariant to these transformations, while color histograms provide a global representation that is less affected by local changes.

Figure 4: Query Image

Figure 5: Images Retrived using Combined features

4.3 CBIR using Pretrained CNN

This method focuses on using a Convolutional Neural Network (CNN) for feature extraction, which is a common approach in CBIR due to CNNs' ability to learn rich and discriminative features from images. CNNs are a class of deep neural networks that consist of multiple layers, including convolutional layers, pooling layers, and fully connected layers. In CBIR, a pre-trained CNN, such as VGG16, is used to extract features from images. These features capture various visual characteristics of the images, such as shapes, textures, and patterns.The VGG16 model is used to extract features from images. Specifically, the 'block5_conv3' layer of the VGG16 model is chosen as it captures high-level features in the images. The extracted features are then flattened to create a feature vector representation for each image. This feature vector represents the image's visual content in a high-dimensional space.To find similar images, the feature vector of a query image is compared to the feature vectors of all images in the dataset. It compares the query image features to the indexed features using cosine similarity. Cosine similarity measures the cosine of the angle between two vectors, with a value closer to 1 indicating higher similarity.

Figure 6: Query Image

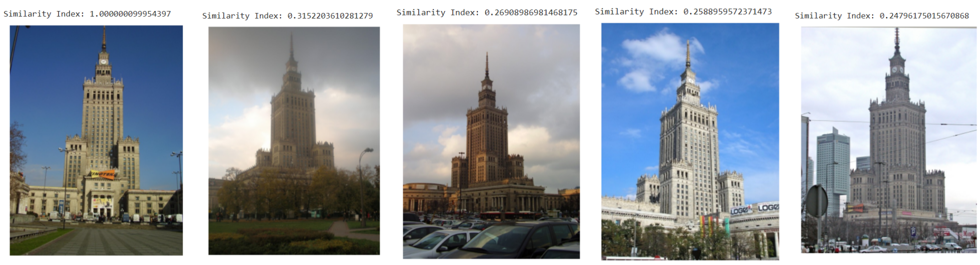

Figure 7: Images Retrived using Pretrained CNN

5 Results

This section presents the results of our image retrieval experiments using different feature extraction techniques and similarity metrics. We evaluate the performance of each approach based on precision and retrieval time.

- **Precision:** Precision is calculated as the number of relevant images retrieved divided by the total number of images retrieved. It measures the accuracy of the retrieval system in returning relevant images. A high precision indicates that a large proportion of the retrieved images are relevant to the query.
- **Retrieval Time:** The retrieval time is the time taken by the system to process the query and retrieve the top similar images. It measures the efficiency of the retrieval algorithm in returning results. A shorter retrieval time indicates a more efficient algorithm.

Technique	Precision	Retrieval Time(in Seconds)
Color Features	0.60	600
Combined Features	0.80	960
Pretrained CNN	0.86	6300

Table 1: Comparison of CBIR Techniques

The experiments compared three methods for image retrieval based on precision and response time. The Pretrained CNN method achieved the highest precision of 0.86 but had the longest response time of 6300 seconds. The Combined Features method balanced precision and response time well, with a precision of 0.80 and a response time of 960 seconds. The Color Features method had the lowest precision of 0.60 but the fastest response time of 600 seconds. These results suggest that the Pretrained CNN method provides the most accurate results but at the cost of longer processing times, while the Combined Features method offers a good balance between accuracy and speed, and the Color Features method prioritizes quick results over accuracy. The choice of method should depend on the specific requirements of the application, with consideration for the trade-off between precision and response time.

6 Conclusion

In summary, this work has shown how various CBIR approaches can be used to effectively retrieve images based on their visual content. The results of the experiments indicate that every technique has advantages and disadvantages, and the best approach should be chosen depending on the particular requirements of the application. Future research in this area could focus on enhancing the efficiency of CBIR systems, improving the accuracy of feature extraction methods, and exploring new similarity measures to further enhance the retrieval performance. Overall, CBIR remains a promising approach to image retrieval, providing intuitive and reliable search capabilities for a variety of applications.

References

[1] Schall, K., Barthel, K. U., Hezel, N., & Jung, K. (2022). GPR1200: A Benchmark for General-Purpose Content-Based Image Retrieval. In MultiMedia Modeling: 28th International Conference, MMM 2022, Phu Quoc, Vietnam, June 6–10, 2022, Proceedings, Part I (pp. 205–216). Berlin, Heidelberg: Springer-Verlag. doi:10.1007/978-3-030-98358-1_17

[2] Shukran, M. A. M., Abdullah, M. N., Yunus, M. S. F. M. (2021). New Approach on the Techniques of Content-Based Image Retrieval (CBIR) Using Color, Texture and Shape Features. *Journal of Materials Science and Chemical Engineering*, 9(1). Retrieved from `https://www.scirp.org/journal/paperinforcitation.aspx?paperid=107008`

[3] Hameed, I. M., Abdulhussain, S. H., Mahmmod, B. M., & Pham, D. T. (2021). Content-based image retrieval: A review of recent trends. *Cogent Engineering*, 8(1). doi:10.1080/23311916.2021.1927469. Retrieved from `https://doi.org/10.1080/23311916.2021.1927469`

[4] OpenCV. (n.d.). Content-Based Image Retrieval. Retrieved from `https://opencv.courses/blog/content-based-image-retrieval/`

[5] Baeldung. (n.d.). Content-Based Image Retrieval (CBIR). Retrieved from `https://www.baeldung.com/cs/cbir-tbir`