Image Search Engine

Aditya Mishra[1], Ashish Kumar[2], Ayush Yadav[3], Aurangajeb Ansari[4], Nidhi Gupta[5]

[1-4]Student, Dept. of Information Technology, Inderprastha Engineering College, UP, India
[5]Asst. Professor, Dept. of Information Technology, Inderprastha Engineering College, UP, India

---***---

Abstract - *In this paper, we put forward a model for a search engine where an image can be uploaded from the local database of the user to retrieve information about it from the existing database. This is similar to the traditional keyword search used by most of the search engines with the only difference being that here an image is uploaded as a query rather than textual keywords. The fact that the image being used as a query makes the search ever more complicated as the content of the image needs to be analyzed and matched to find the information corresponding to the uploaded image. This is most apt for searching for information about images of historical monuments, places, or any specific place or thing that is identifiable.*

1. INTRODUCTION

An image retrieval system is a computer system for searching and retrieving images from a large database of digital images. The purpose of an image database is to store and retrieve an image or image sequences that are relevant to a query. There are a variety of domains such as information retrieval, computer graphics, database management, and user behavior that have evolved separately but are interrelated and provide a valuable contribution to this research subject. Therefore, the design and creation of algorithms and methodologies that can efficiently retrieve image content are the main focus of research in the area of content-based image retrieval. For this reason, many methods rely on the extraction of color or texture descriptors and the organization of that data to determine the similarity of images to one another. Color information is more popular because it can be mapped to a three dimensional coordinate system that closely matches human perception. Color, however, is unsuitable in many cases as there may be images that are grayscale images, which have similar color counterparts for them. Though there are some methods for texture and shape descriptors, they cannot be applied to complex images, which have many tiny details.

2. RELATED WORK

There is a concept of edge detection and the algorithm of the Sobel edge detection technique which lines out the edges of an image at the points where the intensity changes drastically. It also talks about the general concept involved in edge detection techniques and also compares various methods used for edge detection. The main reason for using edge detection is to extract the important information of the image, i.e. it's content.

There is a concept of image processing applied to an image to extract the content from it using a technique, called Color Coherent Matrix (CCV). It defines color coherence as the degree to which pixels of that color a remembers of large similarly colored regions.

There is a technique of retrieving image content by using a combination of Sobel edge detection and Color Coherent Vector matrix algorithms. It focuses mainly on the conceptual understanding of how this proposed technique works and why it is important in refreshing the web documents on the internet.

There is a system of computer-aided browsing for image databases based on the Pathfinder Network. Additionally, they have introduced a new clustering algorithm based on the Hausdorff distance metric for clustering binary images of line segments culled from the edge-analysis of images in a given database. These cluster centers have been used as local shape feature descriptors and the images in the database have then been decomposed into feature vectors and this description has Been used to calculate the similarity of the images in the database to one another.

There is a shape-based retrieval system for use on infrared images. Similar to stand documents retrieval, an indexing structure has been used in the retrieval process, rather than the direct comparison approach, which is more common in content-based image retrieval systems. Retrieval of similar polygons has been performed using a local association hashing method.

3. LITERATURE SURVEY

Using today's high-speed search engines, it is easy to find out information about something we know the name of. But how can one find out about something he doesn't know of. For searching for a monument or person or a place, it is necessary to know what it is called so that a text query can be formed. But in a case where one only has an image of the object/ monument/ person, how can a text query be formed? The design of the search engine proposed here holds a possible solution to this problem: The user uploads an image from the local databases, the one about which he wishes to find some information. This is a unique functionality, which gives information to the user by extracting the content of the uploaded image to match it with that of images on the internet. It collects all those images whose content matches that of the uploaded image. It then ranks those images based on the percentage of the content found similar to the uploaded image and displays their corresponding websites right below the image. Clicking on the web link will take the person to the page that contains information about the image. Keeping the above-mentioned purpose of extracting the content of these images in mind, and comparing them, a combination of the techniques mentioned below have been used:

3.1. An edge detection technique called the Sobel edge Detection has been used to extract the content of the uploaded and web images.

3.2. The next step involves comparing the extracted content with the help of a method called the CCV matric technique, which checks for the content similarity between the images being compared.

3.3. Using an overlap of the mentioned algorithms, a new image content extraction technique comes into existence.

4. METHODOLOGY

When we building image search engine we will first have to index our data set. Indexing a data set is the process of quantifying our data set by utilizing an image. An **image descriptor** defines the algorithm that we are utilizing to describe our image.

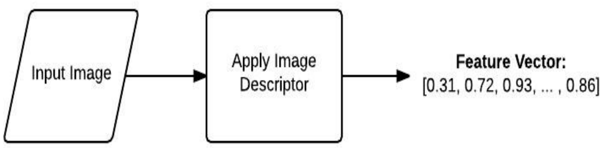

Fig 1: *Working of Image descriptor.*

Here we are present with an input image, we apply our image descriptor, and then our output is a list of features used to quantify the image. Feature vectors can then be compared for similarity by using a distance metric or similarity function. Distance metrics and similarity function. Distance metrics and similarity functions take two feature vectors as inputs and then output a number that represents how "similar" the two feature vectors are.

The Figure below visualizes the process of comparing two images.

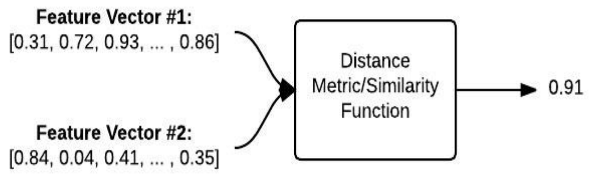

Fig 2: *Comparing two image.*

The four Steps of Any CBIR System: No matter what Content-Based Image Retrieval System we are building, they all can be boiled down into four distinct steps:

i. Defining your image descriptor: At this phase, we need to decide what aspect of the image we want to describe. Are we interested in the color of the image? The shape of an object in the image? Or do we want to characterize texture?

ii. Indexing our data set: Now that we have your image descriptor defined, our job is to apply this image descriptor to each image in your data set, extract features from these images, and write the features to storage (ex. CSV file, RDBMS, Redis, etc.) so that they can be later compared for similarity.

iii. **Defining our similarity metric:** Now we have a bunch of feature vectors. But how are we going to compare them? Popular choices include the Euclidean distance, Cosine distance, and chi-squared distance, but the actual choice is highly dependent on our data set and the types of features we extracted.

iv. Searching: The final step is to perform an actual search. A user will submit a query image to your system (from an upload form or via a mobile app, for instance) and our job will be to extract features from this query image and then apply your similarity function to compare the query features to the features already indexed. From there, we simply return the most relevant results according to our similarity function.

Again, these are the most basic four steps of any CBIR system. As they become more complex and utilize different feature representations, the number of steps grow and you'll add a substantial number of sub-steps to each step mentioned above. But for the time being let's keep things simple and utilize just these four steps.

Fig 3: *Flow chart representing the process of extracting features from each image in the data set.*

We start by taking our data set of images, extracting features from each image, and then storing these features in a database.

We can then move on to performing a search.

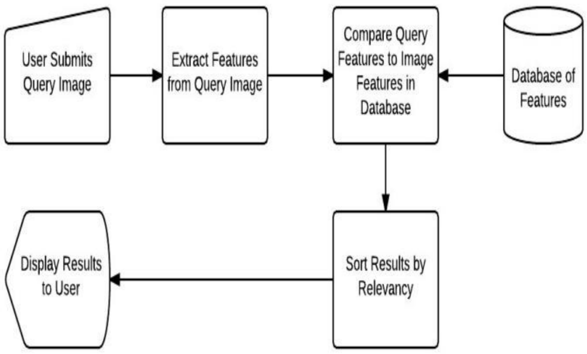

Fig 4: *Performing a search on a CBIR System*.

First, a user must submit a query image to our image search engine. We then take the query image and extract features from it. These "query features" are then compared to the features of the images we already indexed in our data set. Finally, the results are then sorted by relevancy and presented to the user.

5. RESULT

Understand, how we mark a image in a five parts to find a result from a query.

Take a look at the example below where we have submitted an photo of the boats on the water and have found relevant images in our photo collection

Fig 5: Query and result images

In order to build this system, we used a simple, yet effective image descriptor: **the color histogram.**

By utilizing a color histogram as our image descriptor, we will be relying on the color distribution of the image. Because of this, we have to make an important assumption regarding our image search engine.

Assumption: Image that have similar color distributions will be considered relevant to each other. Even if images have dramatically different contents, they will still be considered "similar" provided that their color distribution are similar as well.

6. CONCLUSION

As a conclusion, this paper provides a study of image retrieval work. A wide variety of researches has been made on image retrieval. Each work has its technique, contribution, and limitations. It might not include every aspect of individual works, however, this paper attempts to deal with a detailed review of the most common traditional and modern image retrieval systems from early text-based systems to content-based retrieval.

7. REFERENCES

[1] O. R. Vincent , and O. Folorunso, "A Descriptive Algorithm for Sobel Image Edge Detection," Proceedings of Information Dcience & IT Education Conference 2009.

[2] Divya Ragatha Venkata, Deepika Kulshreshtha, and Divakar Yadav, "Techniques for Refreshing Images in Web Documents," Proceedings for International Conference on Control, Robotics and Cybermetics, October 2011.

[3] Rishi Mukhopadhyay, Aiyesha Ma and Ishwar K. Sethi, "Pathfinder Networks for Content Based Image Retrieval Based on Automated Shape Feature discovery",2012.

[4] Daniella Stan. eID: A system for Exploration of Image Databases, phD thesis, Oakland University, 2002.

[5] Chaomei Chen, George Gagaudakis, and Paul Rosin. Similarity-based image browsing. Proceedings of the 16th IFIP World Computer Congress, August 2000.