ISSN: 2582 - 6379
Orange Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2020, Volume 1 Issue 2
www.ijisea.org

Content Based Image Retrieval Using Feature Extraction with KNN Algorithm

[1]Yasaswi Tatineni, [2]Dhanooj Kumar Tatikonda, [3]Azarudhin Shaik,
[4]Jayasree Nandipu and [5]Dr.Katarapu Naga Prakash
[1234]Student(s), Department of ECE,
[5]Professor, Department of ECE,
Gudlavalleru Engineering College, Gudlavalleru, A.P., India

ABSTRACT

This paper presents, the Content Based Image Retrieval (CBIR) using feature extraction for extraction of digital images. In this technique of image retrieval, the visual features of an image are used in order to search the user-based query images from database and they are extracted without any human intervention. In this paper, KNN algorithm is used for feature extraction. Though KNN is computationally expensive it gives specific number of images as per the user requirement.

Key words: Content Based Image Retrieval (CBIR), K-Nearest Neighbor (KNN).

I.INTRODUCTION

In the recent past the progression in PC and multimedia technologies has prompted the creation of computerized pictures and modest enormous picture storehouses. The size of picture collections has increased quickly because of this, including computerized libraries, medical pictures and so on. To handle this fast development, it is required to create Picture recovery systems which works for an enormous scope. The essential point is to build a powerful system that creates, manages and query picture databases in an exact way. CBIR [1] is the methodology of automatically ordering pictures by the extraction of their low-level visual features, similar to shape, colour, and texture, and these indexed features are exclusively responsible for the recovery of pictures.

A content Based Image Retrieval (CBIR) is an interface between the semantic gaps which is depicted as the distinction between a human mind (significant level) and a PC system (low-level) discernment. In CBIR, visual picture contents are represented to as picture features, which are extracted utilizing automatically done feature extraction methods. CBIR system have two fundamental steps, the initial step involves the pre-processing of pictures from the picture database. The picture features are extracted from the database pictures and are stored as a feature database of pictures. This step includes an enormous number of computation and is a moving advance as how to extract the picture features. Picture feature extraction is generally done on the basis of colour, shape, texture that are low-level features representing to the picture and numerous methods have been implemented for the equivalent. The second step includes the similarity measures, which is a significant step as our outcomes are recovered after the closeness measure check. The user query picture features are coordinated with the stored feature database that gives us the most ideal match as per the user query picture. A few closeness coordinating techniques have been proposed which generally figure the distance between the query picture and the database pictures. The essential utilization of KNN [2] calculation is that this technique gives us the particular number of results according to the user requirement.

ISSN: 2582 - 6379
Orange Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2020, Volume 1 Issue 2
www.ijisea.org

II.PROPOSED METHODOLOGY

1. CONTENT BASED IMAGE RETRIEVAL

Content based image retrieval [3] is also known as query by image content (QBIC) and content-based visual information retrieval (CBVIR). It is the application of computer vision techniques to the image retrieval problem. "Content-based" means the search analyses the contents of image not the metadata such as keywords, tags, or descriptions associated with that image. The term "content" here might refer to colours, shapes, textures, or any other information that can be derived from the image itself. In CBIR, low-level visual features (e.g., colour, shape and texture) are computed from the query and matching of these features is performed to sort-out the result.

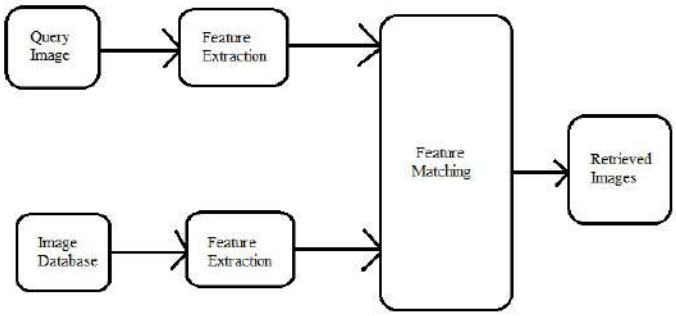

Figure 1 : Shows Basic Block Diagram of CBIR

2.FEATURE EXTRACTION TECHNIQUES

General visual content include color, texture, shape, spatial relationship, etc. Domain specific visual content, like human faces, is application dependent and may involve domain knowledge. Our proposed system is based on General visual content which includes:

a. Color
b. Texture
c. Shape

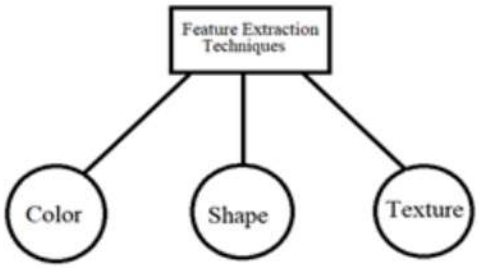

Figure 2: Shows Feature Extraction Techniques

ISSN: 2582 - 6379
Orange Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2020, Volume 1 Issue 2
www.ijisea.org

A. Colour Feature: Colour is the most important visual feature in Content based image retrieval. It depends on the response of human visual system to light and the interaction of light with objects. It is vigorous to background complication and independent of image size and orientation. Colour feature extraction include the key issues like colour space, colour quantization, and the choice of similarity function.

B. Shape Feature: Shape feature [4] is one of the visual features and primitive features for image content description. Shape feature has all the geometrical information of an object in the image which don't change even when the orientation or location of object are changed. Some shape features are the perimeter, area, eccentricity, symmetry, etc. There are plenty of shape descriptors that can be divided into two main categories:

i. region-based and
ii. contour-based

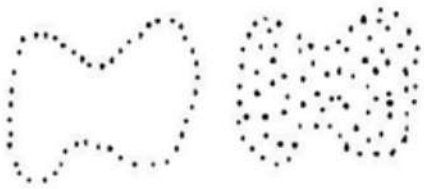

Figure 3: Shows Boundary based and Region based shape representations

C. Texture Feature: Texture feature [5] refers to the visual patterns that have the properties of homogeneity that do not result from the presence of only a single colour or intensity. It is a natural property of virtually all surfaces, including clouds, trees, bricks, hair, and fabrics. Texture feature contains important information about the structural arrangement of surfaces and their relationship to their surrounding environment.

III. MACHINE LEARNING ALGORITHM USED

The ability to perceive and group objects by people is a significant use of the intensity of human sight. After numerous decades the research is going on the best way to incorporate this force into PCs. This is done with the assistance of Machine learning algorithms. we as a whole realize that Machine Learning models makes predictions by gaining from the past information accessible. Example: It is to detect the License plate detection, Pattern recognition, Image classification etc., It is difficult to implement Image classification in machines. It requires utilization of some complex algorithms called K-Nearest Neighbor algorithm [6]. K-Nearest Neighbor is one of the simplest supervised Machine learning algorithms generally utilized for classification. This idea is to scan for nearest match of test information in feature space. It classifies information dependent on how its neighbors are i.e., Firstly, KNN stores all the accessible cases and classifies new cases dependent on similarity measure. To discover the closest neighbors and content comparison is done by utilizing distance measure, there are various distance formulae like Euclidean formulae, City Block Manhattan formulae, Minkowski

ISSN: 2582 - 6379
Orange Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2020, Volume 1 Issue 2
www.ijisea.org

formulae. Generally, Euclidean formulae is considered over the various referenced formulae. In Euclidean distance measure, a image distance measure analyzes the closeness of two pictures in different measurements relating to color, shape, texture features. we will calculate Euclidean distance between query picture and all the pictures in the database. Lastly, the system returns the main K-Closest neighbor pictures that are similar like query picture. The Euclidean separation can be figured as below:

$$D = \sqrt{\sum_{i=1}^{K}(x_i - y_i)^2} \quad \text{---------- (1)}$$

Here, D speaks to the separation between two Picture features.

In KNN, we get all our training information of dataset picked. At that point, whenever new information comes in, we find its separation utilizing Euclidean distance formulae. Let us pick k value to be 3 (for example K=3). Hence, a circle must be attracted spoken to with a blue star in Fig enclosing the closest information focuses over a plane. At that point, we discover K closest neighbours of input query picture.

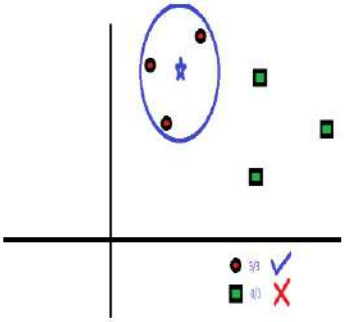

Figure 4: Shows Selecting the nearest Data Points in KNN

IV. RESULTS

DATABASE

Corel database [7] is colour picture database. It contains huge number of pictures of various kinds like animals, seashores, flowers and so on. These pictures are ordered into various classes of having 100 pictures for every classification in database.

A few analysts feel that Corel database meets all the necessities to assess an image retrieval system, in light of its huge size and heterogeneous content. In this paper, we gathered the database DB contains 1000 pictures of 10 distinct classifications. Ten classes are given in the database to be specific Humans, Flowers, Beaches, Elephants, Buildings, Busses, Mountains, Horses, Food and Dinosaurs. Every class has 100 pictures (NG=100) and these have either 256X384 or 384X256 sizes. Figure below depicts the example pictures of Corel 1000 picture database (one picture from every classification).

ISSN: 2582 - 6379
Orange Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2020, Volume 1 Issue 2
www.ijisea.org

Figure 5: Shows Simple Images from Corel 1000 (One Image from Category)

The performance of the proposed method is estimated as far as accuracy and recall by below equations respectively.

$$Precision = \frac{Number\ of\ Relevant\ images\ retrieved}{Total\ Number\ of\ images\ retrieved} \times 100$$

$$Recall = \frac{Number\ of\ Relevant\ images\ Retrieved}{Number\ of\ relevant\ images\ in\ the\ database} \times 100$$

The most common performance measures used in image retrieval are precision and recall. Precision-Recall measure the accuracy of the image retrieval system.

It gives a better indication of what might be a good number of images to retrieve. Precision takes all the retrieved documents into account, but it can also be evaluated at a given cut-off rank, considering only the topmost results returned by the system. Precision is used with recall, the percent of all relevant documents that is returned by the search.

In binary classification, recall is called sensitivity. It can be viewed as the probability that a relevant document is retrieved by the query. Table 1, [8] summarizes up the retrieval results of the proposed technique as far as precision and recall respectively.

ISSN: 2582 - 6379
Orange Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2020, Volume 1 Issue 2
www.ijisea.org

Category	When K=15	
	Precision	Recall
Humans	0.933	0.733
Flowers	1	1
Beaches	0.53	0.48
Elephants	0.705	0.6
Buildings	0.4	0.3
Buses	0.75	0.6
Mountains	0.15	0.1
Horses	1	1
Dinosaurs	1	1
Food	0.666	0.5
Total	7.134	6.313

Table 1: Shows Results in terms of precision and recall on database

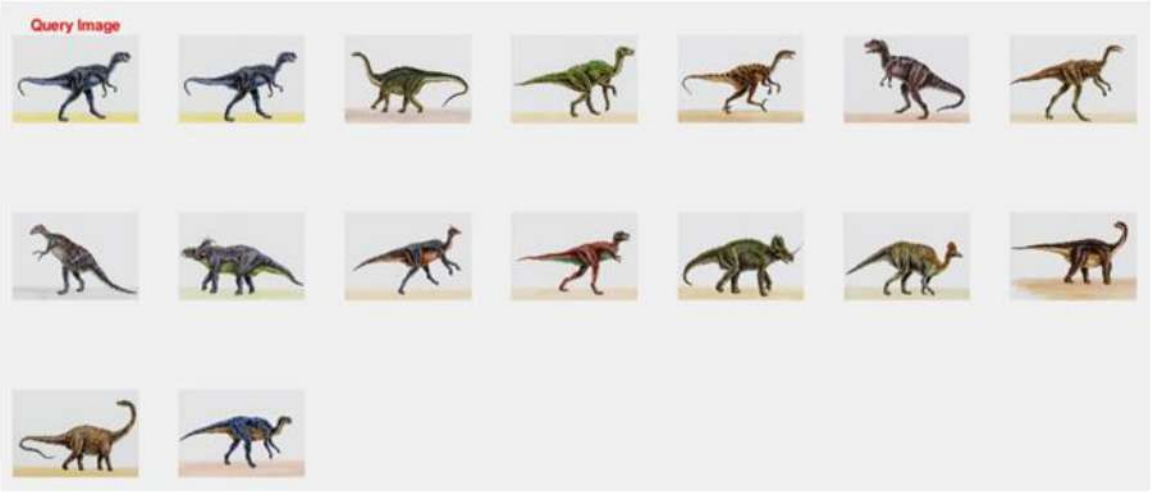

Figure 6 : Retrieved Images Based on the Query Image

V. CONCLUSION

In this research paper, the image retrieval is done using KNN method. Firstly, feature extraction will be done using KNN method in which Euclidean distance method is used to calculate the distance between two points in a coordinate system. And after retrieving the images, precision and recall values are determined.

ISSN: 2582 - 6379
Orange Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2020, Volume 1 Issue 2
www.ijisea.org

REFERENCES:

[1] Aasia Ali, Sanjay Sharma," Content based image retrieval using feature extraction with machine learning", International Conference on Intelligent Computing and Control Systems. ICICCS 2017.

[2] Gongde Guo, Hui Wang, David Bell, Yaxin Bi and Kieran Greer," KNN Model-Based Approach in Classification", supported by the European Commission project ICONS, project no. IST-2001-32429.

[3] Divya Srivastava, Rajesh Wadhvani and Manasin Gyanchandani," A Review: Color Feature Extraction Methods for Content Based Image Retrieval", in IJCEM International Journal of Computational Engineering & Management, Vol.18 Issue 3, May 2015.

[4] Yang Mingqiang, Kpalma kidiyo and Ronsin Joseph," A Survey of Shape Feature Extraction Techniques", July 15, 2008.

[5] Md. Baharul Islam, Krishanu Kundu and Arif Ahmed," in International Journal of Engineering and Technical Research (IJETR), Volume-2, Issue-4, April 2014.

[6] Zhe-Ming Lu, Hans Burkhardt," Fast mage retrieval based on equal-average equal-variance k-nearest neighbor search", on 18th International Conference on Pattern Recognition (ICPR'06)2, 853-853, 2006.

[7] https://sites.google.com/site/dctresearch/Home/content-based-image-retrieval.

[8] Henning Muller, Wolfgang Muller, David McG Squire, Stephane Marchand-Mallet and Thierry Pun," Performance Evaluation in Content Based Image Retrieval: Overview and Proposals", supported by the Swiss National Foundation for Scientific Research (grant no. 2000-052426.97).

www.ijisea.org