Abstract — Traditional methods for image retrieval used meta-data associated with images, commonly known as keywords. These methods empowered many World Wide Web (WWW) search engines and achieved reasonable amount of accuracy. A database shape, color, texture of content based image retrieval (CBIR) and classification algorithm is based on the K-means clustering is proposed in this paper. The algorithm is found the content based image capture with intra cluster and cluster will be calculated a similarity between the shape and texture of image grouping with minimum euclidean distance considerations in taking a picture of the relevant semantics of a database. The image classification using k-means clustering has been applied successfully in shape, colour, texture data base image. By using this algorithm we can classified any type of colour data base image in different field. Performance evaluation methods are now done with precision and recall for different databases. Response time to fine the most signatures of sixth is 85%.

Keywords—gradient edge detection; k-means; re-ranking; prewitt gradient magnitude; CBIR

I. INTRODUCTION.

Web based image re-ranking, is an effective to produce the right image search query by developing a system of hierarchy. Web based image re-ranking, is an effective to produce the right image search query by developing a system of hierarchy. The problem in this case a universal visual semantic signatures space to characterize diverse images from the web is difficult and inefficient. The intent of the classification process is to categorize all objects in a data base colour image into one of several classes, or "clusters". This categorized data is based on according to their some similar features present in an data base image. Image classification is a label process, in which image pixels are categorized into different classes in case of single image, but in case of data base image instead of pixels we take an image as an object and categorized into different classes. Image classification algorithms have been applied to a range of problems, including image segmentation and color quantization, change detection for land cover monitoring, and data mining among others. Image classification is perhaps the most important part of digital image analysis [19]. This problem was resolved by content based image retrieval (CBIR) up to an extent. CBIR uses contents of image, such as shape, color, texture or any other information that can be derived from image itself. Although there are many problems associated with CBIR method. Amongst them semantic gap with image features has received a lot of attention. Images are represented by low-level features and it is important to reduce semantic gap between high-level and low-level features of images to retrieve visual similar images and re-ranking of images. This study proposes Content-based image capture with intra cluster and cluster will be calculated a similarity between the shape and texture of image grouping with minimum euclidean distance considerations in taking a picture of the relevant semantics of a database.

II. RELATED WORK

Image re-ranking [6] [22] is a useful method for web-based image. The search based on only keywords queried by users. It is not efficient and results is imprecise. The web-based image search used by Bing and Google uses image re-ranking. In image re-ranking, users’ intention is captured by one-click on the query image. This helps in providing better search results to the users. In this reference, they used the method in which a query keyword is first used to retrieve of images based on the keyword. Image re-ranking framework automatically learns different from semantic spaces offline for different query keywords. To get semantic signatures for images, their visual features are projected into their related semantic spaces. Images are re-ranked by comparing their semantic signatures and the query keyword during the online stage. This procedure of selecting the query image and then re-ranking requires four steps : A query image is first categorized into one of several predefined intention categories, and a specific similarity measure are used inside each category to combine image features for re-ranking based on the query image. [4],[20-24]. Query keywords are expanded to capture user intention, through the visual content of the query image selected by the user and through image clustering. Image pool is enlarged to contain more relevant images. The query image is also expanded by using keyword expansion. All four of these steps are automatic with only one click in the first step without increasing user’s burden. This makes it possible...
for Internet scale image search by both textual and visual content with a very simple user interface. The explosive growth [3,5,10] and widespread accessibility of community contributed media content on the Internet have led to a surge of research activity in image search. Approaches that apply text search techniques for image search have achieved limited success as they entirely ignore visual content as a ranking signal. We propose an adaptive visual similarity to re-rank the text based search results. A query image is first categorized as one of several predefined intention categories, and a specific similarity measure has been used inside each of category to combine the image features for re-ranking based on the query image. The images are very rich in the content such as in colour, texture, and shape information which are presented in them. Retrieving images based on colour similarity is achieved by computing a colour histogram for each image that identifies the proportion of pixels within an image holding specific values (that humans express as colors) [11,12,16].

The relative level brightness of pairs of pixels are computed such in the degree of contrast, regularity, coarseness and directionality that may be estimated. The shape does not refer to the shape of an image, but to the shape of the particular region that is being sought out. Shape is the most important and most powerful feature used for image classification, indexing and retrievals. Shape information extracted using histogram for edge detection. The edge information in the image is obtained by using the Prewitt edge detection [16,17]. In shape, we will segmentation of two Prewitt gradient magnitude edge detection are two images which at each point contain the horizontal and vertical derivative approximation techniques. Shapes representations can be generally divided into two categories, they are boundary based and region-based, the Prewitt operator performs spatial gradient magnitude measurement in an image. Prewitt Edge Detector uses two convolution kernels, one is to detect changes in vertical contrast (hx) and the other is to detect horizontal contrast (hy). These kernels are designed to respond maximal to the edges, running vertically and horizontally, relative to the pixel grid, one kernel for each of the two perpendicular orientations. The kernels can be applied separately to the input image, to produce separate measurements of the gradient component in each orientation (call these Gx and Gy) [16,17]. This function makes the centroids as compact and separated as possible. When the clusters are of vary in sizes, k-means in order to reduce mean-squared error, divides large clusters into smaller ones even if they contain proper assigned data points. K-means method has the tendency to converge to local optima. To find a global solution, many methods are proposed. In [12,15], authors propose a fast global k-means technique based on geometrical information of the data points. Author in [13,14,15], relates the problem of assignment of points to centroid as black hole and star. The classic k-means method uses euclidean distance to measure the similarity between centroid and data points. Applying distance measure other than Euclidean may stop k-means from converging as its objective is to minimize the “squared” distance from centroid to data object. Spherical k-means, a variant of k-means employs cosine similarity method to compute the similarity between centroids and data points. All the centroids and data points are represented as vectors and L2 normalized. The cluster centroid is (L2) normalized summation of all the vectors in the cluster. Thus the centroid is also on the unit sphere. Lue [1] observes that Euclidian and cosine methods give same result when applied on unit vectors.

J. Mac Queen (1967) and L. Kaufman (1990), Partitional clustering aims at partitioning a group of data points into disjoint clusters optimizing a specific criterion [1,2,5]. When the number of data points is large, a brute force enumeration of all possible combinations would be computationally expensive. Instead, heuristic methods are applied to find the optimal partitioning. A cluster is a collection of data points that are similar to one another within the same cluster and dissimilar to data points in other clusters [3,4,6,7,8]. Clustering is a method of unsupervised classification, where data points are grouped into clusters based on their similarity. The goal of a clustering algorithm is to maximize the intra-cluster similarity and minimize the inter-cluster similarity. Clustering algorithms can be broadly classified into five types: Partitional clustering, hierarchical clustering, density-based clustering, grid-based clustering and model-based clustering [5]. Partitional and hierarchical clustering are the most widely used forms of clustering. In partition clustering, the set of n data points are partitioned into k non-empty clusters, where k ≤ n . In the case of hierarchical clustering, the data points are organized into a hierarchical structure, resulting in a binary tree or dendogram [9,18,19].

### III. METHODOLOGY

Data mining for finding and describing structural patterns in data using python. It is applied in a wide range of domains and its techniques have become fundamental, we will see how to:

- Import and visualize data, classify and cluster data, discover relationships in the data using regression and correlation measures, reduce the dimensionality of the data in order to compress and visualize the information it brings, analyze structured data. In this section, we will describe our proposed method which is an efficient image re-ranking system. The system architecture contains offline part and a digital signature is obtained by the process of encode, decode web-based and use a paint pen. The algorithm used to find data points based on a contribution and similarity search algorithm to optimize intra-cluster and cluster signature in the sixth grade ie, GX, Gy, Gt, Gx ', Gt', Gt '. Images of Gx, Gy, Gt are convolution with the intensity values of 0.05 to obtain the data images Gx ', Gt', Gt '. The sixth classes will be divided into 6 hierarchies. The input process of learning is the learning process with the input images stored in a database with three classes of data, results of pre-process by the method of gradient prewitt which aims to increase (i) feature extraction accuracy of the k-means algorithm (ii) an increase in the uptake of better accuracy, distance and similarity retrieval quick. Content-based image capture with intra cluster and cluster will be calculated a similarity between the shape and texture of image grouping.
with minimum euclidean distance considerations in taking a picture of the relevant semantics of a database.

3.1 Learning Process

Learning process is the input of a collection of learning software image that has been known as the class label. The stages of the process as follows:

A. Application paint pen
   The digital signature is create web based (see Fig 2). The application paint pen create a signature digital.

B. Learning input learned there step the process of digital data signature, consist of:
   1) The digital signature is obtained by the process of encode, decode web-based and use a paint pen (see Fig 3).
   2) Digitalize data in a data store in Portable Network Graphic (PNG) format with image size is 400 x 400 results.
   3) Prewitt gradient magnitude uses two convolution kernels, one is to detect changes in vertical contrast (hx) and the other is to detect horizontal contrast (hy). The data is stored in the directory, the signatures and image results prewitt gradient magnitude.
   4) Save the signatures data with a file name according to the name you have.

C. Classification

The database of signatures are stored in the data base in Figure 4. The data of signatures will be classified with k-mean clustering of the following steps: (1) where K points in space represented by any class i.e. Gx, Gy, Gt, Gx', Gt', Gt'. The sixth classes will be divided into 6 hierarchies. specific value points represent initial group centroids per class, and (2) defines each object to the group that has the closest centroid, (3) when all is set, recalculated the position of the K centroids, (4) determine the distance of each object to the centroid and (5) repeat step 2 and 3 until the centroids no longer move and a cluster of a data, we can see in Figure 5. Because the same data is displayed for ID and a signature that has the same pattern. The pattern shown is the minimum, maximum and average the digital signature of each individual.

IV. RESULTS AND DISCUSSION

For the proposed method, experiments for testing in an environment Windows XP and data mining for finding and describing structural patterns in data using Python. Experiments conducted by various databases for different sizes. Signature image uploads from sub-directory, the sixth classes will be divided into 6 hierarchies and consisting of 6 classes. The dataset is stored in the comma separated values (CSV) format. It is convenient to parse the CSV file and store the information that it contains using a more appropriate data
structure. The dataset has 5 rows. The first 4 rows contain the values of the features while the last row represents the class of the samples.

![Fig.7. A Result Re-rangking a Signatures](image)

IV CONCLUSIONS

In this paper, we have the algorithm how K is combined with a combination of K means algorithm with prewitt filter for case learning for a universal visual semantic signatures space. The images can be stored on user data based system. Experiments with different database sizes show that the number of similarity comparisons are learning for a universal visual semantic signatures space.

The similarity between images is calculated based on local color properties. Images in the database are clustered into groups having similar color contents. This grouping enables searching only images that are relevant to the query image, the integrated clustering algorithm for image classification is tested with a different image that image signatures with a different gradient.

The previous work the image rank with the help of textual based and content based. In these project the image re-ranking on offline data base. The images can be stored on user data based system. Experiments with different database sizes show that the number of similarity comparisons are learning for a universal visual semantic signatures space.

REFERENCES


