Image to Image Search using K-means Clustering

Aayush Kumar Singh
Maharaja Agrasen Institute of
Technology, GGSIPU
Plot No 1 Rohini, Sector 22, PSP
Area, Delhi, 110086

Abhishek Kumar
Maharaja Agrasen Institute of
Technology, GGSIPU
Plot No 1 Rohini, Sector 22, PSP
Area, Delhi, 110086

Kuldeep
Maharaja Agrasen Institute of
Technology, GGSIPU Plot No 1
Rohini, Sector 22, PSP Area,
Delhi, 110086

ABSTRACT
With the increase in the amount of information on the World Wide Web, it has become a difficult task of finding relevant information. More often, textual based search engines were used but there were very few search engines using which similar images can be searched all over the internet with an image as input. This method of finding information is called Content-Based Image Retrieval (CBIR). Commonly used engines for CBIR like Google Images, Yahoo! Images are based on textual annotation of Images. The images that form the result have to be previously tagged in order to appear as result. In this research, K-means clustering for segmentation of images into clusters have been studied and applied, thereby finding similar images without the need of labelling or tagging images in the database.

General Terms
CBIR, Reverse Image Search, Artificial Intelligence

Keywords
Image, Image classification, K-means clustering, Image search

1. INTRODUCTION
Image to Image search is a method of searching images having similar features i.e., shape, texture or colour from a database containing images in large numbers (may even span across every image on the internet) given an image as input. The popular search engines used for this purpose employ a mechanism of manually annotating the images with tags or keywords which is used to compare with the input images to find similar images. This method is successful in doing what is asked but its performance decreases as the size of the database increases and will not be feasible when a large amount of unlabelled images are present in the database.

With the increase in the amount of data that is uploaded on the internet, it will be resource consuming to consider tagging of each image with keywords that can be for the image search process. Also, taking into account images that are unclear to the viewer or cannot be identified may be falsely tagged. To make the process more efficient and accommodating towards unlabelled data in the database, the K-means clustering algorithm is used. K-means is an unsupervised learning algorithm which is used to find clusters in data without knowing which data points belong to a cluster; it has to uncover similarities for itself, where K representing the number of clusters or number of cluster centres. Feature similarity is used as a measure to classify the data into the cluster, with similar data items belonging to a cluster and dissimilar belonging to others. The colour of pixels is considered as a feature for classification of images. Remaining sections of the papers are sectioned as follows: (ii) Literature Background. (iii) Proposed Methodology. (iii) Implementation. (iv) Searching Image. (v) Summary and Conclusions. (vi) Future work.

2. LITERATURE BACKGROUND
This concept of retrieval of images based on content was first introduced by Kato in 1992. In the past few years, it has been implemented for retrieval of images from various types of database such as a human face, vehicle, animal and flower. The techniques used were based on match point, colour histogram and Eigen values [1]. This was later improved to include a fast and efficient application with help of modified Support Vector Machines (SVM) [2]. Support Vector Machines were used as a way for clustering of images.

In 2015, to increase the performance of the system various Relevance Feedback Techniques were selected and tested [3]. The idea of Relevance Feedback is to take results that are initially returned from a given query image, to gather feedback and perform a new query based on the obtained information. Bayesian Relevance Feedback, k-Nearest Neighbour and Query Point Movement were a few of the selected techniques. Existing Shape descriptors mainly contours based and region based both of which can be further divided into global shape descriptor versus local shape descriptors were reviewed as a feature for the purpose of Image Retrieval [4].

In 2016, to reduce the semantic gap between High-Level Features such as text, visual features and Low-Level Features like colour, shape and texture five methods were studied. Two of which were the use of machine learning tools to associate LL features with query concepts and use of object ontology to define high-level concepts [5]. Use of object ontology works in a way by assigning a region attribute such as an image of scenery may contain grass in the lower region. A descriptor is assigned to it specifying the grass's position, colour, shape and size. Search-Based Face Annotation in which the images of people's faces were correctly tagged with respective names were tested used of clustering based algorithm mainly K-means clustering [6].

In 2017, a new technique based on the prediction-error system was presented. The proposed method makes use of a new clustering scheme employing wavelet based counter transform and fuzzy c-means clustering for improving the performance of the system. A multilayer neural network predictor has been employed for predicting the pixel values using neighbourhood pixels. Since only the error values are stored in the database instead of actual pixels, the proposed system has less number of variables and requires less storage space [7]. To improve the retrieval performance of the system a proposed method combined the colour and texture feature to constitute a single feature vector [8].

3. PROPOSED METHOD

Colour, is considered one the most efficient feature for use in an image retrieval system. Due to its additive nature, the basic colour in Red-Green-Blue space can be combined to form all the colours in the spectrum. Therefore, colour was considered as a feature to find similar images based on its distribution.
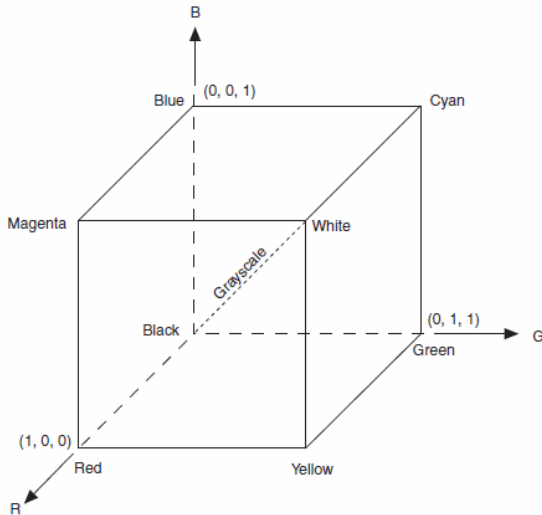

Figure 1: RGB Colour space.

The proposed system works, by extracting the features of an image and clustering them into similar clusters by K-means clustering. K-means was selected due to its easiness to implement, producing more tighter clusters and being computationally faster than hierarchical with images that have a large number of variables i.e., pixels. An image with dimensions of 64 by 64 will have 4096 variables and gets increased by 3-fold when all 3 colours of RGB space are considered. (see figure 1)

In order to reduce the number of variables and effectively increase computation speed, Haar wavelets are used. Haar wavelet compression is an efficient way to perform both lossless and lossy compression. It relies on averaging and differencing values in an image matrix to produce a matrix which is sparse or nearly sparse i.e., a matrix having a large portion of its entries as zeros. It is implemented to image once it is extracted into three component matrix. The working and implementation of the above are discussed in the next section.

4. IMPLEMENTATION

4.1 DATASET USED

This method was tested on two datasets. One of which was the Corel Database. Corel Database for Content-based image retrieval is divided manually into 10,800 images from the Corel Photo Gallery into 80 concept groups, e.g., autumn, cloud, waterfall, iceberg with each group having more than 100 images. Each group was divided in the ratio of 80:20 for training and testing purposes.

The Second dataset was created by us having 1,500 unlabelled images of various clothing items including dresses, bags, shoes, etc. with a distribution as shown below. This dataset was created in order to test that the proposed model was able to work on different kinds of images. But before using them in the model for training, it was passed through various types of pre-processing for results to have minimum error.
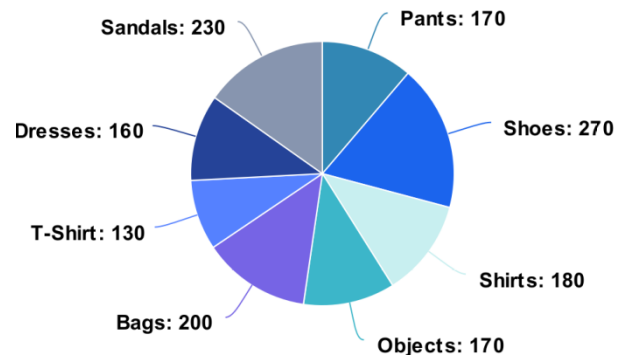

Figure 2: Distribution of different images in the dataset

Prior to passing the images through K-means algorithm, pre-processing of images is done.

4.2 PRE-PROCESSING OF IMAGES

In this stage the raw image is passed through a number of preliminary processing in order to manipulate and hence enhance their respective characteristics .i.e., contrast, brightness, slant, width, height.

4.2.1 Noise Reduction:

Noise reduction is the processing of removing unwanted information or noise from a signal. Images, taken both from digital and film cameras are susceptible to noise. Before passing it to the k-means algorithm, the noises present in the image have to be removed. One of the methods used for this purpose is the use of Linear Smoothing filters.

These filters remove the noise by convolving the original image with masks that represents a pass filter. As a result of this, the image tends to smooth out.

| (a) | (b) |

Figure 3: (a) Image chosen from "objects" with visible noise. (b) Image after noise removal.

4.2.2 Contrast Enhancement:

Contrast is defined as difference in visual properties that make an object differentiate from other objects and background. Contrast enhancement and quality improvement processes are performed to better highlight the information contained in the image which may be not visible.

Contrast enhancement is done by linear contrast stretching. By trial and error, it was found that Minimum-Maximum Linear Contrast Stretching works best for this purpose. In this method, the original maximum and minimum values of brightness are assigned newly specified set of values that utilize the full range of available brightness values. For example, if an image has a minimum brightness of 30 and a maximum of 100, then contrast can be enhanced by stretching

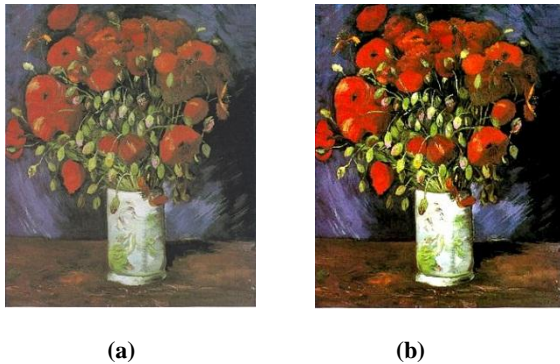the minimum and maximum values to zero and 255 respectively.

(a) **(b)**

Figure 4: (a) Image before contrast enhancement (b) Image after contrast enhancement

4.2.3 *Resizing of Images:*

Images were resized to a default size of 64x64 without hampering the quality in order to decrease the number of variables, as working with a large number of variables require a large amount of memory and computational power. It also decreases size taken by the database and increases efficiency as the smaller the dimensions of the image, less is the resources consumed by it, in being processed.

Progressive scaling is used in which an image is incrementally scaled in smaller steps until the final dimensions are reached. The results produced are acceptable in most cases with very good performance while the only downside is, if scaled less than half then it produces mediocre results since it is only scaled once.
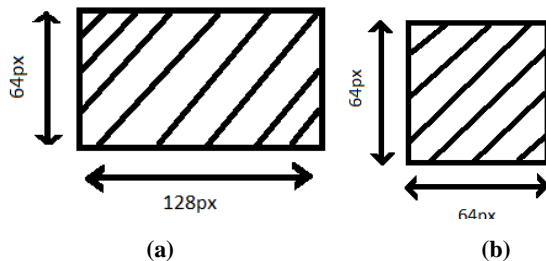

(a) **(b)**

Figure 5: (a) Image of dimensions 128px by 64px (b) Image scaled to 64px by 64px dimensions

4.3 FEATURE EXTRACTION AND K-MEANS

Pre-processed images are then converted to image matrix having the colour intensity of each pixel as values of rows and columns. For e.g., if the second pixel from 3rd row has colour intensity values of 208, then it is normalized to a value between 0 and 1, and stored in m[3][2], where m is image matrix.

Haar wavelet transform is then used to convert the image matrix into a sparse matrix. As an example, consider a 1D matrix, (88 88 89 90 92 94 96 97). The transformation occurs in three steps, the first grouping of all columns in pairs: ([88,88], [89,90], [92,94], [96,97]) and replacing first four columns by average of these pairs and last four with ½ of difference. In the second step, this process is repeated with using the resultant matrix taking an average of the first two columns and ½ of the difference of next pair ignoring the last four columns. The last step is to replace the first column with an average of the first two numbers and the second column

with ½ of the difference of the first two numbers. The resultant matrix obtained (91.75 -3 -0.75 -1.75 0 -0.5 -1 -0.5) has one less number of variable than the original matrix. This process is similarly repeated on the image for all rows and columns of the image matrix.

The image matrixes obtained as a result are passed through K-means clustering algorithm for clustering of an image into k number of clusters. The k-means clustering works follows:

1. Initially, k pixels of an image are chosen. These pixels represent centroids for an image.

2. Each pixel is assigned, starting from top-left of an image to one of the k-clusters that has the closest centroid.

3. When all pixels have been assigned, recalculate the positions of the k centroids.

4. Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups.

The mean values of these k clusters are used as an index and are stored in the database along with their corresponding images as shown below.
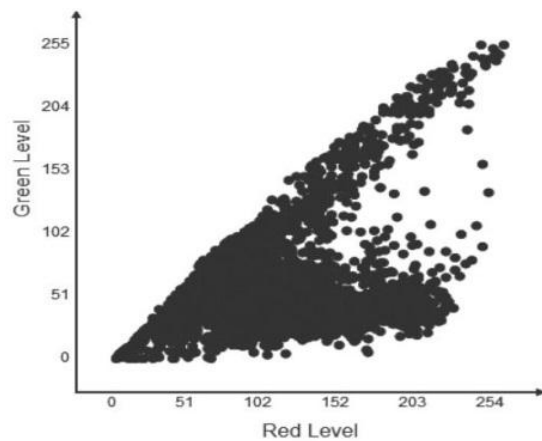

Figure 6: 2-Dimensional Plot of colour values of pixels in Green Red space.

4.4 SIMILARITY MEASURE

Classified images are similar to one another if their centroids have a difference in values up to a threshold limit. If their centroids are greater than their threshold value then the image is considered to be different. This is based on the theory that similar images have equally distributed colour composition or slightly varying composition. E.g., images of a body's CT scan will always have densities of black, grey and white.

5. SEARCHING IMAGES

Images that are input to the system are first passed through the pre-processing stage where the various manipulations were applied to it to decrease the number of variables and improve the quality, for objects in images can be clearly identified with accurate results.

Features of an image were extracted and K-means algorithm was then applied to it in order to iteratively find the k cluster centres of its pixels. The values of k were set to 2, 4, and 6 in order to observe the changes in clusters made by the system. Numbers of iterations for algorithm tested are 50, 100 and 200. These cluster centres are used to search the database's

column consisting of the stored clusters centres and their respective images. This image can be stored in databases with its cluster centres for it to further appear in search results of different queries.

The search results consist of images which best match the colour composition of query image. Here, the images are shown with their similarity measures i.e., how much the query input image matched with the images in the database. The results can be filtered out based on the value similarity measure. E.g., if the filter is set to value of 0.5, images with similarity measures of greater than 0.5 will be obtained as result.
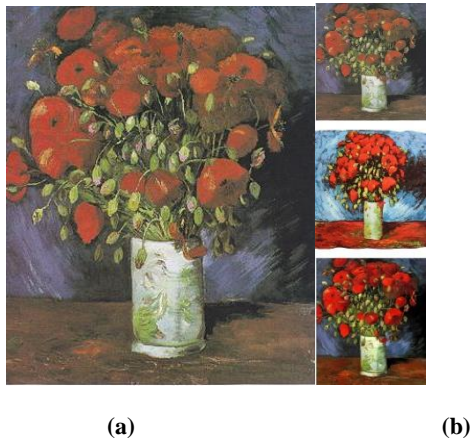

(a)   **(b)**

Figure 7: (a) Input query Image (b) Examples of output corresponding to the given query input image from the dataset used.

6. SUMMARY AND CONCLUSIONS

It was have shown that using k-means clustering images can be segmented into clusters and thereby used to find similar or exact images from a collection of unlabelled or unsupervised images with good results. The results were studied with the different value of k and for different number of iterations. The accuracy of the system for the selected value of k equals 4 and the number of iteration equals 200 works with the system well to give positive results. Various pre-processing work were also have studied that can be used to enhance the quality of images and reduce noise and concluded the above three processes are best suited for the project.

7. FUTURE WORK

Although the current system proposed provides satisfactory results, it can be further improved to enhance speed and accuracy. It can also be trained with various scanned images of the body's parts to find the presence of tumours. Also, instead of using k-means different types of unsupervised clustering algorithm like spectral clustering, expectation-maximization etc. can be implemented.

8. ACKNOWLEDGEMENET

9. REFERENCES

[1] Mohini. P. Sardey, G.K Kharate, "A Comparative Analysis of Retrieval Techniques in Content Based Image Retrieval" ArXiv, 2015.

[2] Deepu Rani, Monica Goyal, "A Research Paper on Content Based Image Retrieval System using Improved SVM Technique" International Journal Of Engineering And Computer Science, Volume 3 Issue 12, Dec 2014

[3] Mowloud Mosbah and Bachir Boucheham "Selection of Relevance Feedback Technique in Context of CBIR" The International Arab Conference on Information Technology (ACIT' 2015)

[4] Pushpendra Singh, V.K. Gupta and P.N. Hrisheekesha, "A Review on Shape based Descriptors for Image Retrieval" International Journal of Computer Application, Volume 125-No.10, Sept 2015

[5] Nikita Upadhyaya and Manish Dixit "A Review: Relating Low Level Features to High Level Semantics in CBIR", International Journal Of Signal Processing, Image Processing and Pattern Recognition, Vol.9, No.3, 2016

[6] Suchita Barkund, Dr. Sulochana Sonakamble, "Search-Based Face Annotation with CBI and Clustering-Based Algorithm", International Research Journal of Engineering and Technology, Vol 3 Issue 06, June 2016.

[7] A. Mohamed Uvaze Ahamed, C. Eswaran and R. Kannan, "CBIR System Based on Prediction Errors", Journal of Information Science and Engineering 33, 347-365 (2017).

[8] Hany F. Atlam, Gamal Attiya, Nawal El-Fishawy, "Integration of color and Texture Features in CBIR", International Journal of Computer Applications (0975 - 8887), Volume 164-No 3, April 2017