IMAGE QUERY BASED SEARCH ENGINE USING CONTENT AWARE IMAGE RETRIEVAL

[1]*Annapurani.K,[2]Ayush Surana, [3]Tazeen Ajmal

ABSTRACT--*Successful Content-Based Image Retrieval is a challenge even today. The present system for image retrieval is quite inefficient and expensive. This paper proposes a system that deals with searching and localizing all occurrences of an image query. The image is constituted using a set of descriptors which makes recognition successful even when there is a slight difference in illumination or a change in point of view. A dictionary of visual words is created through raw descriptor matching by trying to utilize the same concept as text retrieval approach from a document. This content-aware Image retrieval system forms full frame queries from a region, matches it with the image query. This process uses the Scale Invariant Feature Transform (SIFT), a feature detection algorithm to outline features in an image. The query image is recognized in the database by individually comparing each feature and grouping the similar images by calculating Euclidean distance of their respective feature vectors. This process is trained on a pre-determined (nearly 24,000) number of image frames of a group of episodes of the F.R.I.E.N.D.S T.V. show by extracting 24 frames per second and dropping the common frames.*

Keywords-- *image, query, search engine, content, retrieval*

I. INTRODUCTION

In the present era availability of data is not an issue. However, using data in the most efficient way draws our focus. Nobody wants to wait to get the desired result. Enterprises are working hard to make systems that could retrieve the perfect match for an image query in search engines using the least amount of time. Having an ideal image extraction system is crucial since images have an unparalleled impact on multimedia applications. Formerly, image retrieval system was termed as text-based image retrieval. However, it had many drawbacks. Textural description does not capture visual content and images are beyond description of words. New retrieval techniques were proposed to solve these issues. Content Based Image Retrieval (CBIR) (Content Base Image Retrieval) and Scale Invariant Feature Transformation (SIFT) (Scale Invariant Feature Transform) have become household names for image retrieval.

Content based image retrieval (CBIR) tries to discover patterns in images. It helps the image data management system as it is faster than the traditional methods. The retrieved images have various features like color, shape and

[1]**Department of CSE, SRM Institute of Science and Technology, Kattankulathur, Chennai, India, annapook@srmist.edu.in.*
[2]*Department of CSE, SRM Institute of Science and Technology, Kattankulathur, Chennai, India.*
[3]*Department of CSE, SRM Institute of Science and Technology, Kattankulathur, Chennai, India.*

Received: 22 Aug 2019 | Revised: 13 Sep 2019 | Accepted: 15 Nov 2019

dimensions which are automatically extracted since the CBIR is trained that way. The feature description is used to detect objects in the other images present in the database. This is also called image matching.

To obtain higher accuracy in the result, extracting and recording the relevant features from the image is vital. One issue that is pervasive in the domain of computer vision is feature matching with irrelevant images which are bringing the accuracy of the system down. Researchers are being conducted to make feature matching efficient as it forms the backbone of the retrieval system. Features and interesting points are important information that can be extracted from an image to provide a feature description of an object in the image. This description can then be used to locate the object in another image which is usually called as image matching. Image matching is a fundamental aspect of many problems in computer vision, including object or scene recognition, solving for 3D structure from multiple images, and motion tracking. In order to achieve a good matching result, there are many considerations when extracting features and record them from an image.

It attempts to find designs in pictures. It helps the picture information to executive’s framework as it is quicker than the customary strategies. The recovered pictures have different highlights like shading, shape and measurements which are consequently separated since the CBIR is prepared that way. The component portrayal is utilized to identify protests in different pictures present in the database. This is additionally called picture coordinating.

To get a better degree of precision in the outcome, separating and recording the pertinent highlights from the picture is indispensable. One issue that is inescapable in the space of PC vision is including coordinates with unessential pictures which are bringing the precision of the framework down. Inquires about are being led to make highlight coordinating productive as it shapes the foundation of the recovery framework. Highlights and intriguing focuses are significant data that can be extricated from a picture to give an element portrayal of an item in the picture. This portrayal would then be able to be utilized to find the article in another picture which is typically called as picture coordinating. Picture coordinating is a basic part of numerous issues in PC vision, including article or scene acknowledgment, fathoming for 3D structure from different pictures, and movement following. So as to accomplish a decent coordinating outcome, there are numerous contemplations while extricating highlights and record them from apicture.

Matching features across different images is a common problem in computer vision. When the images are similar in nature, which means they have same scales and same orientations, simple corner detector can be used to extract features from both images. However, when the images are different both in scales and rotations, simple corner detector cannot solve the problem. Some corner detectors like Harris detector is rotation invariant, which means, we still can find the same corner even if the image is rotated. It is obvious because corners remain corners in rotated image. However, Harris is not scale invariant because a corner may not be a corner if the image is scaled.

In the present time accessibility of information isn’t an issue. No one needs to hold on to get the ideal outcome. Undertakings are endeavoring to make frameworks that could recover the ideal counterpart for a picture question in web search tools utilizing the least sum of time.

Having a perfect picture extraction framework is essential since pictures unparalleled affect sight and sound applications. Earlier, picture recovery framework was named as content based picture recovery. In any case, it had numerous disadvantages. Textural portrayal does not catch visual substance and pictures are past depiction of words. New recovery strategies were proposed to tackle these issues. CBIR (Content Base Image Retrieval) and SIFT (Scale Invariant Feature Transform) have progressed toward becoming easily recognized names for picture recovery.

In simpler words, our proposed search engine will take a query image, we will then select a region of interest in the query image, run it through the database and all the images that match with our region of interest will be produced as output.

II. LITERATURE SURVEY

A lot of research is being going on to make an impeccable image retrieval system. Over the years a fair amount of traditional practices has been rendered inefficient and innovations have been welcomed. Below is a list of the relevant related works:

2.1. Text Retrieval Approach Image Retrieval

Josef Sivic and Andrew Zisserman [1], proposed an object retrieval system which looks for and determines all the instances of a query image in a video. The query image is a collection of descriptors/features which are illumination and viewpoint invariant so that recognition can proceed successfully regardless of the changes in illumination, view- point and partial occlusion. The matching features are determined by classical vector quantization technique clubbed with indexing and ranking of the files. The outcome is a prompt retrieval of images and a ranked list of images, similar to the output produced by Google.

Lina Al-Quraan, Sawsan Nusir and Belal Abuata [6], studied the performance of three image search engines for Arabic text queries. They also evaluated whether the query in language has any effect on the content based images retrieved. For the first phase, they used ten Arabic text queries and the retrieved images were tested for relevance and the prediction ratio was calculated for each query. For the second phase, image features were analyzed and calculated using the Euclidian distance. First phase results showed that Google has the best retrieval efficiency while the second phase results showed that the image content was not similar for the retrieved images for a specific query.

2.2. Object Retrieval Approach

Hailiang Li, Yongqian Huang and Zhijun Zhang [2], proposed an improvised Region Convolutional Neural Network (R-CNN) that will return the images matching with the query object even in cases where the setting of the image/scene has been modified. They concluded that the ability of Region of Interest (ROI) pooling to extricate features is enhanced by concatenating the feature maps of various convolution layers.

Benjamin Bustos [3], described methods for extracting descriptors from 3D objects automatically. They applied the content aware searching and optimization to determine the similarity between two 3D images. Although the use of 2-D mappings decreased the probability of determining the equality and the system is only able to provide two 3D models that are similar to each other.

Philbin, J., Chum, O., Isard, M., Sivic, J., Zisserman, A. [4], proposed a comprehensive object retrieval system. This system takes an input of a query object by only selecting a part of it, and the system outputs the set of matching images from the database ranked on the basis of fraction of the query image they contain. This was a revolutionary step towards a large scale image retrieval system in proposed invert picture look based recovery framework. It gives the CNN framework a question picture on which the inquiry will be based. Switch picture hunt can fill two needs. It tends to be utilized to look applicable information identified with the inquiry picture just as gathering appropriate pictures identified with question picture. They had the capacity to execute their calculation continuously for up to 100000 pictures.

2.3. Image Retrieval Techniques

Wichian Premchaiswadi, Anchua Tungkatsathan [7], presented a joint queries based image retrieval system. They went on to introduce feature extraction based on color called Auto Color Correlogram and Correlation (ACCC). This helped them extract low level features in images.

Dr. T. Santha, M. Abhayadev [8], explained the image retrieval and Content Based Image Retrieval (CBIR) systems, and their working levels. It also discussed the challenges faced by CBIR. They presented a set of design issues to be taken into consideration before proposing architecture of image retrieving system.

Pushpa M. Chutel, Apeksha Sakhare [9], proposed reverse image search based retrieval system. It provides the CBIR system with a query image on which the search will be based. Reverse image search can serve two purposes. It can be used to search relevant data related to the query image as well as group pertinent images related to query image. They were able to implement their algorithm in real time for up to 100000 images.

A. Araujo, D. Chen, S. Tsai, H. Chen, R. Angst and B. Girod [10], discussed the importance of variability of visual search. Various applications which require indexing of video databases are determined by their visual contents. Hence it is vital to diverse the scope

B. of search. They presented new solutions to reduce storage requirements and improve video search quality.

III. DESIGN OF QUERY SEARCH ENGINE

The architecture of the proposed Image Query Based Retrieval System has been illustrated via the block diagram in Fig 1. The step by step process for retrieval of image has been discussed in the modules given below.

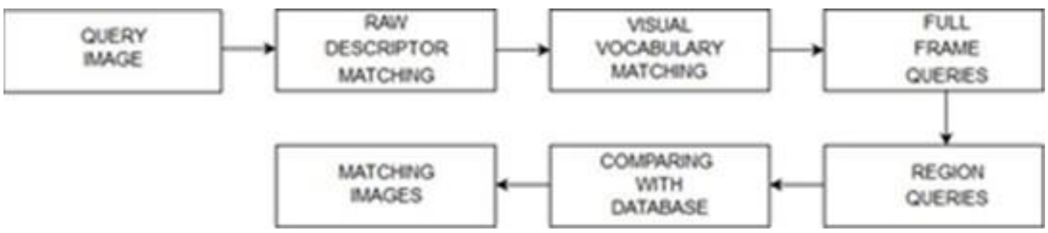

Figure 1: Architecture of the proposed Image query based retrieval system

2.4. Query Image

Query Image could be a single image from the database or a part of an image that we feed into our program. This image is supposed to be that part of our image that we need to search for. Now this image can be a complete image or a part of an image. The database that we are using for this purpose is that of all the frames obtained from

Received: 22 Aug 2019 | Revised: 13 Sep 2019 | Accepted: 15 Nov 2019

International Journal of Psychosocial Rehabilitation, Vol. 23, Issue 04, 2019
ISSN: 1475-7192

a particular episode of the very popular F.R.I.E.N.D.S T.V Series. The database has 6612 images that have been manufactured by extracting 20 fps from the videos. The similar frames from the video have been discarded.

It is the query image in which the region of interest is marked and is run through the database to find out the images in the database that match with the input query image.

2.5. Raw Descriptor Matching

The first step towards solving computer vision task is to portray the image as a group of extracted features. Our eye can decipher all the details from the raw image, however that is not true for computer algorithms. Image features are classified as global and local features. When the image is represented by a single multi-dimensional feature vector, it is called a global feature representation since it describes all the information about the image as a whole. In local feature representation the image is represented based on notable regions.

The global features measure various characteristics of the image like shape, color, dimension, etc. A single vector from each image is extracted and the two images are compared by comparing their feature vectors. Thus, image is extracted by a combination of global and local features which is collectively termed as raw descriptor matching as shown in Fig 2

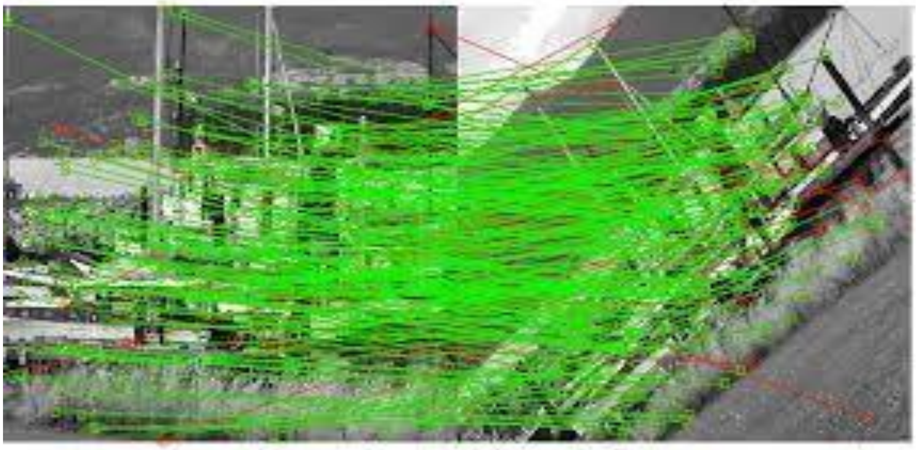

Figure 2: Raw Descriptor Matching

2.6. Visual Vocabulary Matching

Creation of a Visual Vocabulary with the basis of the similarity matching of the different aspects of the things. This vocabulary can be created from the set of different images and then using the text retrieval approach for the purpose of succinctly matching all the details with the similarity of these features. They are best ways to understand the different visual vocabularies and analyze them for the purpose of Image Extraction and Searching.

The region of interest chose on the query image is run through the database. The database contains 6612 pictures. The descriptors set apart on the picture based on varying level of vitality are utilized to form vocabulary. The separation between the descriptors is discovered by ascertaining k-means. The descriptors are bunched based on data given by k-means work. The query image is then run through database and the similitude of the region of interest is found out.

This is the penultimate module of the query based search engine. The yield of this module is fed to full casing questions that produce the yield. The false positives are disposed of, all things considered, here. Be that as it may, no machine is impeccable and some false positives still appear in the outcomes. This is discussed in detail in the Results section.

2.7. Full Frame Queries

Full frame is basically the act of capturing an image by fixing the width and height at its maximum value. Full frame query is therefore focusing on the image as a whole. The search query image is analyzed by the search engine as a whole and the features are defined for the full frame query image. The full frame queries are separated into a lot of smaller queries. This progression helps in breaking down the complicated subtleties of the input image and aides in decision out the insignificant pictures from the database. The matching areas are characterized for pictures when the full edge inquiry matches for a picture in the database. It is a significant advance towards decreasing the quantity of calculations for a query image.

2.8. Region Queries

The full frame query is broken down into a set of region queries. This step helps in analyzing the intricate details of the search image and helps in ruling out the irrelevant images from the database. The region queries are defined for images when the full frame query matches for an image in the database. It is an important step towards reducing the number of computations for a query image.

Region query is fundamentally the demonstration of catching an image by fixing the width and length at its most extreme ends. Region queries are therefore concentrating on the picture all in all. The query image picture is examined by the web index all in all and the highlights are characterized for the full frame query image. The full edge inquiries are isolated into a lot of smaller queries. This movement helps in separating the confused nuances of the information. It is a noteworthy development towards diminishing the amount of calculations for a query image.

2.9. Comparing With Database

This is the final step where we sort the relevant images and produce the output. Scale Invariant Feature Transformation (SIFT) algorithm is applied in this step to find the relevant images. SIFT computes the best orientation for each key-point region. It uses local image gradients at selected scale and rotation to describe each key-point. SIFT allows us to take two pictures of similar content, find key-points which are invariant to scale and rotation, remove outliers and transforms those images. Using this algorithm images in the database containing different views for a selected region of interest of query image are found.

IV. IMPLEMENTATION

The project was implemented using Matlab R2018a. A total of 8 Matlab Files have been implemented to get the desired results. The following sections contain the code and the output of the project.

2.10. Get Query Descriptors

This is the first module of implementation. The query image is fed as input. A dialog box is prompted on the screen where region of interest from the query image is to be selected. Once the region of interest is selected, the co-ordinates of the region of interest is recorded and is run through the database. On compiling this file the dialog box containing the query image appears as depicted by fig .3. The user is asked to select a region of interest having 4 co-ordinates. On selecting the region of interest itis displayed on screen as depicted by fig .4.

At the point when the region of interest is chosen in the query image, the co- ordinates of the region of interest is recorded. This is sustained to the following module. According to the vitality of the picture diverse size of descriptors are set apart on the picture. Varying level of vitality clusters the descriptors. It additionally gives various false positives, which is rectified in the following module by picturing the vocabulary that is, introducing pictures in the diminishing order of similarity index.

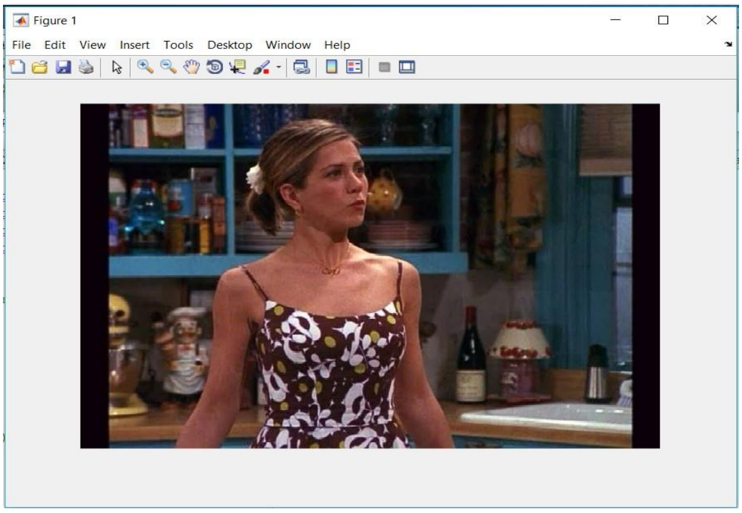

Figure 3: Query image

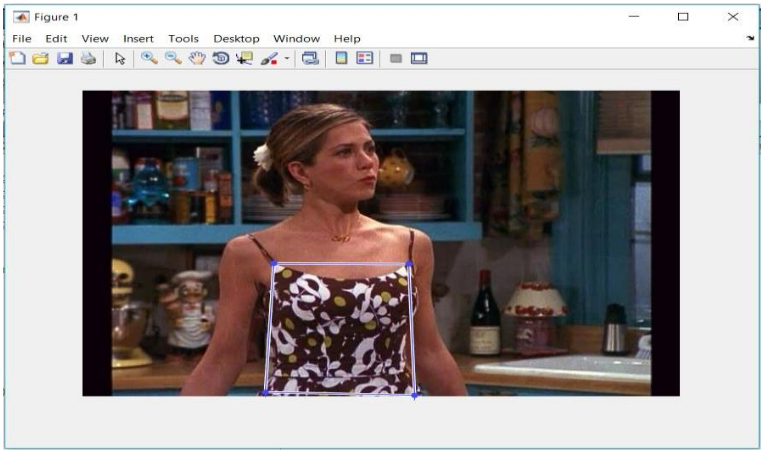

Figure. 4: Query image with region of interest selected

2.11. Raw Descriptor Matches

When the region of interest is selected in the query image, the co-ordinates of the region of interest are recorded. This is fed to the next module. As per the energy of the image different size of descriptors are marked on the image. Varying degree of energy clusters the descriptors. It also gives a number of false positives, which is corrected in the next module by visualizing the vocabulary that is, presenting images in the decreasing order of similarity index.

Fig .5. shows the query image with the selected region of interest marked with a blue box. The output of this module is shown in fig .6. where the output image with descriptors is displayed. As per the energy of the image different size of descriptors are marked on the image. Varying degree of energy clusters the descriptors. It also gives a number of false positives, which is corrected in the next module by visualizing the vocabulary that is, presenting images in the decreasing order of similarity index.

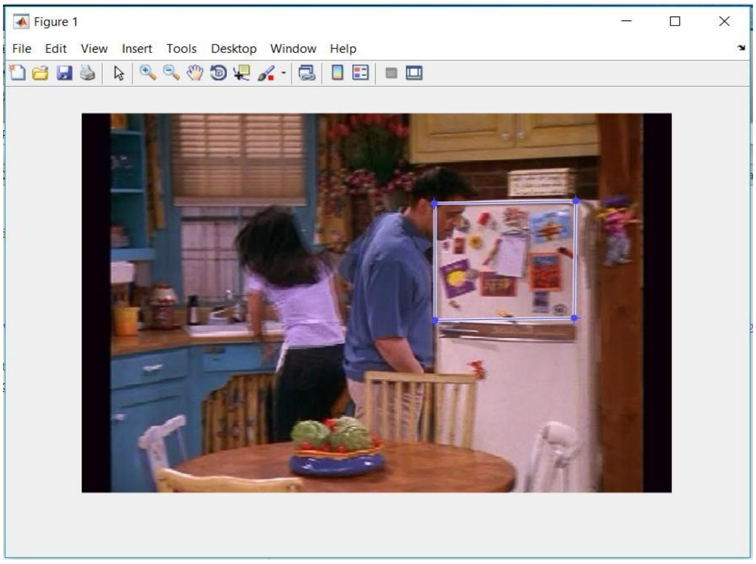

Figure 5: Query image with selected region of interest

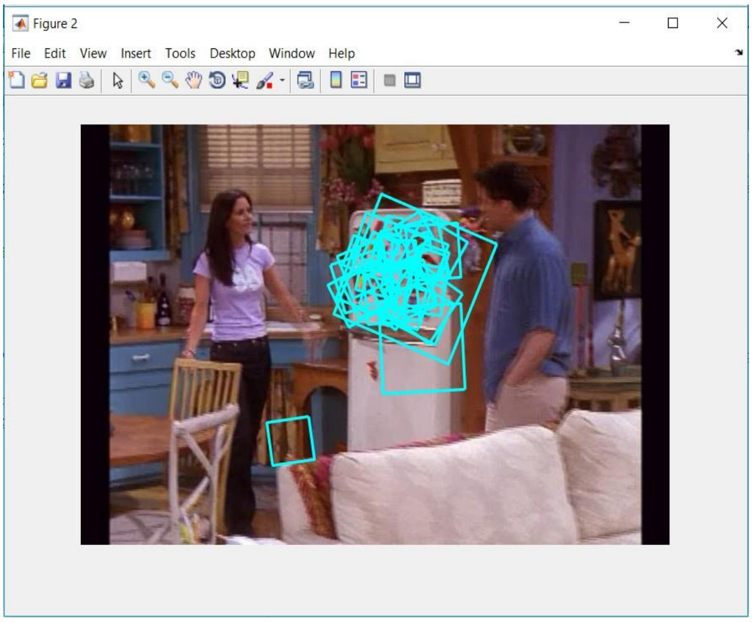

Figure 6: Output image with descriptors

2.12. Visualize Vocabulary

The region of interest selected on the query image is run through the database. The database contains 6612 images. The descriptors marked on the image on the basis of varying degree of energy are used to visualize vocabulary. The distance between the descriptors is found out by calculating k-means. The descriptors are clustered on the basis of value provided by k-means function. The query image is then run through database and the similarity of the region of interest is judged as depicted in fig .7. The fig .8 is a snapshot of the command window of Matlab that shows all the frames that have been read.

This is the penultimate module of the query based search engine. The output of this module is fed to Full frame queries that produce the output. The false positives are eliminated to a great extent here. However, no machine is perfect and some false positives still show up in the results. This is described in detail in the Results section.

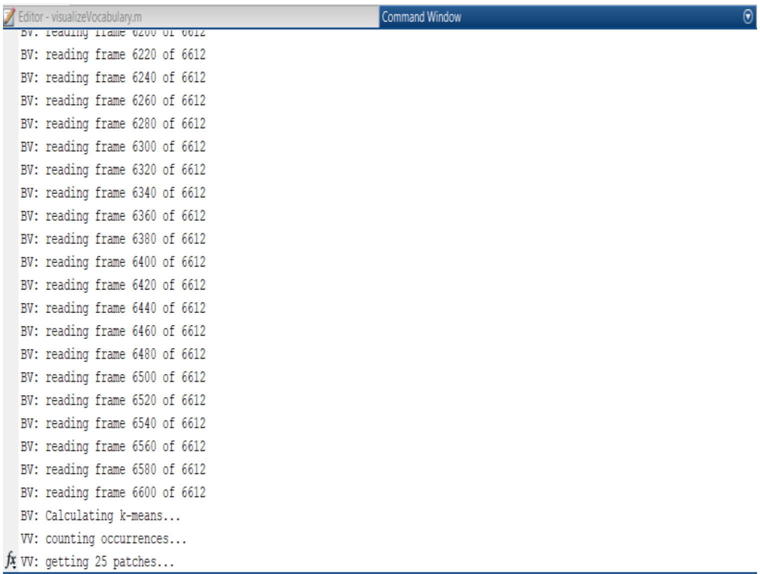

Figure 7: Reading frames

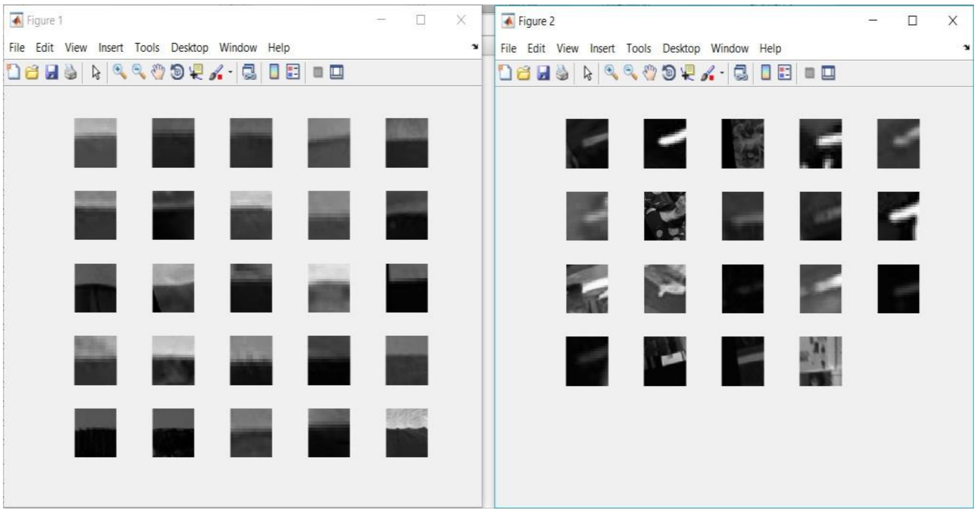

Figure 8: Visualizing vocabulary

2.13. Full Frame Queries

The final module of the query based search engine. After sorting the most relevant images found out by visualizing the vocabulary, the final module is executed. We have restricted the match images to 5, that are the five most similar images to the query image is displayed as output to the search engine. 4 different query images have been taken for the execution of the search engine. The region of interest is marked on them using a yellow box and the match rank 1 to 5 has been depicted by fig .9, fig 10, fig 11 and fig 12.

The output of the search engine does show some false positives, which is explained in detail in the Results section. As the size of the database is increased it is observed that the larger the size of the database the lesser is the number of false positives detected and hence higher the accuracy. However, it takes a lot of time to build. With the help of GPU's the time taken to execute can be taken care of. So for easier depiction of the project the database size has been limited to 6612. The final output of the query based search engine is as follows

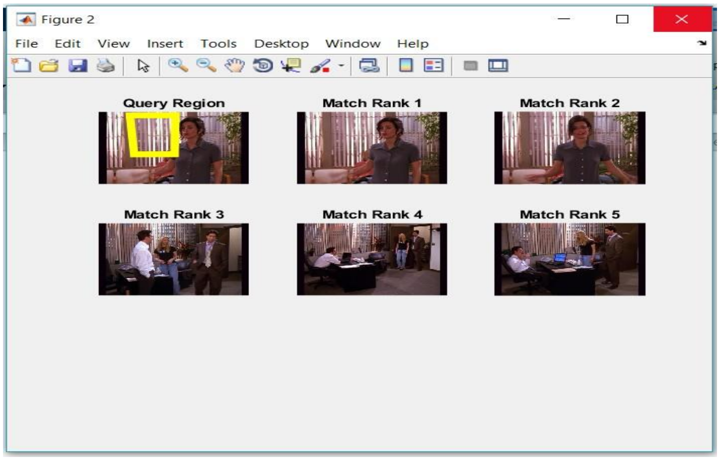

Figure 9: Output for query image 1

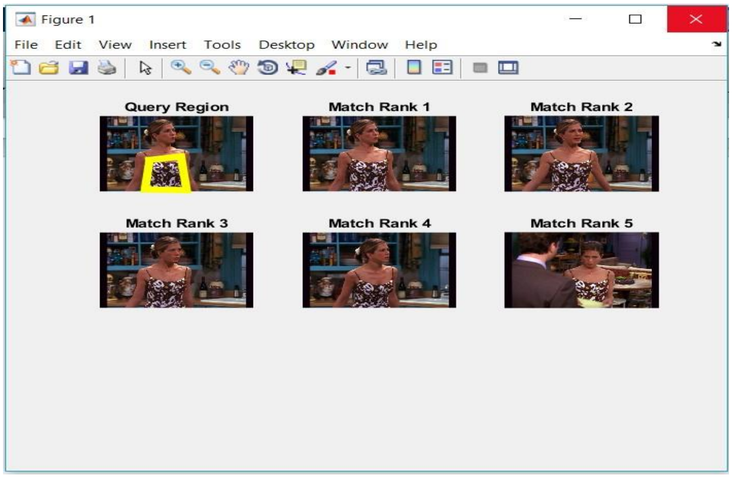

Figure 10: Output for query image 2

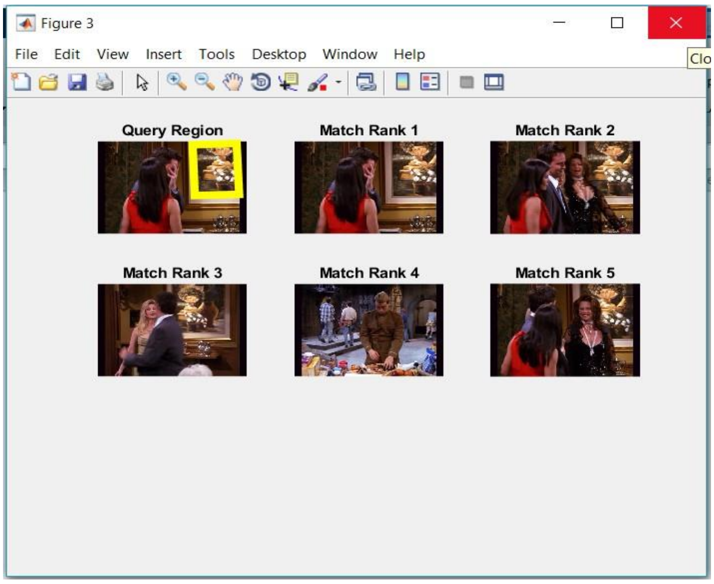

Figure 11: Output for query image 3

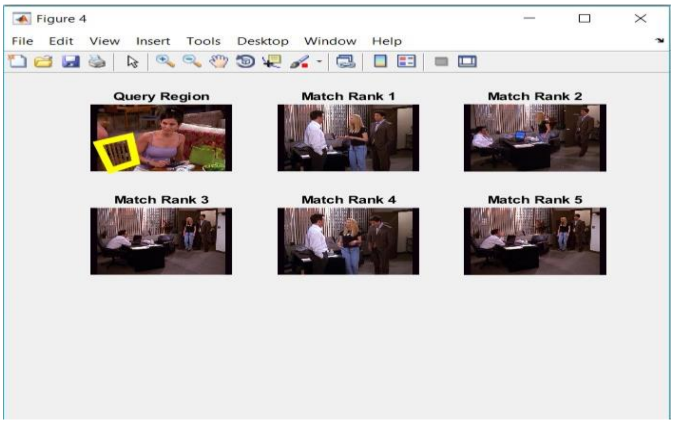

Figure 12: Output for query image 4

V. RESULTS

The figure 13: contains the query image in which the region of interest is outlined by yellow border.

Figure 13: Input Image

Received: 22 Aug 2019 | Revised: 13 Sep 2019 | Accepted: 15 Nov 2019

The results can be understood by studying the table.1

Table 1: Contingency Table

	Relevant Images	Non-Relevant Images
Retrieved Images	True Positive	False Positive
Non-Retrieved Images	False Negative	True Negative

TRUE POSITIVE (TP): Model correctly predicts positive class

TRUE NEGATIVE (TN): Model correctly predicts negative class

FALSE POSITIVE (FP): Model incorrectly predicts positive class

FALSE NEGATIVE (FN): Model incorrectly predicts negative class

The fig .14 shows TP, FP, FN outlined by blue boxes on the image.

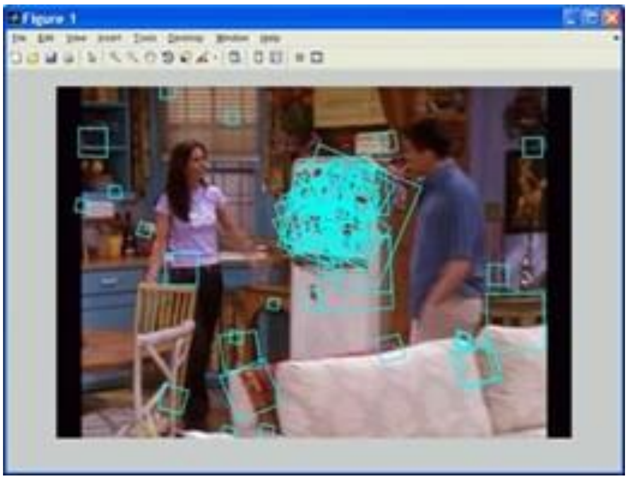

Figure 14: Output Image

The definition of precision and Recall is given as in fig.15.

$$\text{Precision}=\frac{True\ Positive}{True\ Positive+False\ Positive}$$

$$\text{Recall}=\frac{True\ Positive}{True\ Positive+False\ Negative}$$

Figure 15: Precision and Recall

To find the precision of our system we have used the dataset that contains 24000 images of F.R.I.E.N.D.S Tv show. When the system was trained and tested against 24000 images we found out that it was able to find about 6 out of 8 images. When the number of images increased to 50000, the system was able to find out 14 out of 15 images appropriately. Thereby we found out that the efficiency increased from 75% to nearly 94% with only a mere increase of 30000 images.

Hence we can conclude that the efficiency of the Image Query Based Search Engine increases with the increase in images in the dataset as depicted by the efficiency graph in Fig 16. For the time being we have used 4 episodes of the T.V. show. It has more than 100 episodes and therefore we can extend the database as much as we want. The more number of images used, the more rigorous and efficiency the training will be.

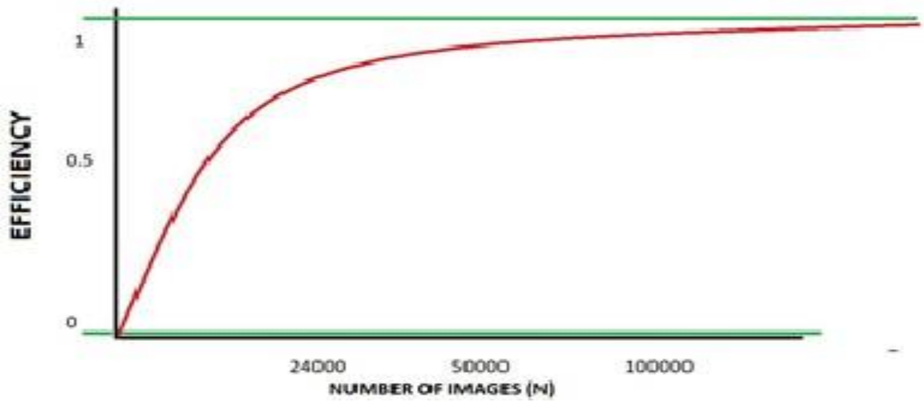

Figure 16: Efficiency Graph

VI. CONCLUSION

Managing huge amounts of data is the talk of the hour these days. A lot of organizations are taking actions to propose the most effective technique especially when it comes to visual data. The purpose of this paper is to define an efficient and truly realizable approach for image query based search engine. The query image is presented with the help of descriptors so that identification of relevant images is done efficiently even when there is a slight change in viewpoint or illumination of image and in cases of partial occlusion. A dictionary of visual words is created through raw descriptor matching by trying to utilize the same concept as text retrieval approach from a document. This content-aware Image retrieval system forms full frame queries from a region, matches it with the image query. It uses the Scale Invariant Feature Transform (SIFT) for feature detection in images. This process is trained on a pre-determined number of images until we can make a system with automatic feature learning.

The own dataset is created using a set of videos of a famous T.V. show makes our work unconventional. The efficiency of the image retrieval is found to be 94%, when more images are trained. For the time being, the dataset used is a set of images. We can extend the same concept and run this query based search engine on a video as input. This application can make video searches fast. There wouldn't be a need to go through the whole video looking for a particular scene. One tap and the results will be in front of us. Hence, with a few enhancements this work can solve a lot of our problems.

REFERENCES

1. JOSEF SIVIC AND ANDREW ZISSERMAN , A TEXT RETRIEVAL APPROACH TO OBJECT MATCHING IN VIDEOS. IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION (ICCV 2003)
2. HAILIANG LI, YONGQIAN HUANG AND ZHIJUN ZHANG, AN IMPROVED FASTER R-CNN FOR SAME OBJECT RETRIEVAL. JOURNAL OF LATEX CLASS FILES, VOL.14, NO.8, AUGUST2015.
3. BENJAMIN BUSTOS ,DANIEL KEIM AND DIETMAR SAUPE ,CONTENT- BASED 3D OBJECT RETRIEVAL.IEEECOMPUTER SOCIETY 2007.
4. PHILBIN, J., CHUM, O., ISARD, M., SIVIC, J., & ZISSERMAN, A. (2007). OBJECT RETRIEVAL WITH LARGE VOCABULARIES AND FAST SPATIAL MATCHING. 2007 IEEE CONFERENCE ON COMPUTER VISION AND PATTERN

RECOGNITION.

5. RICHANGHONG,MEMBER,IEEE,ZHENZHENHU,RUXINWANG,MENGWANG,MEMBER, IEEE, AND DACHENG TAO, FELLOW, IEEE. MULTI-VIEW OBJECT RETRIEVAL VIA MULTI-SCALE TOPIC MODELS. IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 25, NO. 12, DECEMBER 2016
6. LINAAL-QURAAN, SAWSAN NUSIRAND BELALA BUATA , CONTENT BASED WEB IMAGE SEARCH ENGINE EVALUATION USING ARABIC TEXT QUERIES. TRANSACTIONS ON IMAGE PROCESSING, DECEMBER
7. WICHIAN PREMCHAISWADI, ANUCHA TUNGKATSATHAN ,ON-LINE CONTENT-BASED IMAGE RETRIEVAL SYSTEM USING JOINT QUERYING AND RELEVANCE FEEDBACK SCHEME. WESAS TRANSACTION OF COMPUTERS ISSUE 5 VOLUME 9 MAY 2010.
8. DR.T.SANTHA, M.ABHAYA DEV , CONTENT-BASED IMAGE RETRIEVAL PUBLIC AND PRIVATE SEARCH ENGINES .UGC SPONSORED NATIONAL CONFERENCE ON ADVANCED NETWORKING AND APPLICATIONS, 27TH MARCH 2015.
9. PUSHPA M. CHUTEL, APEKSHA SAKHARE, REVERSE IMAGE SEARCH ENGINE USING COMPACT COMPOSITE DESCRPTOR. INTERNATIONAL JOURNA L OF ADVANCE RESEARCH IN COMPUTER SCIECNE AND MANAGEMENT STUDIES. VOLUME 2, ISSUE 1, JANUARY 2014.
10. A. ARAUJO, J. CHAVES, H. LAKSHMAN, R. ANGST, AND B. GIROD, “LARGE SCALE QUERY-BY-IMAGE VIDEO RETRIEVAL USING BLOOM FILTERS,” ARXIV, VOL. 1604.07939, 2016