Content based video retrieval system for distorted video queries

Ms Mohana C[1] , Keerthana B S [2] , Saimanimala S [3], Vandana K [4],
Department of Electronics and Communication Engineering, Dr T Thimmaiah Institute of Technology

Abstract: In today's world, huge amount of video data is being generated on a daily basis. The video information on any topic is much easier for the user to understand than textual information and hence most of the users prefer videos over textual information. Video databases are growing ever fast. In this case, searching for a particular video in a huge and varied collection of videos becomes challenging. Also, a user might have an image and may want to retrieve videos containing similar information.. Then in such cases, searching a huge database becomes difficult and highly time consuming. The proposed system reduces the manual task of searching similar videos thereby saving the user's time and effort. The video frames can be annotated to perform search for retrieving videos containing similar content. But this technique will not take into account the context information of the video. The proposed system retrieves video of similar content by taking into account texture, edge features as the context.

Keywords: *video retrieval, colour, texture feature.*

I. INTRODUCTION

Content Based Video Retrieval (CBVR) is used to retrieve desired videos from a large collection of videos on the basis of features that are extracted from the videos. A video has features like color, texture. These feature helps to describe the contextual information contained in the videos. Large volumes of videos are acquired and stored on computers due to frequent use of digital video devices in various areas. Video retrieval system is used for searching, browsing and restoring videos from excessive database of digital videos. The ability to manipulate and access stored videos is widely used by the users of different fields in variousways

The users find it difficult to locate the desired video in varied and huge collection of videos. Videos containing useful information are under-utilized unless CBVR systems are capable of retrieving desired videos by selecting relevant videos and filtering out the undesiredvideos. Search for solutions regarding the problems related to video retrieval has become the wide area for research and development. Content based video retrieval and analysis is the one of the most important and recent research areas in the Image processing domain. Content based video retrieval can be used for multiuser systems for video search and browsing which are useful in web applications Video retrieval system has acquired much importance because of the richness of the visual information contained in the video, thus Text based video retrieval methods have become inefficient methods. Content based video retrieval methods are applied in order to improve the effectiveness of the content-based methods. And these methods can play essential rules whichare used in media collection and enhance retrieval accuracy. Therefore, content-based video retrieval (CBVR) plays a vital role instead of text-based retrieval in multimedia systems. With few advances' multimedia technologies, digital TV and information highways, more and more video data have been captured, produced and stored. However, without appropriate techniques that can make the video content more accessible, allthese data are hardly usable research Hence, Content-based video retrieval systems have established as powerful tools for finding specific content in ever-growing large-scale video collections Video retrieval tools are typically built around a retrieval engine that returns a ranked video list according to various features.

Content-based video retrieval system for distorted video queries is a relatively new research area in the field of computer vision and multimedia information retrieval. The first content- based video retrieval systems were developed in the early 1990s, but they were limited to retrieving videos based on textual metadata and manual annotations. In the early 2000s, researchers began to explore the use of video content analysis techniques for developing content-based video retrieval systems. The focus was mainly on retrieving videos based on low-level visual features such as color, texture, and shape. However, the performance of these systems was limited due to the lack of robustness to variations in viewing conditions, such aslighting changes, camera motion.

In the present days, the users prefer videos for illustration and understanding purposes. The typical applications include video lectures, products demonstration and installation. It is estimated that the web constitutes around 85% of the multimedia content. Nearly 300 hours of video are uploaded to YouTube every minute, Face book users generate more than 8 billion video views per day and Google's Nest reports more than 100 hours of surveillance video uploads per minute.

This poses a real challenge in handing these voluminous video data. The efficient storage and access of multimedia data means, we can avoid the customary manual process of searching and retrieving them based on metadata. The retrieval system can be made automatic and powerful by facilitating content-based search. Traditional textual based video retrieval technique faced some drawbacks in the process of indexing and retrieving the videos from large databases. The process has been time consuming as it takes long time to analyze the metadata and manually assign a textual data with it. It also involves a different insight abouta certain video and hence the results may not match the user's expectations, since it is a subjective process.These challenges have paved the way for the next generation video retrieval system. Videos are very rich in visual content like color, texture, shape and motion, these can be used to resolve the drawbacks of the traditional retrieval system These retrieval systems that function on the content rather than the metadata of the video are called as content-based retrieval systems with few advances multimedia technologies, digital TV and information highways, more and more video data have been captured, produced and stored.

Content based video retrieval can be used for multiuser systems for video search and browsing which are useful in web applications. Video retrieval system has acquired much importance because of the richness of the visual information contained in the video, thus Text based video retrieval methods have become inefficient method

II. LITERATURE REVIEW

[1] Ahmad Alzu'bia et.al [1] proposed, **"Content-Based Image Retrieval with Compact Deep Convolutional Features"**, Convolutional neural networks (CNNs) with deep learning have recently achieved a remarkable success with a superior performance in computer vision applications. Most of CNN-based methods extract image features at the last layer using a single CNN architecture with order less quantization approaches, which limits the utilization of intermediate convolutional layers for identifying image local patterns. As one of the first works in the context of content-based image retrieval (CBIR), this paper proposes a new bilinear CNN-based architecture using two parallel CNNs as feature extractors The activations of convolutional layers are directly used to extract the image features at various image locations and scales. The network architecture is initialized by deep CNNs sufficiently pre-trained on large generic image dataset then fine-tuned for the CBIR task. Additionally, an efficient bilinear root pooling is proposed and applied to the low-dimensional pooling layer to reduce the dimension of image features to compact but high discriminative image descriptors. Finally, an end-to-end training with back propagation is performed to fine-tune the final architecture and to learn its parameters for the image retrieval task.

[2] Salahuddin Unar **"A decisive content-based image retrieval approach for feature fusion visual and textual images"**, Image content analysis plays a dynamic role in various computer vision applications. These contents can be either visual (i.e., color, shape, texture) or the textual (i.e., text appearing within images). Both the contents involve fundamental characteristics of an image and thus can be an enormous asset for any intelligent application. For content-based image retrieval (CBIR) systems, most of the art methods are either annotated text based or the visual search based Due to high demand of multitasking, there is a great need of a system that can combine visual as well as textual features. Consequently, this work proposes a decisive CBIR approach that combines visual and textualfeatures to retrieve similar images. Firstly, the method classifies the query image as textual andnon-textual. If any text appears within the image, then the query image is classified as textual, and the text is detected and formed as Bag of Textual words. If the query image is classified as non-textual, the visual salient features are extracted and formed as Bag of Visual words. Next, the method fuses the visual and textual features, and top similar images are retrieved based on the fused feature vector. It supports three modes of retrieval

[3] R. Rani Saritha, **"Content based image retrieval using deep learning process"**, Content-based image retrieval (CBIR) uses image content features to search and retrieve digital images from a large database. A variety of visual feature extraction techniques have been employed to implement the searching purpose. Due to the computation time requirement, some good algorithms are not been used. The retrieval performance of a content- based image retrieval system crucially depends on the feature representation and similarity measurements. The ultimate aim of the proposed method is to provide an efficient algorithm to deal with the above-mentioned problem definition. Here the deep belief network (DBN) method of deep learning is used to extract the features and classification and is an emerging research area, because of the generation of large volume of data. The proposed method is tested through simulation in comparison and the results.

The ultimate aim of the proposed methodis to provide an efficient algorithm to deal with the above-mentioned problem definition. it requires the retrieval of the most relevant video from a large collection as well as localizing the start and end timestamps of a segment that matches the text query best from the video.

[4] El Mehdi Saoudi **"A distributed content-based video retrieval system for large datasets"**, With the rapid growth in the amount of video data, efficient video indexing and retrieval methods have become one of the most critical challenges in multimedia management. For this purpose, Content-Based Video Retrieval (CBVR) is nowadays an active area of research. In this article, a CBVR system providing similar videos from a large multimedia dataset based on query video has been proposed. This approach uses vector motion-based signatures to describe the visual content and uses machine learning techniques to extract key frames for rapid browsing and efficient video indexing. The proposed method has been implemented on both single machine and real-time distributed cluster to evaluate the real- time performance aspect, especially when the number and size of videos are large. Experiments were performed using various benchmark action and activity recognition datasets and the results reveal the effectiveness of the proposed method in both accuracy and processing time compared to previous studies. convolution is essentially sliding a filter over the input rather than looking at an entire image at once to find certain features it can be more effective to look at smaller. This layer consists of a set of learnable filters that we slide over the image.

[5] Xiao Sun , **"VSRNet: end-to-end video segment retrieval with text query"**, Users are sometimes interested in specific segments of an untrimmed video when using the video search engine. Targeting at this demand, we explore a novel research topic of text query-based video segment retrieval (VSR). Different from the conventional videoretrieval task or localizing

text descriptions in a single video, it requires the retrieval of the most relevant video from a large collection as well as localizing the start and end timestamps of a segment that matches the text query best from the video. A direct solution is to perform video-level matching first, and then apply description localization among such video candidates. Such two-stage based methods are not able to utilize complementary information of each stage, and are time- consuming in inference. . In this paper, we propose VSRNet, an end-to-end framework that efficiently retrieves video at segment granularity with two branches. In the first branch, individual videos and texts are mapped to a common space for stand-alone ranking. In the second branch, we propose a supervised text-aligned attention mechanism and calculate the response of every frame to the text query, from which the frameswith high scores are aggregated as segment proposals. Extensive experiments conducted on Activity Net Captions.

Software Requirements

The software requirements are:
Anaconda/PyCharm/Jupyternotebook.

III. METHODOLOGY

The system consists of a database of video files from which key frames are extracted. From the key frames that are obtained from the videos, features like color, texture, edge are extracted from each of the image**(fig 1).** Further, when a user inputs a test image or a query image for the videothe user wants to retrieve, the color and texture features are extracted from the query image.

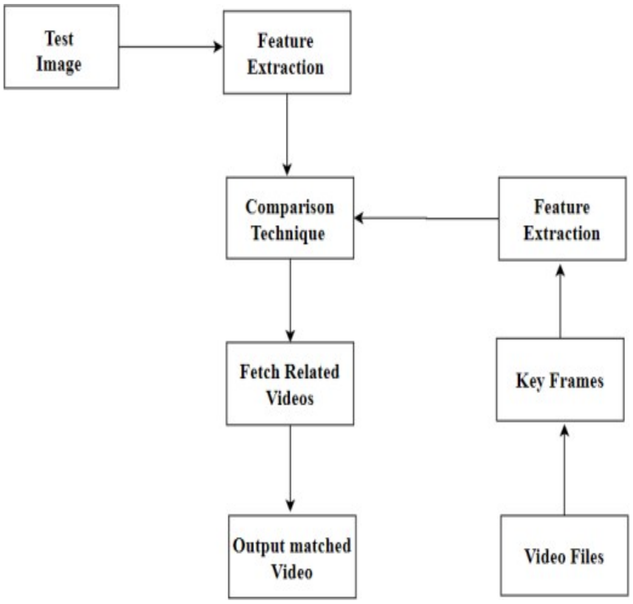


Fig 1. Block Diagram of Proposed Method

The general idea of our CBVR system is:

1. For the Videos stored in the database, Key frames are extracted using Key Extraction Method by implementing katna.
2. Feature Extraction is done using two methods:
- Texture Feature: It is extracted using GLCM for the key frames and inputimage.
- Color Feature: It is extracted using Color Histogram for the key Frames andinput image.
- Edge Feature: It is extracted using Canny Edge Detection.
3. Similarity Measure which is done using Euclidean Distance which compares the inputimage and key frames extracted.

Working of Key Frame

Since there is lot of redundancy in the video for simplicity, we can select key frames from thevideo so that the key frame will represent the whole of the video**(fig 2).** A key frame is a representative image for a video. Key frame extraction plays an important role in the analysis of large amount of video files. Key Frame extraction reduces the useless information of the video. The key frame extraction is done using katna. The number of key frames to be extractedfrom a

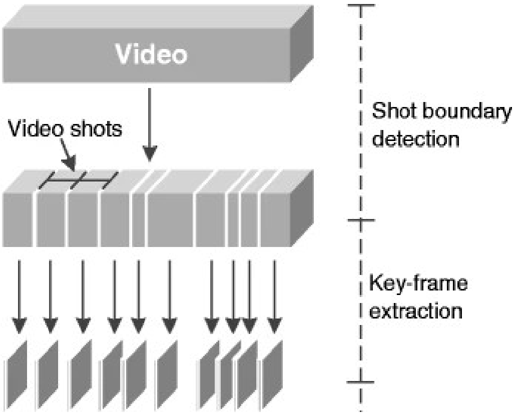


Fig 2. The key frame

video can be limited. There are two modules: Video Module and Image Module. The first step in the key frame selection process is to identify potential key frames from the video query.

Steps involved in Key frame extraction:

1. Video segmentation: The first step in keyframe extraction is to divide the video into smaller segments. This can be done using techniques such as shot boundary detection,which detects changes in the visual content of the video, such as changes in camera angle or scene transitions.

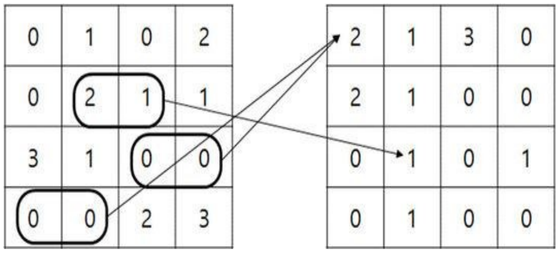
2. Feature extraction: Once the video has been segmented, features are extracted from each frame in the segment. These features could include color histograms, edgehistograms, motion vectors, or other visual features that can be used to describe the content of the frame.

3. Keyframe selection: Keyframes are selected based on some criteria that capture the most representative frames of the video segment. There are different criteria that can be used to select keyframes such as minimum distance, maximum diversity, and clustering-based methods.

4. Keyframe summarization: After keyframe selection, the final step is to create a summary of the video content using the selected keyframes. This can be done by concatenating the keyframes in the order they appear in the video, or by using a more sophisticated approach that considers the relationships between the keyframes.

Working of GLCM

In content-based video retrieval systems for distorted video queries, Gray-Level Co-occurrence Matrix (GLCM) is a texture analysis method that is commonly used to extract texture features from video frames



0	1	0	2
0	2	1	1
3	1	0	0
0	0	2	3

2	1	3	0
2	1	0	0
0	1	0	1
0	1	0	0

Fig 3:Original matrix and GLCM

Steps involved in GLCM extraction:

1. Gray-level quantization: The next step is to quantize the image into discrete gray- level values. This can be done using a fixed quantization scheme or an adaptive scheme based on the image histogram.

2. GLCM computation: The GLCM is then computed by analyzing the spatial relationships between pairs of pixels in the quantized image (fig 3).

3. Feature extraction: Features are then extracted from the GLCM by computing statistical measures such as contrast, correlation, energy, and homogeneity. These measures can be used to describe the texture properties of the image.

4. Feature selection: Finally, a subset of the extracted features may be selected based on some criteria such as relevance, redundancy, or computational complexity. This can help to reduce the dimensionality of the feature space and improve the performance ofsubsequent image analysis tasks.

Working of Color Features

Colour features are one of the essential visual features used in content-based video retrieval systems for distorted video queries. The colours of objects and scenes in video frames provide a vital clue to their identity and can be used to index and retrieve similar videos based on theircolour characteristics**(fig 4).** In a content-based video retrieval system, colour features are extracted from each key frame and stored in a database. There are several ways to extract colour featuresfrom video frames, such as colour histograms, colour moments, and colour correlograms.

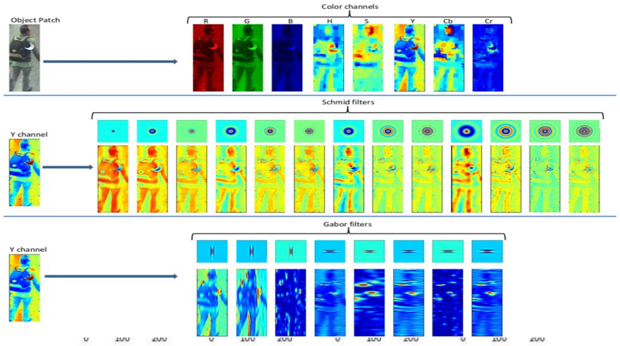


Fig 4: Colour feature extraction

Steps involved in Colour Feature extraction:

1 Colour space conversion: The first step in colour feature extraction is to convert the image or video from the original colour space

2 Colour quantization: The next step is to quantize the colours in the image or video into a set of discrete colour bins.

3 Colour histogram computation: Once the colour bins are defined, a colour histogram is computed to represent the distribution of colours in the image or video.

4 Colour moment computation: Colour moments are statistical measures of the colour distribution in an image or video.

5 Colour correlation matrix computation: The colour correlation matrix represents the correlation between different colour channels

6 Feature selection: Finally, a subset of the extracted colour features may be selected based on some criteria such as relevance, redundancy, or computational complexity.

Working of Edge Feature Extraction

Edge features are derived from the edges in an image or video frame**(fig 5).** Edges represent the boundaries between regions of the image or video frame with different colors or textures.

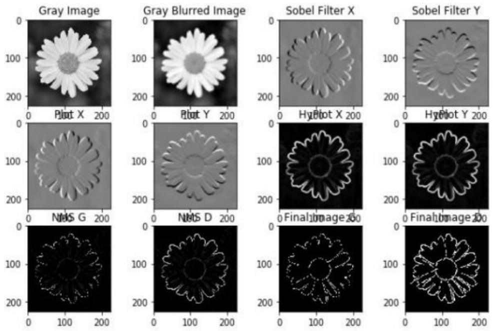


Fig 5: Edges detected for an image

Steps involved in Colour Feature extraction:

1. **Gaussian Blur:** The first step is to apply a Gaussian filter to the input image or video frame to remove any noise and smoothen the image.
2. **Gradient Calculation:** Next, the image gradient is calculated by convolving the Gaussian-filtered image with a Sobel kernel
3. **Non-maximum Suppression:** In this step, the gradient magnitude is calculated at each pixel and compared with the magnitudes
4. **Double Thresholding:** This step involves setting two thresholds: a low threshold and a high threshold. The high threshold is used to detect strong edges, and the low threshold is used to detect weak edges
5. **Edge Tracking by Hysteresis:** In this step, the weak edges are tracked by connecting them to strong edges. If a weak edge pixel is connected to a strong edge pixel, it is considered part of an edge, and if not, it is discarded

Working of Similarity Measuring

Content-based video retrieval systems typically rely on a similarity measure to compare the features of the video being queried against those of videos in the database

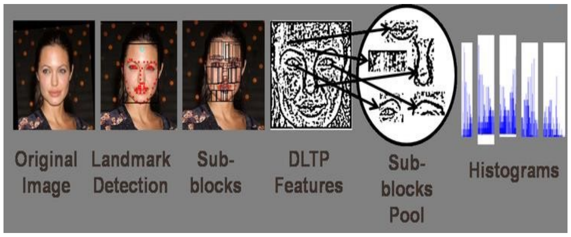


Fig 6:Similarity measurement and classification method for face verification

Steps involved in Similarity Measuring:

1 Feature Extraction: Features are extracted from the videos being compared. These features can include color histograms.

2 Feature Representation: The features extracted from each video are represented as avector in a high-dimensional feature space

3 Distance Metric Selection: A distance metric is selected to compute the distance between the two feature vectors

4 Similarity Computation: The similarity measure is computed by applying the distancemetric to the feature vectors.

5 Ranking: The retrieved videos are ranked based on the similarity score, and the top- ranked videos are presented to the user as the most relevant matches.

IV. RESULTS & SIMULATION OUTPUT

PyCharm

PyCharm is an integrated development environment (IDE) used for Python programming. It isdeveloped by JetBrains and offers advanced features for coding, debugging, and testing Pythoncode.

I Pre-processing

The first step in the process is to pre-process the video data. This involves extracting keyframesfrom the video and converting them into a suitable format for feature extraction.

II Feature Extraction

Once the keyframes have been extracted, the next step is extract features from them. One ofthe most common methods for feature extraction in video retrieval systems is the Canny edge detection algorithm

III Feature Matching

After the features have been extracted, the next step is to match them with the features of the query video. This is done by calculating the similarity between the features of the keyframes in the query video and the features of the database videos and temperature has been performed on 120nm.

IV Query Processing

Once the similarity scores have been calculated, the next step is to process the query and retrieve the most relevant videos from the database. This can be done by ranking the videos based on their similarity scores and presenting the top results to the user.

V. IMPLEMENTATION

Fig 7 : **Input frame**.

extractor.py

Convent video to image frames for further processing by OpenCV

Dividend each image frame to blocks

Extract features for each block color-based method which extracts the average intensity ofred, green and blue.

Reduce the dimension of features by discrete cosine transform features at top are moreimportant

Save features of all frames in JSON

Searcher.py

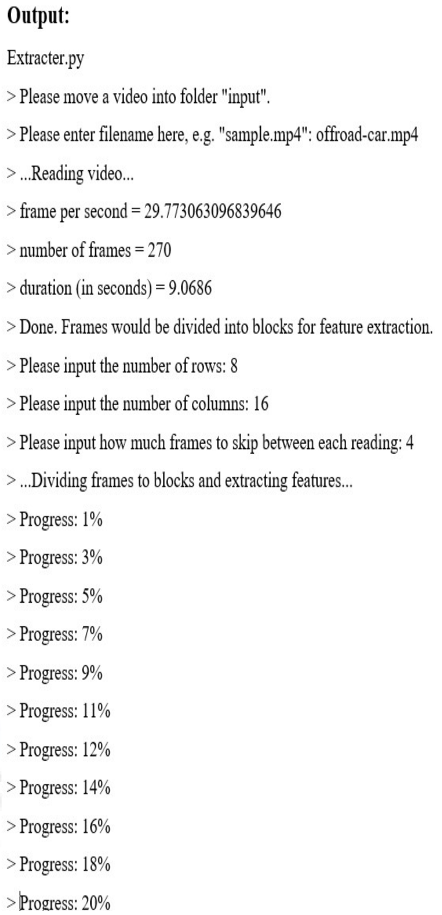
Convent video to image frames for further processing by OpenCV

Dividend each image frame to blocks

Extract features for each block color-based method which searches the average intensity ofred, green and blue.

Reduce the dimension of features by discrete cosine

transform features at top are moreImportant
Saves features of all frames.



```
Output:

Extracter.py
> Please move a video into folder "input".
> Please enter filename here, e.g. "sample.mp4": offroad-car.mp4
> ...Reading video...
> frame per second = 29.773063096839646
> number of frames = 270
> duration (in seconds) = 9.0686
> Done. Frames would be divided into blocks for feature extraction.
> Please input the number of rows: 8
> Please input the number of columns: 16
> Please input how much frames to skip between each reading: 4
> ...Dividing frames to blocks and extracting features...
> Progress: 1%
> Progress: 3%
> Progress: 5%
> Progress: 7%
> Progress: 9%
> Progress: 11%
> Progress: 12%
> Progress: 14%
> Progress: 16%
> Progress: 18%
> Progress: 20%
```

Fig 8: Output of the extractor video

VI CONCLUSION AND FUTURE SCOPE

Conclusion

The search for a particular video in a huge and enormous set of videos require a lot of time andeffort. The proposed system reduces the delay by automating the task of video retrieval by saving the users time, effort and energy. The proposed extracts multiple features from an image to aid in the process of video retrieval. To retrieve the desired videos correctly, only one feature like the color feature may not be enough. Hence, the system also considers the texture feature. Multiple feature extraction from an image would provide better accuracy compared to single feature. The accuracy of the system is about 70%. To achieve a better accuracy, the system could be trained with a larger and varied set of videos with higher configuration system and high-speed processing systems.

Future Scope

The future scope of Content Based Video Retrieval System for Distorted Video Queries includes the integration of AI and machine learning algorithms, real-time processing, improved accuracy, cloud-based storage and processing, and personalized video recommendations through collaborative filtering. As the technology continues to evolve, we can expect more advanced and sophisticated systems that are capable of handling larger volumes of video datawhile providing more accurate and relevant results to users

REFERENCES

[1] Content-based image retrieval with compact deep convolution feature" Author links open overlay panel AhmadAlzubi, Abbes Amira, Naeem Ramzan School of Engineering and computing, University of West of Scotland, Paisley PA12BE, UK College of Engineering Qatar University,22-12 IEEE.

[2] B. Salahuddin Umar, Xingyuan Wang, Champing Wang, Yu Wang, published apaper on "Content Based Image Retrieval" approach for feature fusion invisual and textual images,33- 14 IEEE.

[3] In 2021 Xiao Sun, Xiang Long, Dongxiang Hemsley Wenzhou Lian published a paper on "VSRNet: end-to-end video segment retrieval with text query",23-11

[4] M. Miyahara, Y. Asada, P. Daehwa and A. Matsuzawa, In 2021 EI MEHDI Saudi and Said Jai-Andalusia, published a paper on "A distribution Content-Based Video Retrieval system for large datasets",112-34-55 2021 IEEE

[5] Asif Ansari and Muzammil H Mohammed "Content Based Video Retrieval System methods, Techniques and Challenges", International Journal of Computer Application (0975-887) Volume 112-No.7, February 2015

[6] Janarthanan.Y, Balaji J.M and Srinivasa Raghavan's "Content Based Video Retrieval and Analysis using Image Processing", International Journal of Pharmacy and Technology, December 224-2-2016.