Journal of Theoretical and Applied Information Technology
15th December 2023. Vol.101. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

CONTENT BASED VIDEO RETRIEVAL USING LOW-LEVEL FEATURES

SUBRAMANYAM KUNISETTI[1], SUBAN RAVICHANDRAN[2]

[1] Research scholar, Department of CSE, Annamalai University, Chidambaram, Tamil Nadu.
[2]Associate Professor, Department of IT, Annamalai University, Chidambaram, Tamil Nadu.

E-mail id: [1]subramanyamkunisetti@gmail.com,[2]rsuban82@gmail.com

ABSTRACT

Over the past ten years, researchers have studied the Content Based Video Retrieval System, based on numerous applications, advancements, and technologies. In video retrieval systems, there is still a requirement for the high-level semantic elements as and the processing of low level materials. As a result, it inspires and motivates a lot of academics to learn more about the content Retrieval and to make more useful and effective while creating system applications. Analysis of video for retrieval of key aspects is regarded as earlier work. In this instance, input videos from YouTube are watched for analysis. After that, foreground segmentation is carried out to find a lot of tiny subset objects since each subset must identify the foreground video class. Given for feature extraction is the split region. In this study, four distinct methods for extracting features—chromatic moment, blur, color variety, and reflection features are taken into consideration. The high dimensionality characteristics are removed from the retrieved features using Principal Component Analysis since they may affect classification accuracy. The feature vectors are taken into account while combining all of the parts. The Nave Bayes classifier is used to complete the classification process. Metrics including accuracy, precision, recall, and F-measure are used to gauge how effectively video retrieval is performed. The predicted model outperforms the current strategy when the proposed model and the in-depth learning approach are compared.
Keywords: *Content Retrieval, Foreground Segmentation, Image Retrieval, Chromatic Moment, Naïve Bayes Classifier.*

1. NTRODUCTION:

The method of searching and retrieving videos based on their visual content as opposed to just their metadata or textual content is known as content-based video retrieval. It entails assessing video visual aspects and qualities such as color, texture, structure, motion, and semantic content in order to locate pictures that are visually similar to or related to a particular query[1]. Content-based video retrieval tries to automatically analyze and interpret the visual content of videos, in contrast to conventional video retrieval techniques that rely on human annotations or textual descriptions. It enables users to search for films based on their actual content rather than how they are labelled or described by extracting and comparing visual elements. Content-based video retrieval systems typically comprise two major steps i.e., feature extraction and similarity matching. With usage of deep learning methods like CNNs or 3D CNNs, different visual features are recovered from video frames or sequences during the feature extraction stage. These features record various elements of the video, such as things that are in the scene, events that are taking place, or patterns of movement. Multimedia information systems are becoming more crucial due to the growth of the Internet, and as well as increasing multimedia data visualization. A multimedia database should be efficiently used to extract, store, as well as to retrieve video data[2]. Utilizing video content for video archiving and searching presents a challenge to multimedia administrators. Many approaches came in recent years to index and retrieve films based on content properties. The features of the video content, like object, motion, speech, etc., can be extracted. The retrieved features are analysed in the similarity matching process to assess the degree of similarity or distance between videos. This comparison can be based on multiple similarity measures, such as Euclidean distance, cosine similarity, or learnt distance metrics using deep learning models[3]. Videos with comparable visual elements or substance are deemed relevant and are retrieved as results. Content-based video retrieval has several applications, including video search engines, video recommendation systems, video surveillance, video summarization, and video content management. It allows users to search movies based on visual similarity, identify related or comparable videos to a particular query, and explore video collections based on their real content. The efficiency and accuracy of content-based video retrieval have been considerably enhanced by the developments in deep learning algorithms, such as CNNs, RNNs, and 3D CNNs. These methods allow for more accurate and robust retrieval based on visual information since they can learn from video data and derive meaningful representations from it.

In order to search videos, organize and retrieve videos with content base, effective techniques are required given the exponential growth of video content on platforms like social media, video sharing websites, and surveillance systems. Automatic visual feature extraction and representation from video data is now possible because of deep learning algorithms, which have shown to be potent tools in this field. Convolutional and recurrent neural networks, two types of deep learning models, have shown unique efficiency in tasks like object and activity recognition and scene comprehension. These models may learn complex patterns and representations from raw video data by using the hierarchical and compositional characteristics of neural networks, enabling more accurate and robust content-based video retrieval. The manual annotation and textual information used by traditional video retrieval techniques can be time-consuming and prone to mistakes. Deep learning-based techniques, on the other hand, can discover discriminative features automatically from video data, removing the need for considerable manualmat labelling. These models can collect both geographical and temporal information, allowing for more extensive video analysis and retrieval. RNNs are used to capture temporal dependencies and motion patterns, CNNs to collect spatial features from individual frames, and 3D CNNs to develop spatiotemporal representations are just a few examples of deep learning techniques for video retrieval. Siamese networks are used to learn video-to-video similarity metrics, whilst variational autoencoders (VAEs) can build compact representations of video material. For a more complete visualization, two-stream networks combine geographical and temporal information. Deep learning approaches enable content-based video retrieval systems to rapidly analyze and index huge video libraries, enabling accurate video search and recommendation, and support a variety of applications such as video surveillance, video summarization, and content-based video analysis. This study's motivation is to more effectively tackle the challenge of video retrieval and expand space available for searches in the database. A combination of characteristics and video content forms the basis for the video searching. This might improve video retrieval accuracy. The proposed model's important contribution is outlined as:

1) YouTube Database is taken as input. The video frames are divided into foreground segments and background segmentation.
2) Four distinct methods like Chromatic moment, blur, color variety, and reflection properties, features are extracted.
3) Principle Component Analysis (PCA) is used to combine the mean values obtained utilizing feature extraction methods. The term "feature fusion" refers to this method.

In this area, ongoing research addresses the drawbacks of video understanding, representation, handle large-scale video datasets, increase the efficiency and accuracy of deep learning models, and provide strategies that can generalize effectively across a variety of video content. The use of deep learning in video retrieval is evolving, leading to breakthroughs in analysis, search, and determining future of content management and comprehension.

For the majority of video databases, retrieval of data is unfeasible, necessitating query-based search. The assumption of a suitable and comprehensive collection of tags may not be accurate given the constantly growing size of video databases, making content-based video retrieval approaches essential. Users have been drawn to content-based video retrieval because, in contrast to text-based search, they use multimedia material far more frequently in daily life. However, content-based retrieval is extremely important when dealing with massive databases or data warehouses. CBVR system considers the following steps:

1. Analysis of input video composition.
2. Key frames recognition and segmentation.
3. Region of Interest for prominent features extraction method is to be determined.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4. Developing a database that is segmented and uses a key frame scheme in accordance with the application's nature.
5. Comparing and extracting information content from databases using database queries.
6. Using the query videos, retrieve videos from a huge multimedia database.

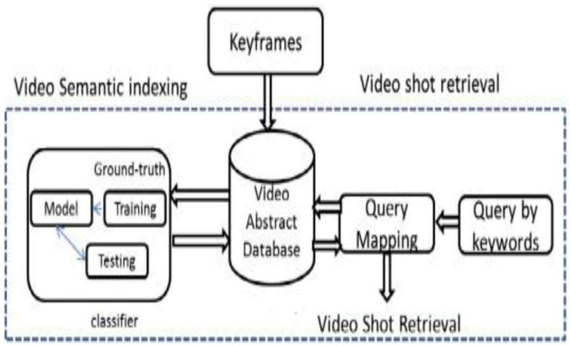

Figure 1: System architecture

2. RELATED WORK:

The video retrieval is the subject of numerous studies in recent years. These researches have made significant advances in accuracy and processing speed, but their diversity of content identification and experimental datasets remains limited. A word detector using video input was put forth by Yu Youngjae et al., who claimed that it could be trained end-to-end alongside any language models without a need of outside knowledge sources [15]. In 2016 Large Scale Movie Description Challenge, the idea was put to the test. [16] handled audio-based video retrieval by taking audio similarity measures between video pairs. Wavelet-based signatures are used in the CBVR system presented by Robles Oscar D. et al. in [4]. This method computes a multiresolution plot using the Haar transform. Signatures were described in XML2, a common description language. 62 movies with 817 recordings were used as the database for the experiments.

Local feature descriptors are employed by Dipika H. Patel [5], who is renowned for their rapid and reliable detector search capabilities. The distribution matrix and Hessian-like techniques are also offered, which enhances system performance. SURF proved to be more useful to the author than earlier methods like SIFT, PCA shift. There are a few techniques used in video database management system technology that have been mentioned and discussed in reviews [6,7,8,9,10]. These techniques include color correlation invariance, image segmentation, keyframe extraction.

[11] developed his Kekre transform approach similar to discrete walsh and haar transform. In essence, to lower the size of the feature vector commonly use factor coefficients via the modified frames. We use similarity metrics such mean squared error (MSE), accuracy, and recall. When utilizing the Haar transform, the average precision and recall are observed to be 0.

To test the effectiveness of image, Chao et al. [12] predicted JPEG with CBIR query photos and they came to the conclusion that using a Hessian Affine detector a powerful JPEG yields good and reliable performance. Color and texture features, as well as the extraction of features utilizing the Gabor filter and color histogram. [13] The Multi-Scale technique attempted to do away with the necessity for scale selection, while still enhancing the quality of natural photographs. [14] explains a method for enhancing algorithms by using the HIS color space. To enable the technique to function properly, the pixels were spitted using a classification method.

The video segmentation presented by [21] uses various change detection, adaptive threshold techniques, to determine threshold noise and to this noise-robust approach is utilized. Comparing the resulting segmentation results to those from other techniques, they exhibit low computation convolution and great efficiency.

Key frame summarizing utilizing histogram-based techniques are both handled by a system presented by Zhong Qu et al. in [96]. The segmentation of video shots include techniques including Histogram intersection, etc. Depending on image entropy for each HSV frame each of these shots was given consideration with a key frame. The testing outcome demonstrates that the method provides an improvement at certain degree.

In order to identify the significant events in aired videos, [19] provide video summarizing framework. The system derives significant incidents and significant actors from player popularity and sentiments. Authors have therefore successfully identified the crucial players and crucial events.

A System with level features is presented by Santhosh Kumar K. et al. in [20]. The feature vector is chosen using methods like Mean, Standard deviation, and Entropy. Utilizing the Euclidean Distance method, the similarity measurement is processed. The adoption of a multi-feature technique increased precision and recall by 18 to 16%, respectively, and the results attained up to 60% on average.

The automatic signage video detection was proposed by S.S. Gornale et al. [21]. Traffic signs were the primary focus of the work's content. SVM and KNN are then used to classify these signs using the LBP feature vector.

Using a statistical distance method and the behavioral content of the videos, L Zelnik Manor et al. [22] provide video sequence extraction. The system's nonparametric nature allows it to manage a variety of dynamic occurrences. The measure is applied by the authors for a variety of multimedia applications. They observed event-based accuracy is less in these models and more advanced.

In order to offer video shot boundary detection, Nikita Sao [23] first introduces graph theory, that determines set of nodes based on graph groups. Their discontinuity is then elevated and contrasted with a threshold value. The nodal analysis is effective, with positive outcomes of boundary detection.

A method for classifying objects into many classes offline has been put out by Chandra Mani Sharma et al. [24]. Multiclass Exponential Boosting, Stage-wise Additive Modelling, and Haar-like features have all been employed. The approach has a detection accuracy of 98.30% when processing 25 frames per second.

Using the adaptive threshold method, [25] described a method for video segmentation. Levels semantic annotation is used to represent the video. When compared to a conventional system that retrieves videos using color features, the experimental results are 13% to 19% higher.

A multilayer feature and Latent Structured SVM (LSSVM)-based action recognition system was suggested by Xingxiao Wu [26]. For the representation of local spatiotemporal situations, the Gaussian mixture model (GMM) is used. The LSSVM approach is employed to ensure compatibility between features and labels. The UCF50 database, YOUTUBE, and UCF Sports databases were used for the trials. As a result, they were able to get 86.91% for SVM and 88.04% for LSSVM.

A CBVR video segmentation is presented by S. Padmakala et al. in [27]. LBP and color characteristics are used in the segmentation process, while OFR is used to determine the best frame. The distance between these features is compared using feature weighting. The video is retrieved only if the distance between the matches are smaller than threshold value. Various experiments were found to be effective and well-done by the authors.

By combining frequency analysis with a 2-D method, [28] described a system, When used in conjunction with the current underpinning approach of SURF implementation, experimental analysis demonstrates that the proposed strategy performs better.

For video recovery, [29] employed feature extraction technique. For online data retrieval, the suggested approach has achieved recall rates and precision of 80%. The experiment revealed that the performance of the multiple feature CBVR system was superior to that of the single feature CBVR system.

A video search and indexing system built on SVM has been reported by Paul et al. [30]. The system makes use of state transition rectification techniques and estimates transition quality. The Dynamic Power Management approach measures energy efficiency. The SVM result had a precision rate of approximately 83.83%, while the query results from indexing the graph had an accuracy of 80%.

S.No	Methods adopted	Database	Research Findings
31	Object tracking, detection and segmentation of motion	Feature database	The authors of egoMotion conclude that because static object annotation is lacking from the FBMS, state-of-the-art performance cannot be attained despite the method's successful results for YouTube, SgeTrackv2, and egoMotion.
32	Semantic gap between low and high levels	Feature library	According to the trial findings, recall and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

			precision both averaged up to 60%. Additionally, the employment of a multi-feature technique increased precision and recall by 18 to 16%, respectively.
33	text identification and localization	Feature library	The multi model and cross video retrieval is the focus of this work. These approaches have some flaws that have been found.
34	OCR approach	Feature library	The erroneous rate is eliminated in this case using the stroke width transform. The spelling is checked and integrated at the very end by OCR. The precision may be superior, but it disappears when the noise level is larger.
35	automatic system	Feature library	Keywords and text content are utilized for testing. The study of multiple videos shows that accuracy is improper.
36	automatic approach	Feature library	Slides from a video were used to extract the information. However, because of a higher recall value, the performance is really worse.

The efficiency of classification is evaluated by accuracy. It is assessed using Eq. (1):

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

In order to assess classification accuracy, F-measure uses recall and precision.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F-measure = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (4)$$

3. METHODOLOGY:

The various phases involved in proposed model comprises of:

a. Base video segmentation
b. Neutral, blur, reflection and color feature extraction
c. Classification using Naïve Bayes

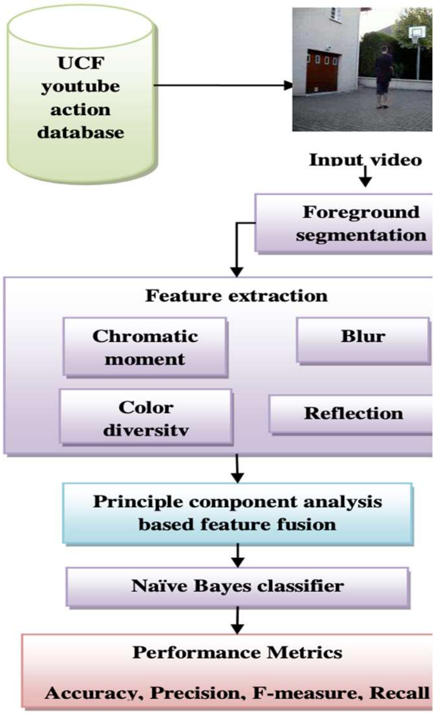

Fig 2: Proposed model

3a. Proposed model: It contains following steps:

1: Start the process.
2: Perform video groups *'V'*.
3: grouping the video set($V=v_1,v_2.....,v_n$)
4: While // if the probability value is greater than the extracted features.
5: Start v_1 of the video group's feature V.
6: pertinent features and original features are to be mapped.
7: If max $P(V\,q||Q\,c)$ then
8: Based on the user query, add a feature to the set of videos that were found.
9: skip unnecessary feature set.
10: End

When a user query founds a classifier, the key feature along with query feature has to be mapped. In this case, matching procedure uses the posterior probability, and a classifier must retrieve the video with greater probability. A Classifier must match features with a specific group in equation (5). The characteristic that corresponds to this group is obtained and extracted. The generated query shoul be identical to video cluster and defined in Eq.(6). A standard query cluster is enhanced in this case to determine matching correlation for video retrieval and this improved model is shown in Eq.(7).

$$V_q = [V_w||V_s||V_z] \tag{5}$$

$$y = arg \max_{Q\epsilon\{1,2,....,q\}} P(Q_c)P\ (V_q||Q_c) \tag{6}$$

$$CCR = arg \max_{Q\epsilon\{1,2,....,q\}} C(V_q,Q_c)P(V_q||Q_c) \tag{7}$$

$$P(Q_c^f) = \frac{|Q_c|}{|V_n|} \tag{8}$$

$$P(V_q||Q_c) = \frac{1}{\sqrt{2\pi\sigma_Q^2}} e^{(-v_z-\mu_Q)^2}/2\sigma_Q^2 \tag{9}$$

where Q = 1,2,...,q represents multiple clusters, CCR is the cluster correlation with the relevant features membership function, P is the probability of videos.

3b. Feature extraction:

Four distinct models—color, opacity, color diversity, and feature reflection were used for feature extraction. The reason behind selecting these features is that which will give more accuracy. To reduce data with various dimensions affecting classification accuracy, the mean values were taken into consideration for merging, and feature merging with principle component analysis (PCA) was carried out. A video's color is typically regarded as a significant distortion during both video recording and video playback. Because of the inadequate color pre-production qualities, this distortion can make it possible to tell the difference between false and live footage. In comparison to live video, printed video may generally have less contrast and color saturation. On-screen video, in comparison, displays greater contrast and brightness than live video. The color contrast between the false video and the live video is used in this instance to evaluate the video-based color distribution. The mean of the video, standard deviation and color distribution is evaluated by using below equations:

Journal of Theoretical and Applied Information Technology
15th December 2023. Vol.101. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$E_i = \frac{1}{N}\sum_{j=1}^{N} p_{ij} \qquad (10)$$

$$\sigma_i = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(p_{ij} - E_i)^2} \qquad (11)$$

$$s_i = \sqrt[3]{\frac{1}{N}\sum_{j=1}^{N}(p_{ij} - E_i)^3} \qquad (12)$$

A entire pixel ‘N’ is one where the value of the ith color channel's pixel is. These statistical moments are regarded as chromatic moment features for each channel. As a result, the color feature dimension is 3*3. The feature exhibit certain unique characteristics in local and smaller regions of films.

4. EXPERIMENTAL RESULTS:

A simulation was carried out in PYTHON by using,
Processor: Intel Core i7
operating system: 64-bit
RAM: 8 GB
Better classification outcomes are obtained using Bayes technique. The input video frame is shown in Fig. 3, it was obtained from the dataset and the foreground segmentation process is represented in Fig. 4.

Figure 3: image from input video

Figure 4: Foreground segmentation

Table 2: Feature values

Mean	Skewness	Blur	Reflection	Color diversity
7.289	40.171	0	0.067	6.683
36.540	86.823	2.504	0.096	32.14
17.808	62.940	0	0.069	22.82
43.524	94.206	2.431	0.095	31.585
47.050	96.615	1.809	0.089	41.414
15.282	58.47	4.86	0.09	15.01
7.3078	40.35	6.864	0.07	6.971
162.45	119.67	0.608	0.090	169.8
10.106	44.727	5.616	0.045	10.51
6.396	38.92	0	0.105	6.389
11.99	50.67	5.29	0.063	12.62

Table 3: Performance measures

Data size	Accuracy	F-Measure	Precision	Recall
FF with NB	98	99	98	99
Deep learning	98	97	98	96
Extensive feature set	-	60	6678	60
OCR	-	71	75	66
OCR and ASR	-	70	76	68

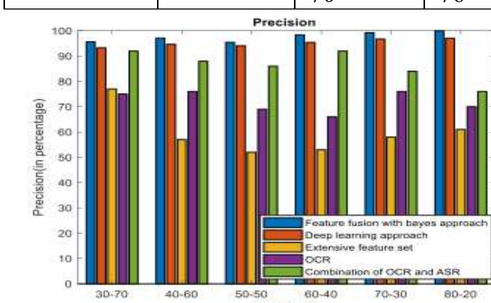

Figure 5: Precision

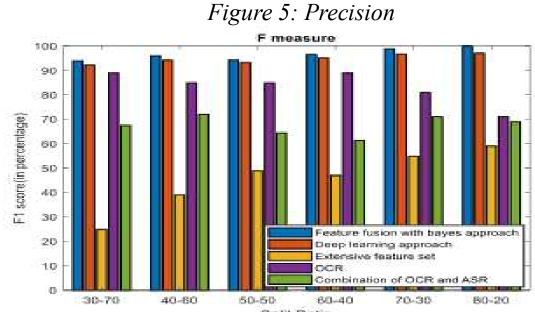

Figure 6: *F-Measure*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

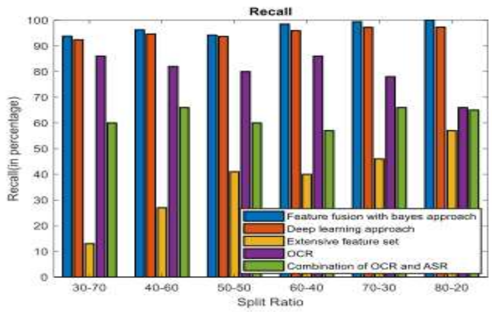

Figure 7: Recall

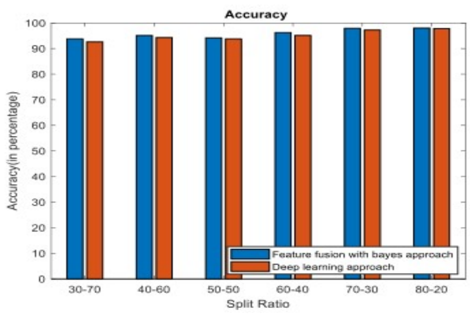

Figure 8: Accuracy

There are significant inter class and intra class variances in the database video. They are real-time environments that have been captured. The lighting, background, partial occlusion, and illumination are some of the current issues with this video collection. The videos are gathered in a variety of situations. The selection of an object region from the key frames is then done using foreground segmentation. Feature fusion NB classifier is used to extract and learn the features from frames and segmentation.

Recall and precision values are weighted averaged. Values range between 0 and 1, whereas 1 represents the high score and 0 represents less value. An accuracy of the suggested FF in comparison to NB and Deep learning is shown in Table III. The F-measure of FF with NB, DL, EFS, OCR, OCR with ACR is shown in Table 4.

5. CONCLUSION:

In this study, we used a novel approach for video recommendation. This approach uses low-level features to provide users with individualized recommendations. We have suggested a novel video retrieval from an untrimmed set of videos in this study.

The proposed approach gives more accuracy than all the existing systems.To obtain video from the accessible database, the suggested model takes Youtube action database is taken into account in this calculation. Foreground segmentation is done using video frames, then feature fusion is done next. Based on features collected from different parameters are combined. With linked videos, the likelihood that the element will be removed rises. For obtaining movies, a Nave Bayes classifier is utilized in the end. The videos in this instance are partitioned or clustered. To extract the functions, NB maps the correlation between the features.

6.FUTURE WORK:

This work done using core parameters. The accuracy may be varied if we use different parameters.

REFERENCES:

[1]. Yang, H.; Siebert, M.; Luhne, P.; Sack, H.; Meinel, C. Lecture video indexing and analysis using video ocr technology. In Proceedings of the 2011 Seventh International Conference on Signal Image Technology & Internet-Based Systems, Dijon, France, 28 November–1 December 2011

[2]. Accattoli, S.; Sernani, P.; Falcionelli, N.; Mekuria, D.N.; Dragoni, A.F. Violence detection in videos by combining 3D convolutional neural networks and support vector machines. Appl. Artif. Intell. **2020**, 34, 329–344.

[3]. Huang, J.; Rathod, V.; Sun, C.; Zhu, M.; Korattikara, A.; Fathi, A.; Fischer, I.; Wojna, Z.; Song, Y.; Guadarrama, S.; et al. Speed/accuracy trade-offs for modern convolutional object detectors. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7310–7311.

[4]. Robles, Oscar D, Pablo Toharia, Angel Rodríguez, and Luis Pastor. "Towards a content-based video retrieval system using wavelet-based signatures." In 7th IASTED International Conference on Computer Graphics and Imaging-CGIM, pp. 344-349. 2004.

[5]. Dipika H Patel, "Content-Based Video Retrieval using Enhanced Feature Extraction", International Journal of Computer Applications (0975-8887), (vol119),pp.4-8, 2015.

[6]. Dr.H.B.Kekre, Dr. Dhirendra Mishra, Ms. P. R. Rege, "Survey on Recent Technique in

Content Based Video Retrieval", International Journal of Engineering and Technical Research (IJETR), (vol3), pp.69-73,2015.

[7]. Muhammad Nabeel Asghar, Fiaz Hussain, Rob Manton,"Video Indexing: Survey", International Journal of computer an Information Technology,(vol3),pp.148-169,2014.

[8]. P.Geetha and Vasumathi Narayanan, "A survey of Content-Based Video Retrieval", Journal of Computer Science, (vol4), pp.474-486, 2008.

[9]. Avinash N Bhute and B B Meshram, "System Analysis and Design for Multimedia Retrieval Systems", The International Journal of Multimedia and Its Applications (IJMA), (vol5), pp.25-44, 2013.

[10]. Haitao Jiang, Abdelsalam (Sumi) Helal, Ahmed K. Elmagarmid, Anupam Joshi, "Scene change detection techniques for video database systems", ©Springer-Verlag, pp.186-195, 1998.

[11]. Dr. H.B. Kekre, Dr. Sudeep D. Thepade and Saurabh Gupta, "Content Based Video Retrieval in Transformed Domain using Fractional Coefficients", International Journal of Image Processing (IJIP), (vol7), 201.3

[12]. Chao, Jianshu, Anas Al-Nuaimi, Georg Schroth, and Eckehard Steinbach. "Performance comparison of various feature detector-descriptor combinations for content-based image retrieval with JPEG-encoded query images." In Multimedia Signal Processing (MMSP), 2013 IEEE 15th International Workshop on, pp. 029-034. IEEE, 2013.

[13]. Baris Sumengen, et.al. "Multi-Scale Edge Detection and Image Segmentation", IEEE Signal processing Conference, pp.1-4, 2005

[14]. N Ikonomakis, K. N. Plataniots and A.N. Venetsanopoulos,"A Region Based Color Image Segmentation Scheme", Proc.SPIE3653, Visual Communications and Image Processing , (vol3653), 1999.

[15]. Yu, Y.; Ko, H.; Choi, J.; Kim, G. End-to-End Concept Word Detection for Video Captioning, Retrieval, and Question Answering. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 3261–3269. [Google Scholar]

[16]. Avgoustinakis, P.; Kordopatis-Zilos, G.; Papadopoulos, S.; Symeonidis, A.L.; Kompatsiaris, I. Audio-based Near-Duplicate Video Retrieval with Audio Similarity Learning. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 5828–5835.

[17]. Yasira Beevi C P et.al.2009, "An efficient Video Segmentation Algorithm with real time Adaptive Threshold Technique ", International Journal of Signal Processing, Image Processing and Pattern Recognition (vol2), pp.13-27

[18]. Zhong Qu, Lidan Lin, Tengfei Gao and Yongkun Wang, "An Improved Keyframe Extraction Method Based on HSV Colour Space", Journal Of Software, (vol8), pp.1751-1758, 2013

[19]. Tjondronegoro, Dian, Xiaohui Tao, Johannes Sasongko, and Cher Han Lau. "Multi-modal summarization of key events and top players in sports tournament videos." In Applications of Computer Vision (WACV), 2011 IEEE Workshop on, pp. 471-478. IEEE, 2011.

[20]. Santhosh kumar K and Mallikarjuna Lingam.K, "Content Based Video Retrieval Using Low and High Level Semantic Gap", International Journal for Research in Emerging Science and Technology, (vol2), pp.149-153, 2015.

[21]. Shivanand S Gornale, Ashvini K Babaleshwar, Pravin L Yannawar," Detection and Classification of Signage's from Random Mobile Videos Using Local Binary Patterns", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.10, No.2, pp. 52-59, 2018.

[22]. L. Zelnik-Manor and M. Irani, "Event-based video analysis", IEEE Conference on Computer Vision and Pattern Recognition, CVPR'01, (vol2), pp.123–130, 2001.

[23]. Nikita Sao & Ravi Mishra, "Video Shot Boundary Detection Based On Nodal Analysis of Graph Theoritic Approach", International Journal of Management, Information Technology and Engineering (BEST: IJMITE), (vol2), pp.15-24, 2014.

[24]. Chandra Mani Sharma, Alok Kumar Singh Kushwaha, Rakesh Roshan, Rabins Porwal and Ashish Khare,"Intelligent Video Object Classification Scheme using Offline Feature Extraction and Machine Learning

based Approach", IJCSI International Journal of Computer Science Issues, (vol9), pp.247-256, 2012.

[25]. Hamdy K. Elminir, Mohamed Abu ElSoud, Sahar F. Sabbeh, Aya Gamal "Multi feature content based video retrieval using high level semantic concept IJCSI International Journal of Computer Science Issues, (vol9), pp.254-260, 2012.

[26]. Xingxiao Wu, Dong Xu, LixinDuan, JieboLuo, "Action Recognition Using MultilevelFeatures and Latent Structural SVM", IEEE Transactions on Circuits and Systems for Video Technology, (vol23), pp.1422-1431, 2013.

[27]. Padmakala, S., G. S. AnandhaMala, and M. Shalini. "An effective content based video retrieval utilizing texture, color and optimal key frame features."Image Information Processing (ICIIP), 2011 International Conference on. IEEE, 2011.

[28]. Navdeep Kaur, Mandeep Singh,"Content-Based Video Retrieval with Frequency domain Analysis using 2- D Correlation Algorithm", International Journal of Advanced Research in Computer Science and Software engineering ,(Vol4), pp.388-393,2014.

[29]. D.Asha, Madhavee Lata, V.S.K. Reddy, "Content based video retrieval system using Multiple Features", International Journal of pure and Applied Mathematics, (vol. 118), pp.287-294, 2018.

[30]. A. Paul, B.-W. Chen, K. Bharanitharan, and J.-F. Wang, "Video search and indexing with reinforcement agent for interactive multimedia services," ACM Trans. Embed. Comput. Syst, (vol12) pp. 25:1–25:16, 2013.

[31]. Benjamin Drayer and Thomas Brox, "Object detection, Tracking, and Motion Segmentation for Object-level Video Segmentation", Computer Vision and Pattern Recognition, pp:1-17 2016.

[32]. Santhosh kumar K and Mallikarjuna Lingam.K, "Content Based Video Retrieval Using Low and High Level Semantic Gap", International Journal for Research in Emerging Science and Technology, (vol2), pp.149-153, 2015.

[33]. Van Nguyen N, Ogier J-M, Charneau F. PEDIVHANDI: multimodal indexation and retrieval system for lecture videos. Lect Notes Comput Sci.2013; 7725:382–393.

[34]. Yang H, Siebert M, Luhne P, et al. 2011 Lecture video indexing and analysis using video OCR technology. Proceedings of IEEE Seventh International Conference on Signal video Technology and Internet-based Systems (SITIS), Dijon, France; 2011. p. 54–61.

[35]. Li K, Wang J, Wang H, et al. Structuring lecture videos by automatic projection screen localization and analysis. IEEE Trans Pattern Anal Mach Intell. 2015; 37(5):1233–1246.

[36]. Baidya E, Goel S. 2014 Lecturekhoj: automatic tagging and semantic segmentation of online lecture videos. Proceedings of IEEE International Conference on Contemporary Computing (IC3), Noida, India; 2014. P.37–43.

[37]. W. Wang, J. Shen, and F. Porikli, "Saliency-aware geodesic video object segmentation," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2015, pp. 3395–3402.