WEB BASED AUDIO-VIDEO RETRIEVAL ALGORITHM USING SUPPORT VECTOR MACHINE CLASSIFIER

N. Srinivas*
Research Scholar, Dept of Computer Science & Engineering, Rayalaseema University, Kurnool, A.P.
Dr.M.V. Ramana Murthy
Retired Prof, Mathematics & Computer Science, Osmania University, Hyderabad.
Professor, MGIT, Gandipet, Hyderabad. E-mail:mv.rm50@gmail.com

**Corresponding author E-mail:srinivas.bhaskar3@gmail.com*

ABSTRACT: Video data retrieval proves to be a challenging and vital problem. Software's are utilized to model, imitate and more significantly surpass human vision with the help of unlike software technology. There is an exponential rise in the digital content that is being produced quickly, mainly videos. In the existing video summarization is done through via the Audio Video Information Summarization Algorithm (AVISA) method. However web based video retrieval is not to be discussed. In proposed Audio Video Retrieval algorithm (AVRA) is necessary for efficient search of these media. The video summarization the web based data's are to be in indexing part. After that indexing using of SVM classifier to retrieve video from database based on query input from the user. To address this issue an AVRA framework is developed using MATLAB.

Keywords: Audio Video Retrieval algorithm (AVRA), Audio Video Information Summarization Algorithm (AVISA), Support Vector Machine (SVM).

1. INTRODUCTION

In this electronic world enormous volume of helpful digital info such as audio, images and video data despite textual data subsists online and is accessible to public, professionals, government authorities and researchers extremely simple and accessible at sensibly low-cost because of speedy development in accessibility of user friendly as well as inexpensive multimedia acquisition devices at a large scale such as high resolution camera in handy cams, mobile phones, and other modern digital devices, accessibility of greater capacity storage devices such as hard disks, memory cards, and so on., large-scale use of internet by quickly rising amount of applications utilized by digital devices to upload enormous volume of multimedia info, modern web technology as well as internet infrastructure [1]. Video data contains numerous info for those utilizing multimedia systems and applications such as publications, broadcasting, digital libraries, education, and entertainment [2]. These applications are helpful only while video retrieval systems are adequate enough to take out videos and other significant info from huge databases as rapid as probable [3]. On the other hand, it is thought-provoking for the previous web search engines to look for video over the web as a result new methods are needed that are able to manipulate the video info in keeping with the content [4]. For multimedia mining, sets of multimedia data are maintained as well as organized with the help of the methods such as classification and annotation of videos [5].

Numerous web based video retrieval systems work by indexing as well as searching videos dependent upon texts related to them on the other hand this method doesn't do well as the texts don't comprise sufficient info of the videos. As video retrieval is not efficient utilizing traditional query-by-text retrieval method, Audio Video Retrieval (AVR) is taken as one among the finest practical solutions for improved retrieval quality. Because of use of rich video content, there is a remarkable scope in region of video retrieval for improving the performance of traditional search engines. This is directing the region of AVR into a direction likely to produce more efficient video search engines in future [6].

In this rest of the paper is organised as follows. Section 2 discuss about the related works for the proposed methodologies. Section 3 discuss about the out proposed method AVRA with query performance, indexing and retrieval scheme. Section 4 discuss about the result and implementation. Section 5 discuss about the conclusion and future work.

2. LITERATURE REVIEW

Brindha et al., [7] presented the Content-based video retrieval system focuses on supporting a user to take out targeted video series in a big database. In order to retrieve videos, numerous search engines utilize textual annotations. These kinds of engines give a low-level abstraction when the user searches for high-level semantics. Linking this kind of semantic gap in video retrieval is a significant problem. In this research, texture, colour and shapes are taken to be low-level features as well as motion is a high-level feature. Colour histograms transform the RGB colour space into YcbCr and take out hue as well as saturation values from frames. Filter mask is used and gradient value is calculated subsequently colour extraction. Gradient as well as threshold values are matched up to draw the edge map. In order to eliminate the needless connected components, edges are smoothed for sharpening. These different shapes are taken out as well as maintained in shape feature vectors. Lastly, an SVM classifier is utilized for classification of low-level features. For high-level features, depth images are taken out for motion feature identification and by means of echo state neural networks (ESN), classification is carried out. ESN are a supervised learning method and keep an eye on the principle of recurrent neural networks. ESN are familiar for time series classification and as well shown their efficient performance in gesture detection.

Dyana et al., [8] have proposed a Gabor filter based representation of motion trajectory, for the motion based video retrieval. They have proposed a spatio-temporal representation of the trajectory, which is concerned about the process of identifying a set of salient points from the peaks (locally) of the Gabor filter responses. The feature set (formed by the frequency, temporal location and turning direction at each salient point) has provided a semantic representation of the trajectory. Their approach has been a global trajectory representation where matching is carried out on the basis of the edit distance and has been revealed to carry out well even for partial trajectory matching.

Zawbaa et al., [9] presented Machine Learning (ML) based event detection as well as summarization system for highlighting significant events for the period of soccer matches. The research method primarily segments the complete video stream into small video shots, and after that it categorizes the ensued shots into diverse shot-type classes. Subsequently, the system uses two machine learning algorithms, called; Support Vector Machine (SVM) and Artificial Neural Network (ANN), for highlighting significant segments with logo appearance and identifying the caption region offering info regarding the score of the game. Then, the system identifies vertical goal posts in addition to goal net. Lastly, the most significant events for the period of the match are emphasized in the ensued soccer video summary. The research technique significantly decreases workload and improves the precision of summarizing soccer video matches regarding recall as well as precision performance measurement criteria.

Bashir et al., [10] provides a new motion trajectory-based compact indexing as well as efficient retrieval technique for video sequences. Supposing trajectory info is already existing, we denote trajectories as temporal ordering of subtrajectories. This method resolves the issue of trajectory depiction while merely partial trajectory info is there because of obstruction. It is attained by a hypothesis testing-based technique used to curvature data calculated from trajectories. The subtrajectories are denoted by their principal component analysis (PCA) coefficients for ideally compact depiction. Diverse methods are incorporated to index as well as retrieve subtrajectories, comprising PCA, spectral clustering, and string matching. We take up a query by example mechanism where an example trajectory is provided to the system as well as the search system yields a ranked list of most alike items in the dataset.

Borkar et al., [11] proposed method of Content Based Video Retrieval (CBVR) the Ordered-Dither Block Truncation Coding (ODBTC) technique is employed which generates appropriate image contents. Combinations of Void-and-cluster half-toning and Block Truncation Coding (BTC) offers low complexity in algorithm and provides better video image quality. Dither array Look-Up-Table (LUT) is a distinctive feature of ODBTC which reduces the difficulties by providing look up values of segmented blocks. ODBTC encoded streams are used for generation of two distinct features including of color features namely Color Co-occurrence Feature (CCF) and Bit Pattern Features (BPF). After quantizing and bit-mapping from ODBTC encoder, BPF is obtained by LUT. In the presented system, CBVR is achieved by Block Truncation (BT) of expected video information to be retrieved. Proposed system provides good remedy for CBVR for large digital video-data processing in the fields of Image and Video Processing.

Patel et al., [12] have described about the Video Retrieval Based on Textual Queries presented an approach that enables search based on the textual information present in the video. Regions of textual information are indented within the frames of the video. Video is then annotated with the textual content present in the images. A learning framework presented an automatic Content-Based Retrieval and Classification of Video Content where construction of a high-level video index is visualized through the synthesis of its set of elemental features. This is done through the medium of support vector machines (SVM) that relate every set of data points in the multidimensional feature space to one among the classes in the course of training. In Content-Based TV Sports Video Retrieve Based on Audio - Visual Features and Text. Here are some techniques utilized for feature extraction as well as its retrieval application.

Bibi et al., [13] proposed a novel approach for video retrieval in augmented reality based on image queries. At first, it extracts the key frames from the videos. Secondly, it employ a novel frame based feature extraction method, namely Ternary Histogram of Oriented Gradient (THOG). Thirdly, it utilize the Double-Bit Quantization based hashing to perform the nearest neighbor search efficiently, which is responsible to generate the candidate list of videos. Finally, the similarity measure is performed to re-rank the list of videos from the candidate list.

3. PROPOSED METHODOLOGY

In this work proposed a generic Retrieval algorithm by using both audio and visual information. The proposed Audio-Video Retrieval Algorithm (AVRA) consists of three major steps. The first step is responsible for automatic structuring of videos. It majorly focuses on the extraction of Video Key Frame Indexing, Texture, Color, audio features, etc and gets the summarization of the video used for AVRA. In the second step the summarization of the video is to be stored in the database is known as indexing it also gets from the feature extraction terms. In the third step gets the query result from the user for the retrieval using the classification algorithm it is known as SVM.

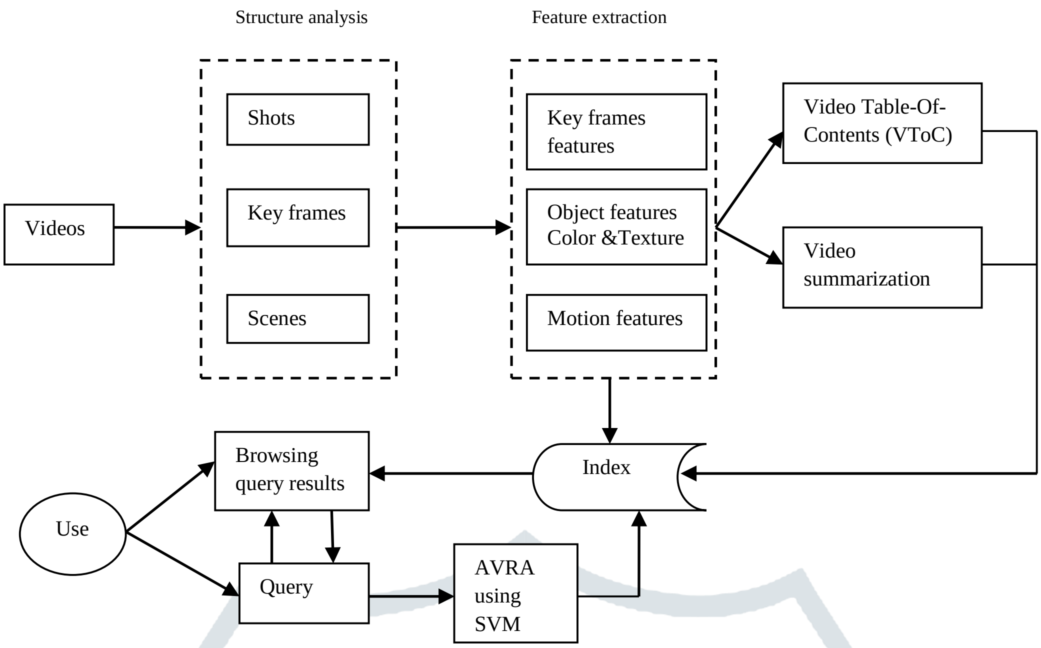

Figure 1. Overall Structure of AVRA System

3.1. Structure Analysis

In the video structuring process, different video features, for example, audio energy, color histograms, shape, texture and video caption text, are utilized to produce the video components. These components are discussed in the previous section so short description is here. Video is a structured medium wherein actions as well as events take place in time to make a story. A video must therefore be regarded as a document and not as an unstructured sequence of frames. Before the video data into the database it must be integrated according to its characteristics be structured. Videos can be composed of different streams be, have one or more audio tracks and, if necessary, subtitles and other text overlays. All this information is in the structural analysis significant. It should be noted that the video itself in Scenes and camera settings, so-called shots. Shots can be used as syntactic units; scenes are regarded as semantic units.

Video Shot - It is an unbroken sequence of frames recorded from a single camera. It is the building block of a video.

Video Scene - It is defined as a collection of shots related in the video content and temporally adjacent. It depicts and conveys the concept or story of video.

Key Frame - It is the frame which can represent the salient content of a shot.

3.2. Feature Extraction

Classification as well as retrieval visual features embedded in video data is used for efficient video indexing. Three key features to be taken out are texture, color and motion for efficient video indexing. These features are denoted by color histogram, Gabor texture features and motion histogram correspondingly [14]. The most beneficial info in the videos comprises key frames, features of the objects and the motion features.

A). Key Frame Features: It encompasses texture, color and shape based static features. These are most important visual properties and are basic concerns in low level image as well as computer vision problems. Numerous color features are color histograms, color moments, color correlograms attained from some Gaussian models.

B). Object Features: It comprise the texture, dominant color, size, and so on., of the image areas relating to the objects. These features are utilize to take out videos probably to comprise alike objects.

1) Color-Based Features: The image denoted in this research is a static image form vision sensors (webcam). The image is a continuous function of light intensity on two-dimensional field. An image must be provied numerically with discrete values so as to be processed by a computer. A digital image is denoted by a two-dimensional matrix f (x, y) encompassing M columns and N rows. There are numerous models one among which is the hue, saturation, value (HSV) model in color image processing. With the help of this model, an object with some color could be identified as well as to decrease the effect of light intensity from the outside. Tests were carried out by six types of colors, ie yellow, black, green, brown, blue, and white.

Digital image processing utilizing computer vision technology is now extensively utilized as a research object. Portion of image processing is to utilize color based processing. Color analysis in the introduction of digital imagery HSV as well as normalized HSV model. One form of use of HSV model is as recognition. Utilizing this model as facial recognition contains the benefit of being modest in programming; the process is quick as a result it is accurate.

The color feature is an 8 × 4 2D color histogram in HSV color space. The V component is not used because of its sensitivity to lighting conditions. The H component is quantized finer than the S component due to the psychological observation that the human visual system is more sensitive to Hue than to Saturation. Each hue is organized in a radial slice, around a central axis of neutral colors that ranges from black at the bottommost to white at the topmost. The HSV representation models the way paints of diverse colors combine together, with the saturation dimension similar to numerous shades of luminously colored paint, and the value dimension similar to the combination of those paints with variable amounts of black or white paint.

The frame is a continuous function (continuity) with the intensity of light in the field of two dimensions. An image should be provided with discrete values so as to be processed with a digital computer. Frame digitization is the depiction of a continuous function into discrete values. A digital frame is denoted by a two dimensional matrix f (x, y) encompassing M columns and N lines, where intersections amid columns and rows are known as pixels (picture element) or the least element of a frame.

2) Texture-Based Features: The Gray Level Co-occurrence Matrix1 (GLCM) and associated texture feature calculations are image analysis techniques. Given an image composed of pixels each with an intensity (a specific gray level), the GLCM is a table of how frequently diverse sets of gray levels co-occur in an image or image section. This feature discussed in the previous method.

C). Motion Features: Motion is the basic characteristic differentiating dynamic videos from still images. Motion info denotes the visual content with temporal variation. Motion features are nearer to semantic notions compared to static key frame features as well as object features. Video motion comprises background motion produced by camera motion as well as foreground motion produced by moving objects. Therefore, motion-based features for video retrieval is split into two types: camera-based and object-based. For camera-based features, diverse camera motions, for instance “zooming in or out,” “panning left or right,” and “tilting up or down,” are approximated and utilized for video indexing. Video retrieval utilizing only camera-based features contains the limitation that they could not define motions of key objects.

The Trajectory-based motion modelling approach for human action recognition is represented. The first describe a method that utilizes global reference points to cancel camera motion specifically for improving the TrajShape descriptor. After that introduce a trajectory-based motion representation which uses each trajectory as a local reference point. This depiction includes location as well as motion relationships of the patch trajectories as well as their local descriptors, and is not subtle to camera motion.

The clustering project precedes a bottom up manner by merging two closest curves. Finally object appearance matching process is performed, using appearance to look for similar objects in a video, in terms of their occurrences. Speed up robust feature can use for object appearance matching. The whole system is separated by two main stages one is index stage and search stage. At the indexing stage, it is also a processing stage, where detect and extract moving objects, smoothen their moving trajectories and construct indices accordingly. At the search stage, it is to deal with a user’s requests at run time, where the inputs could be trajectories or object’s appearances. In Indexing stage first step is back ground subtraction; to achieve this all the frames are median filtered along the time direction to find the background of the video, each frame is compared against the background in terms of CIE 94 color space to form a Gaussian mixture model, and the optical flows of their intensity values are used.

The basic idea is to measure how many characters are needed to convert from one string into another through the operations of insertion, deletion and change. It improves the accuracy and performance in searching stage at run time. There is a possibility of object appearance in numerous forms in video due to different angle, and similar object presented in one scene. Here the tracking result have some erroneous. To solve this problem back of words feature can use, hierarchically cluster the features to form a vocabulary tree. The limitation of this project is during the object tracking, the object may be affected by lightening or other environmental noise it can produce the wrong trajectory file. Based on this trajectory may create an inappropriate result.

3.3. **Query of a Video**

As soon as video indices are got, video retrieval is carried out. On getting a query, a similarity measure is utilized, dependent upon the indices, to look the candidate videos according to the query. The retrieval outcomes are getting by the feedback, etc. In the following, query of a video retrieved based on query by object, query by text, query by example, Query by shot, Query by clip and Query by Faces & Texts these are all some retrieval process from the users.

Query by Object: The object frame is provided. The incidences of objects in video database are identified and places of the object identify success of the query [15].

Query by Text: Since it is familiar for content based video retrieval, example frames are utilized as query to take out appropriate videos in a database of videos (query by example) on the other hand it contains a restraint that motion info of the video being looked for is not used. It depends upon the appearance info. As well, identifying video clip for the fascinated conception could turn out to be very difficult utilizing example frame. Textual query provides additional natural interface and claims to be improved method for querying in video databases [16].

Query by Example: Query by example is superior when visual features of the query are utilized for video retrieval. Low level features are attained from key frames of the query video and they are matched up to segregate the alike videos utilizing their key frames visual features.

Query by shot: Certain systems use the complete video shot since the query rather than key frames. This is a healthier choice on the other hand with a greater computational cost.

Query by clip: A clip is utilized for improved performance of video retrieval as matched up with the method while a shot is utilized as a shot don't denotes adequate info regarding the complete context. The entire clips that contains a substantial similarity with the query clip are taken out.

Query by Faces and Texts: Faces as well as Texts could be utilized as a query to take out a video segment comprising frames labelled for a particular kind in keeping with faces as well as texts. An appropriate method is utilized to look for the video queried by the query clip utilizing info got from faces as well as texts in frames of the query clip.

3.4. AUDIO-VIDEO INDEX AND RETRIEVAL

Video summarization [17] eliminates the redundant data in videos and creates an abstract depiction or summary of the contents that is revealed to users in a clear manner to enable browsing. Video summarization complements video retrieval [18], by making browsing of taken out videos sooner, particularly while the over-all size of the retrieved videos is huge: The user could browse via the abstract depictions to place the preferred videos. A thorough review on video browsing interfaces as well as applications could be found in [19].

Video retrieval is concerned with how to return similar video clips (or scenes, shots, and frames) to a user given a video query. One is to first extract key frames from the video data, then use image retrieval techniques to obtain the video data indirectly. Although easy to implement, it has the obvious problem of losing the temporal dimension. The other technique incorporates motion information (sometimes object tracking) into the retrieval process. Although this is a better technique, it requires the computationally expensive task of motion analysis. If object trajectories are to be supported, then this becomes more difficult. Here we view video retrieval from a different angle. We seek to construct a video Index to suit various users' needs. However, constructing a video Index is far more complex than constructing an index for books. For books, the form of an index is fixed (e.g., key words). For videos, the viewer's interests may cover a wide range. Depending on his or her knowledge and profession, the viewer may be interested in semantic level labels (building, car, people), low level visual features (color, texture, shape), or the camera motion effects (pan, zoom, rotation). In the system described here, we support the following three Index categories:

- Visual Index
- Semantic Index
- Camera motion Index

The techniques that should be apply on video data for its searching and retrieval, one among the most significant steps is to transform video from unstructured formatted data sets to the formatted data set. This video has one complex type as changing stream of images that serves as dynamic feature extraction and retrieval of video contents. Video as a whole is very large data to mine that creates the need of some processing to get data in the suitable format for searching. Video data is consisting of spatial, temporal data, images, audio clips, etc. features. Based on applications requirement these features are used for searching and retrieval. Commonly, video is the hierarchical construction, having number of frames, streams, segments, scenes, clips and full length video. Every video clips unit has its own features which are useful for getting particular useful video clips.

The query specification component translates user input data to system for extraction of feature content description and temporal structure. In current video systems interaction based on video content is done using key frames. Hence, interactive video retrieval can be classified similar to image retrieval. Even though there are no fundamental differences between image retrieval and shot retrieval with respect to the interaction classification, in practice different accents are found in applications for the two media by abstraction. The users input consist of facts in predefined fixed format and query by pictorial example. The retrieval data has to be classified by the SVM classifier to get the matched video queried by the users.

3.4.1. THE PRINCIPLE OF SUPPORT VECTOR MACHINES

Support vector machine is the most successful machine learning technology in pattern recognition and computer vision, which is based on statistical theory. Theoretically the optimal results could be got by SVM as for samples could be separated linearly; in the cases samples could not be separated linearly, it takes advantage of kernel transformation and could obtain satisfactory results. The basic idea of is as following [20].

Assumed that samples$(x_i, y_i), i = 1, \dots, l$, $x \epsilon R^d$, $y_i \epsilon \{-1,1\}$ could be distinguished linearly. SVM is devoted to finding out the maximum-margin hyperplane that differentiates samples belonging to different categories. The optimal hyper-plane equation may be established as (1).

$$\omega . x + b = 0 \qquad (1)$$

In practice, samples could be separated linearly. Through a nonlinear mapping $\varphi(.)$, SVM will solve [20]:

$$\frac{1}{2}||w||^2 + C\sum_{i=1}^{m} \xi_i \qquad (2)$$

Eq.(2) is subject to (3) as below.

$$y_i\big(\varphi(x_i)\big).w + b) \geq 1 - \xi_i, \qquad \xi_i \geq 0; i = 1,2, \dots, l \qquad (3)$$

In (2-3) w and b specify a maximum margin linear classifier in feature space and the variables ξ_i are the alleged slack variables. Parameter C is a positive and regularization parameter set by the user, which keeps balance of model complexity and training error. Because it is hard to handle the inequality constraints directly, the

Lagrange theory is adopted by presenting Lagrange multipliers for the quadratic optimization problem, the subsequent dual representation comes into being [21].

$$max[W(a)] = \sum_{i=1}^{l} a_i - \frac{1}{2}\sum_{i-1}^{l}\sum_{j-1}^{l} a_i\, a_j\, y_i y_j k(x_i, x_j) \quad (4)$$

Subject to: $0 \leq a_i \leq C$, and $\sum_{i-1}^{l} a_i\, y_i = 0$ (5)

It is shown that minimization of the first term in (3) is equal to minimizing the VC-dimension, and minimization of the second term amounts to minimizing the wrong classification error. The above minimization problem could be viewed as a constrained quadratic programming problem. The solution induces a decision equation of (6) below.

$$f(x) = sgn\left[\sum_{i-1}^{l} a_i\, y_i k(x_i, x_j) + b\right] \quad (6)$$

In (6), the function k(.,.) is formulated as (7)

$$k(x, x_i) = \langle \varphi(x), \varphi(x_i) \rangle \quad (7)$$

In general, only a small part of the ai coefficients are nonzero. The corresponding pairs of xi entries are labeled as support vectors, which decide the decision function. $k(x, x_i)$ is the equivalent nonlinear kernel function, RBF kernel function is the commonly used one, which is defined as (8) below.

$$k(x, x_i) = exp\left(\frac{-||x_i - x_j||^2}{\sigma^2}\right) \quad (8)$$

The parameter σ in kernel function denotes the characters of training data and has significant impact on the performance of SVM classifier. In sum, parameters C and σ decide the performance of SVM, which remains a problem to be solved. In this classifier gives the matched audio-video retrieval data based on users feedback results.

4. RESULTS AND DISCUSSION

In this section we present results obtained from analyzing mechanism with videos. Various video types are analyzed (news, talk shows, music, movies, advertisement, sport events and cartoons) in order to test proposal in speech based videos (movies, music and talk shows), high action videos (shows, cartoons and movies) and simple motion videos (shows, movies and sport events). Each video has been tested in its original length. For obtaining results, MATLAB software is used in this work. MATLAB is used for programming the algorithms .The recorded sounds were read in MATLAB, using wavered command. And then various algorithms were applied on these signals to discriminate voice and unvoiced segments of the speech. In the results and discussion section, three techniques such as proposed AVRA system, existing Ordered-Dither Block Truncation Coding (ODBTC) [11] and Ternary Histogram of Oriented Gradient (THOG) [13] results are compared. The entire approaches implemented in this research work are assessed on an extensive range of thresholds. For every set of algorithm and threshold set, we compute the number of boundaries appropriately identified, no of false positive and no of missed boundaries. It is a general means to select Recall and Precision as the suitable evaluation condition.

Recall = (correct) / (correct + missed) (9)

Precision = (correct) / (correct + false positive) (10)

F Measure is defined as the harmonic mean of Precision (P) and recall (R)

F-measure =2. (P.R)/(P+R) (11)

In numerous applications, a tradeoff should be made amid the recall and precision. It could or could not be suitable to take out additional shot boundaries that would or else be missed at the cost of taking out of each system.

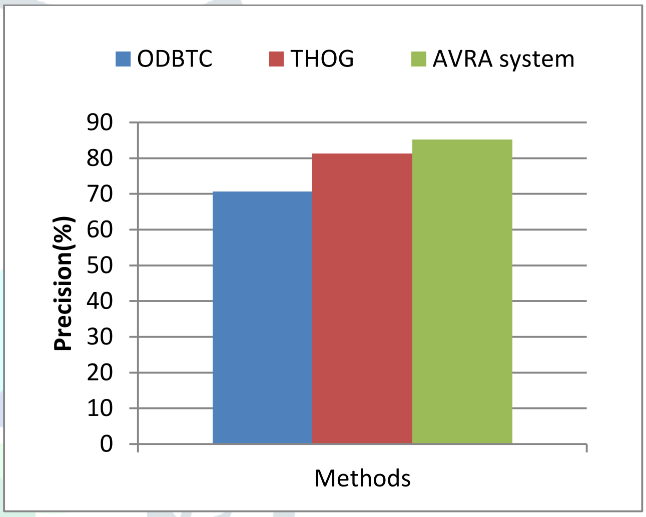

Figure 2. Precision comparison vs. video retrieval methods

In this graphical representation gives the improved results in precision parameters using proposed method compare than the existing methods. In proposed method AVRA System gives the 85.2% of the result that is shown in the figure 2.

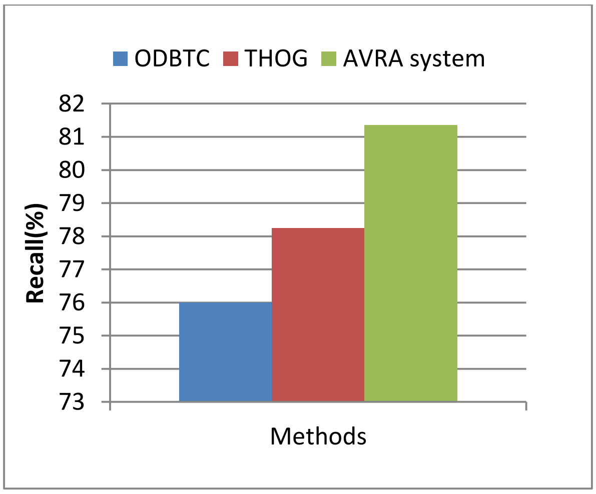

Figure 3. Recall comparison vs. video retrieval methods

In this graphical representation gives the better performance in Recall parameters using proposed method compare than the existing methods. As shown in the figure 3, proposed method AVRA System gives the 81% of the result.

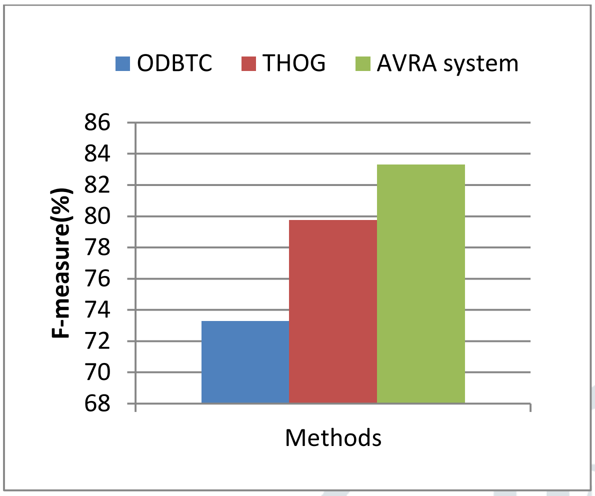

Figure 4. F-measure comparison vs. video retrieval methods

In this graphical representation gives the best analysis in F-Measure metrics using proposed method compare than the existing methods. In proposed method AVRA System gives the 83.28% of the result that is shown in the figure 4.

5. CONCLUSION AND FUTURE WORK

In this paper, we introduce an automatic video indexing and retrieval algorithm for a web based video processing. The retrieval algorithm uses MPEG-7 visual descriptors to produce a video index for retrieval. In this work proposed a generic Retrieval algorithm by using both audio and visual information. The proposed Audio-Video Retrieval Algorithm (AVRA) system consists of three major steps. The first step is responsible for automatic structuring of videos. In the second step the summarization of the video is to be stored in the database is known as indexing it also gets from the feature extraction terms. In the third step gets the query result from the user for the retrieval using the classification algorithm it is known as SVM. Moreover, classification technique is utilized to web based video processing retrieval and indexing in stored video data. Then, AVRA system produces the retrieval for comparing video pairs. The experiments confirm the effectiveness of AVRA system for various query types. In future work will focus on the appropriate preference values for users in retrieval of the video frames.

REFERENCES

1. Ja-Hwung Su, Yu-Ting Huang, Hsin-Ho Yeh, Vincent S. Tseng , "Expert Systems with Applications", 37, pg 5068-5085, 2010 .
2. Shanmugam, T.N. and Rajendran, P., 2009. An enhanced content-based video retrieval system based on query clip. International Journal of Research and Reviews in Applied Sciences, 1(3), pp.236-253.
3. Chen, L.H., Chin, K.H. and Liao, H.Y., 2008, An integrated approach to video retrieval. In Proceedings of the nineteenth conference on Australasian database-Volume 75(pp. 49-55). Australian Computer Society, Inc..
4. Patel, B.V. and Meshram, B.B., 2012. Content based video retrieval. arXiv preprint arXiv:1211.4683.
5. Tseng, V.S., Su, J.H., Huang, J.H. and Chen, C.J., 2008. Integrated mining of visual features, speech features, and frequent patterns for semantic video annotation. IEEE Transactions on Multimedia, 10(2), pp.260-267.
6. Hoi, S.C. and Lyu, M.R., 2007, A multimodal and multilevel ranking framework for content-based video retrieval. IEEE International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. Vol. 4, pp. IV-1225.
7. Brindha, N. and Visalakshi, P., 2017. Bridging semantic gap between high-level and low-level features in content-based video retrieval using multi-stage ESN–SVM classifier. Sādhanā, 42(1), pp.1-10.
8. Dyana, A. and Das, S., 2009. Trajectory representation using Gabor features for motion-based video retrieval. Pattern Recognition Letters, 30(10), pp.877-892.
9. Zawbaa H M, El-Bendary N, Ella Hassanien A, Kim T (2012) Event detection based approach for soccer video summarization using machine learning. International Journal of Multimedia and Ubiquitous Engineering, Vol. 7, no. 2, pp 63-80.
10. Bashir, F. I., Khokhar, A. A., & Schonfeld, D. (2007). Real-time motion trajectory-based indexing and retrieval of video sequences. IEEE Transactions on Multimedia, 9(1), 58-65.
11. Borkar, S.V. and Katariya, S.S., 2017, Content based video retrieval scheme using halftone block truncation coding. 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), pp. 1089-1094.
12. B V Patel, B B Meshram, “Content Based Video Retrieval Systems”, International Journal of UbiComp (IJU), Vol.3, No.2, 2012.
13. bibi Joolee, J. and Lee, Y.K., 2018, Video Retrieval Based on Image Queries Using THOG for Augmented Reality Environments. International Conference on Big Data and Smart Computing (BigComp), pp. 557-560.
14. Liapis, S. and Tziritas, G., 2004. Color and texture image retrieval using chromaticity histograms and wavelet frames. IEEE Transactions on multimedia, 6(5), pp.676-686.
15. J. Sivic and A. Zisserman, “Video Google: Efficient visual search of videos,” in Toward Category-Level Object Recognition.. Berlin, Germany: Springer, 2006, pp. 127–144.
16. Bennett, I.M., Phoenix Solutions Inc, 2009. System and method for natural language processing of sentence based queries. U.S. Patent 7,624,007.
17. B. T. Truong and S. Venkatesh, “Video abstraction: A systematic review and classification,” ACM Trans. Multimedia Comput., Commun. Appl., vol. 3, no. 1, art. 3, pp. 1–37, 2007.
18. Z. Xiong, X. S. Zhou, Q. Tian, Y. Rui, and T. S. Huang, “Semantic retrieval of videoReview of

research on video retrieval in meetings, movies and broadcast news, and sports," IEEE Signal Process. Mag., vol. 23, no. 2, pp. 18–27, 2006.

19. K. Schoeffmann, F. Hopfgartner, O. Marques, L. Boeszoermenyi, and J. M. Jose, "Video browsing interfaces and applications: A review," SPIE Rev., vol. 1, no. 1, pp. 018004.1–018004.35, 2010.
20. U. Maulik, D. Chakraborty, "A novel semisupervised SVM for pixel classification of remote sensing imagery," International Journal of Machine Learning & Cybernetics, vol. 3, no. 3, 2012, pp. 247-258.
21. Z.-Y. Jia, J.-W. Ma, F.-J. Wang, W. Liu, "Hybrid of simulated annealing and SVM for hydraulic valve characteristics prediction," Expert Systems with Applications, vol. 38, no. 7, 2011, pp. 8030-8036.