July 2022

CONTENT BASED RETRIEVAL OF LECTURE VIDEO REPOSITORY: LITERATURE REVIEW

Arun Adrakatti
arun@iiserbpr.ac.in

K R Mulla Dr.
Visvesvaraya Technological University, Belagavi

CONTENT BASED RETRIEVAL OF LECTURE VIDEO REPOSITORY: LITERATURE REVIEW

Arun F. Adrakatti
Asst. Librarian, Indian Institute of Science Education Research (IISER), Berhampur, Odisha, INDIA
Email: arun@iiserbpr.ac.in

K.R. Mulla
Librarian, Visvesvaraya Technological University, Belagavi, Karnataka, INDIA
Email: krmulla@vtu.ac.in

Abstract:

Multimedia has a significant role in communicating the information and a large amount of multimedia repositories make the browsing, retrieval and delivery of video contents. For higher education, using video as a tool for learning and teaching through multimedia application is a considerable promise. Many universities adopt educational systems where the teacher lecture is video recorded and the video lecture is made available to students with minimum post-processing effort. Since each video may cover many subjects, it is critical for an e-Learning environment to have content-based video searching capabilities to meet diverse individual learning needs. The present paper reviewed 120+ core research article on the content based retrieval of the lecture video repositories hosted on cloud by government academic and research organization of India.

Keywords: CBVR, Video Retrieval, Multimedia Information Retrieval, MIR, Video Repository, Institutional Repository.

1. INTRODUCTION

For thousands of years people have realized the importance of archiving and finding information. With the advent of computers, it became possible to store large amounts of information; and finding useful information from such collections became a necessity. The field of Information Retrieval (IR) was born in the 1950s out of this necessity. Over the last forty years, the field has matured considerably. Several IR systems are used on an everyday basis by a wide variety of users. Information retrieval is become a important research area in the field of computer science. Information retrieval (IR) is generally

concerned with the searching and retrieving of knowledge-based information from database.

Semantic based video retrieval is a most critical issue in the multimedia search engine related research. Multimedia content annotation improves the accuracy of semantic based content retrieval system where multimedia content annotation can be done in two ways: One is content based annotation and another is context based annotation. The content based annotation of images and videos consist of both low level and high level features which would be derived from the detailed pixel intensity information of images whereas context information would provide the semantic details about the images/videos.

The increased in availability and usage of on-line digital video has created a need of automated video content analysis techniques, including indexing and retrieving. Automation of indexing significantly reduces the processing cost while by minimizing tedious work. Traditional video retrieval methods based on video metadata, fail to meet technical challenges due to large and rapid growth of multimedia data, demanding effective retrieval systems. One of the most popular solutions for indexing is extracting the features of video key frames for developing a Content Based Video Retrieval (CBVR) system. CBVR works more effectively as these deals with content of video rather than video metadata. Various features like color, texture, shape can be integrated and used for video indexing and retrieval.

2. MULTIMEDIA INFORMATION RETRIEVAL:

Research on multimedia information retrieval (MIR) has recently witnessed a booming interest. Content-based multimedia information retrieval (MIR) has become one of the most active research areas in the past few years. While these approaches establish the viability of MIR based on visual features, techniques for incorporating human expertise directly during the query process to improve retrieval performance have not drawn enough attention. In their widely acclaimed work, (Rui et al., 1997) discuss the limitation and introduces a human-computer interaction based approach to MIR in which the user guides the system during retrieval using relevance feedback.

(Kankanhalli & Rui, 2008) discussed existing impact of multimedia information retrieval (MIR) in applications trends in the applications which can inform the MIR community on

directing intellectual resources towards MIR problems which can have a maximal real-world impact.

The needs and expectations regarding multimedia content access have grown rapidly with the fast development of multimedia technology and the explosion of multimedia content around us. This imposed high demands on the level of sophistication of multimedia information retrieval (MIR) solutions. Although the potential to develop the MIR technology that meets such high demands has also rapidly grown over the years, we are not there yet with adequate solutions.

(Wan & Liu, 2008) discusses the applications and importance of content-based information retrieval technology in digital libraries. The content-based information retrieval will be a significant trend for the development of digital libraries. (Wilkins, 2009) Content Based Multimedia Information Retrieval (CBMIR) is characterized by the combination of noisy sources of information which, in able to achieve strong performance. In this thesis we focus on\nthe combination of ranked results from the independent retrieval experts which comprise a CBMIR system through linearly weighted data infusion.

(Rho & Hwang, 2009) address the problem of parsing audiovisual data for content-based multimedia information retrieval, and proposed a novel scheme for determining video scenes by analyzing both audio and video data. (Zhao, 2011) studied the features of content-based multimedia information retrieval, then analyzes the system structure of content-based multimedia information retrieval and retrieval procedures, lastly the author discusses the exist limitations and prospects of content-based multimedia information retrieval.

(Hanjalic, 2012) discussed new MIR grand challenge and position it with respect to the current efforts in the field. Then, some possibilities for realizing the utility-by-design approach will be highlighted and translated into a number of recommended research directions. (Zhou et al., 2012) presented a novel ranking framework for content-based multimedia information retrieval. It shows that the task of CBMIR can be done more effectively using the relevance features than the original features. Furthermore, additional performance gain is achieved by incorporating our new ranking scheme which modifies instance rankings based on the weighted average of relevance feature values. (P. Zhang et al., 2013) focuses on how to retrieve personalized multimedia information based on user

interest which can be mined from user profile. (Karpathy & Li, 2015) presents a retrieval system that aims at addressing challenge by learning the main concept of every image in the medical database and proposed system contains two modules: a classification/annotation and a retrieval module.

(Raieli & Raieli, 2016) explained the story of MIR fundamental principles, from early years of questioning on documentation to today's theories on semantic means. New issues for a LIS methodology of processing and searching multimedia documents are theoretically argued, introducing MIR as a holistic whole composed by content-based and semantic information retrieval methodologies. (Kambau & Hasibuan, 2017) designed Unified Concept-based Multimedia Information Retrieval (UCpBMIR) technique to tackle those difficulties by using unified multimedia indexing. (Kambau & Hasibuan, 2018) critical reviewed evolution of IRS follows three paces: content-based, context-based and concept-based. Each pace takes on indexing system and retrieval techniques to optimize information retrieved. The challenge is how to come up with a retrieval technique that can process unified MIRS in order to retrieve optimally the relevant document.

Video content is evolving enormously with the heavy usage of internet and social media websites. Proper searching and indexing of such video content is a major challenge. The existing video search potentially relies on the information provided by the user, such as video caption, description and subsequent comments on the video. In such case, if users provide insufficient or incorrect information about the video genre, the video may not be indexed correctly and ignored during search and retrieval.

(Szulc, 2019) The aim of the study is to identify the features of the effective Multimedia Information Retrieval System (MMIR). To achieve this goal, a literature analysis of the subject searched in the databases was carried out: Library and Information Science Abstracts (LISA) and Library, Information Science & Technology Abstracts (LIST). The following areas of research were discussed: criteria for information retrieval, standards for the description of the content of multimedia objects, features of an effective MMIR, implemented projects and experiments. The specified features of MMIR, such as completeness, description of behavior, including changing requirements, can be used when defining and analyzing requirements in the software development process. At the

end of the article the possibilities and areas of application of MMIR in the work of libraries were indicated.

(MacFarlane et al., 2020) introduced a five-step process model that extracts feature from multimedia objects from both knowledge organization and machine learning, merging them together to create an index of those multimedia objects. Further, steps in creating an application to utilize the multimedia objects and maintaining and updating the database of features on those objects.

(Saxena et al., 2020) described content-based image retrieval model has been presented which provides a methodology where image can be retrieved on the basis of similarity on multiple features.(Rinaldi et al., 2020) presented novel approaches for image semantic retrieval and a possible combination for multimedia document analysis. Several results have been presented to show the performance of our approach compared with literature baselines.

(Wagenpfeil et al., 2021) defined a graph projection into a 2D space and the corresponding algorithms for indexing and retrieval. The project demonstrate experimentally that the effectiveness of retrieval also increases substantially, as the Graph Code facilitates more levels of detail in feature fusion. These levels of detail also support an increased trust prediction, particularly for fused social media content. In our mathematical model, we define a metric triple for the Graph Code, which also enhances the ranked result representations. Thus, Graph Codes provide a significant increase in efficiency and effectiveness, especially for multimedia indexing and retrieval, and can be applied to images, videos, text and social media information.

3. **CONTENT BASED RETRIEVAL OF LECTURE VIDEO REPOSITORY**:

(Ardizzone & La Cascia, 1997) described a system allowing content-based annotation and querying in video databases. No user action is required during the database population step. The system automatically splits a video into a sequence of shots, extracts a few representative frames from each shot and computes r-frame descriptors based on color, texture and motion. Queries based on one or more features are possible.

(Chang et al., 1999) proposed extraction algorithms can be hierarchically applied to obtain a tree-structured key frame hierarchy that is a multilevel abstract of the video. The key frame hierarchy enables an efficient content-based retrieval by using the depth-first

search scheme with pruning. Intensive experiments on a variety of video sequences are presented to demonstrate the improved performance of the proposed algorithms over the existing approaches.

(D. Zhang & Nunamaker, 2004) proposed a natural language approach to content-based video indexing and retrieval to identify appropriate video clips that can address users' needs. The method integrates natural language processing, named entity extraction, frame-based indexing, and information retrieval techniques to explore knowledge-on-demand in a video-based interactive e-Learning environment. A preliminary evaluation shows that precision and recall of this approach are better than those of the traditional keyword based approach.

(Zhu et al., 2005) introduced a video similarity evaluation scheme at different levels (key-frame, shot, group, scene, and video.) By integrating the video content hierarchy and the video similarity evaluation scheme, hierarchical video browsing and retrieval are seamlessly integrated for efficient content access. (Jawahar & Chennupati, 2005) approached that enables search based on the textual information present in the video. Regions of textual information are identified within the frames of the video. Video is then annotated with the textual content present in the images. Videos containing the query string are retrieved from a video database and sorted based on the relevance. Results are shown from video collections in English, Hindi and Telugu.

(Masihi & Charkari, 2005) presented a query by example mechanism, to make queries from a video database. During a search for similarity matching, in a video database and introduce an algorithm for similarity matching, which finds the similarity between video sequences with different sizes.

(Mittal & Gupta, 2006) present a learning framework 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). The support vector machines associate each set of data points in the multi-dimensional feature space to one of the classes during training. (Repp & Meinel, 2006) represented a general architecture and a new retrieval method for an educational system that is based on a knowledge base of existing recorded lectures. The extraction of metadata from the multimedia resources is one of the main parts of this paper. The recorded lectures are transcripted by an out-of-the-box speech recognition software. The speech recognition software generates a time

stamp for each word. (Lew et al., 2006) surveyed and reviewed articles on content-based multimedia information retrieval and discusses their role in current research directions which include browsing and search paradigms, user studies, affective computing, learning, semantic queries, new features and media types, high performance indexing, and evaluation techniques and discuss the major challenges for the future.

(Yu et al., 2007) introduced a compound model consisting of several atom search modules, i.e., textual and visual, for news video retrieval. First, the analysis on query topics helps to improve the performance of video retrieval. Furthermore, the multimodal fusion of all atom search modules ensures to get good performance. (Smeaton, 2007) a briefed review of the state of the art of video analysis, indexing and retrieval and pointed to research directions which we think are promising and could make searching and browsing of video archives based on video content, as easy as searching and browsing (text) web pages. (Bolettieri et al., 2007) discussed the architecture of a Digital Library for enabling the reusing of audiovisual documents in an e-Learning context. The objective is to demonstrate the reuse of digital content, as video documents or Power Point presentations, by exploiting existing technologies for automatic extraction of metadata and search interface assists the user of the system in the retrieval the multimedia objects in the collection, by combining full-text retrieval on text extracted and metadata, and similarity search on the MPEG-7 visual descriptors.

(Kotecha & Rajurkar, 2009) proposed scheme a promotion is made on ordinal measures by using motion features and color features beside sum of intensities. New video retrieval system is proposed for similarity matching, which finds the similarity between video sequences with different sizes. Moreover, the system is able to find the similarity between a query video and a part of another video sequence. (Carlos A. F. Pimentel Filho & C. A. Saibel Santos, 2010) proposed approach was applied to a database containing more than 34 hours of broadcast news videos. Visual features extracted from the summarized version of the videos were then used for video content indexing and proposes a method for key frame extraction that summarizes video content in a static storyboard, specifically projectec, for key frame retrieval and video access. Thus, the selected key frames are processed in order to extract statistical features as well as wavelet coefficients to represent the video's essence in a very short amount of data while preserving its main content characteristics.

Many universities adopt educational systems where the teacher lecture is video recorded and the video lecture is made available to students with minimum post-processing effort. These cost-effective solutions suffer from the limited amount of annotations associated with the video content, which strongly limits the usability of the service when students need to retrieve specific portions of video, e.g., to revise unclear aspects covered in the past lectures. This paper presents, as a real case study, the system developed and implemented in our university for video lecture annotation and indexing. The original video recordings, which last around 1.5 hour, are first partitioned into smaller segments and then annotated by mapping their content with the entities in a multilingual knowledge base. To this purpose, the proposed approach analyzes both the transcription of the teacher's speech and the text appearing in the video.

(Yang et al., 2011) developed a novel video segmenter for automated slide video structure analysis and a weighted discrete cosines transformation based text detector. A dynamic image contrast adaption serves the purpose of enhancing the text image quality to make it processible by existing common OCR software. Time-based text occurrence information as well as the analyzed text content are further used for indexing. (Patel, 2012) surveyed and reviewed the interesting features that can be extracted from video data for indexing and retrieval along with similarity measurement methods and identify research issues in area of content based video retrieval systems.

(Yang et al., 2012) presented an automated framework for lecture video indexing in the tele-teaching context. The major issues involved in our approach are content-based lecture video analysis and integration of proposed analysis engine into a lecture video portal. In video visual analysis, we apply automated video segmentation, video OCR technologies for extracting lecture structural and textual metadata. Concerning Automated Speech Recognition analysis, we have optimized the workflow for the creation of a German speech corpus from raw lecture audio data. This enables us to minimize the time and effort required for extending the speech corpus and thus improving the recognition rate. Both, OCR and ASR results have been applied for the further video indexing. In order to integrate the analysis engine into the lecture video portal, we have developed an architecture for the corresponding tasks such as, e.g., data transmission, analysis management, and result visualization etc. The accuracy of each individual analysis method has been evaluated by using publicly available test datasets.

(Uke, 2012) proposed a general system for segmenting and organizing method for educational videos that is based on a knowledge base of existing recorded lectures and approach concentrates on the slides instead of a presenter from the projected electronic slides. The proposed method aims in using the video streams of the lecture delivered captured from camera and get the slides converted into images after applying slide segmentation. These independent images have heading and subheadings which will be used for organizing the lecture video. The organizing and segmenting videos of the technical videos according to slide changes not only structure the content of video according to topics, but also facilitate the synchronization of video, audio and electronic slides for effective indexing, browsing and retrieval at later stage.

(Balagopalan et al., 2012) presented a domain independent, statistical approach for automatic keyphrase extraction from audio transcripts of video lectures and identify new features in audio transcripts, that capture key patterns characterizing key phrases in lecture videos. The extracted keyphrases are used further as features for automatic topic based segmentation of the video lectures. This process of automatic keyphrase extraction and segmentation results in a section-wise annotated video lecture which can be effectively viewed in a lecture browser.

The amount of digital material in video lecture archives is growing rapidly, causing the search & retrieval process to be time-consuming and almost impractical. Indeed, after the search, students receive a list of videos and often must use VCR-like functions to find the specific piece of video that covers the searched topic. Therefore, a more efficient method for video retrieval in digital video lecture archives is needed.

Searching for a Video in World Wide Web has augmented expeditiously as there's been an explosion of growth in video on social media channels and networks in recent years. At present video search engines use the title, description, and thumbnail of the video for identifying the right one. In this paper, a novel video searching methodology is proposed using the Video indexing method. Video indexing is a technique of preparing an index, based on the content of video for the easy access of frames of interest. Videos are stored along with an index which is created out of video indexing technique. The video searching methodology check the content of index attached with each video to ensure that video is matching with the searching keyword and its relevance ensured, based on the word count of searching keyword in video index.

(Hanjalic, 2012) proposed a utility-by-design approach, by which utility is targeted explicitly and embedded deep in the foundations of MIR solutions and motivates new MIR grand challenge and position it with respect to the current efforts in the field. Then, some possibilities for realizing the utility-by-design approach will be highlighted and translated into a number of recommended research directions.

(M.Kamde et al., 2013) Implemented CBVR system is experimented based on integration of texture, color and edge features for video retrieval. These features are combined in various ways like entropy- edge, entropy- color for result refinement. Dataset is created with the videos from different domains like e-learning, nature, construction etc. By the combination of these features in different ways, we achieved comparative results. Obtained result shows that combining of two or many features gives better retrieval.

(Haojin Yang & Meinel, 2014)" presented a speech and text based video retrieval and Video search system using Optimal Character Recognition (OCR) and Automated Speech Recognition (ASR). The converted the video into key-frames and extract the Audio and Text using OCR and ASR. This in turn will improve the user's aptitude to quickly review this material. This will make user go through only information that they needed. However, the text in the video may vary in dimension, orientation, style, background, contrast and variations in rhythm, volume of and noise in speech and the differentiating between the key-speeches and unnecessary other sounds used during the recording as well, makes data extraction extremely challenging.

To extract the audio information from the visual content is a very challenging because it needs the extraction of high level semantic information from low level visual data. The summarization technique for videos has been proposed to improve the browsing faster for large video collections to produce more efficient content indexing and access.

Creating video recordings of events such as lectures or meetings is increasingly less expensive and easy. Thus the Video data is increasing in a great deal on World Wide Web (www) and so thus the need of more efficient and correctly functioning method of video indexing, grouping and video retrieval in WWW or Large video archives is necessary.

Recording lectures and putting them on the Web for access by students has become a general trend at various universities. The amount of lecture video data is growing

rapidly on the World Wide Web (WWW). Lecture videos contain text information in the visual as well as audio channels: the presentation slides and lecturer's speech. So it becomes a need for an efficient method for video retrieval in WWW or within large lecture video archives.

E-Learning is the use of educational technology, communication and information technologies and electronic media in education. E learning contains various types of media including images, video, audio, streaming videos, animation, web based learning, video based learning, audio based learning, E books etc. Distance learning can be done without school or collages, anyone can learn from their home or office. E- Learning industry is economically remarkable and it was work out in 2000 to be above 50$ billion corresponding to traditionalist estimates. Lecture Audio, video data on internet is growing rapidly. Hence there is immediate need for method by which we can retrieve audio, videos on internet.

(DUC NGO et al., 2014) discuss content based video retrieval, an overview of a video retrieval framework and consider two important applications of video retrieval nowadays which are video retrieval based on human face and video retrieval based on generic object categories. The goal was to develop approaches which require lowest annotation cost or computational cost while achieving competitive accuracy so that they can facilitate building scalable and comprehensive video retrieval systems. (Kothawade & Patil, 2015) presented an approach for automated video indexing and video search in large lecture video archives. Initially applied automatic video segmentation and key-frame detection and then applied video Optical Character Recognition technology on key-frames to extract textual metadata and Automatic Speech Recognition on lecture audio tracks.

(Roshdi & Roohparvar, 2015) represented the various models and techniques for information retrieval. In this Review paper we are describing different indexing methods for reducing search space and different searching techniques for retrieving information. (Kate et al., 2015). presented a technology for video search in lecture video archive. Initially, introduced segmentation of videos and key frame detection for offering rules for navigation of video contents. By applying ASR (Automatic Speech Recognition) on lecture audio and OCR (Optical Character Recognition) on video content we can extract metadata. OCR can be used in Data entry for business document, Automatic Number

plate Recognition, Extracting business card information into a contact list and so on.ICR (Intelligent character recognition) focuses on handwritten documents as well as cursive character one at a time usually it involves in Machine Learning. Speech recognition system can classify into continuous or discrete system which can be speaker independent, speaker dependent or adaptive. Discrete system focuses on a separate acoustic model for each single word, sentence, phrase etc. are said to be isolated word speech recognition (ISR). CSR (Continuous Speech Recognition) System focuses on user who speaks sentences continually.

(Radha, 2016) presented an analysis process of lecture video retrieval with automated video indexing and video search for lecture video databases. The video retrieval extracts the relevant metadata from the two main parts of lecture videos, namely the visual screen and audio tracks. In the proposed work, the combined text information from the visual slide frame and audio signal were used to retrieve the video from the lecture video database. The performance of this combined video retrieval system shows a significant improvement in performance when compared with an individual system built using audio and text in video systems respectively.

(Balasubramanian et al., 2016) proposed a multimodal metadata extraction system which extracts an optimal set of key phrases and topic based segments that effectively summarize the content of a lecture video. The proposed content-descriptive metadata extraction technique has been evaluated using actual lecture videos from different sources, and our results show that our multimodal approach is effective in summarizing the lecture's content, potentially improving the user experience during retrieval and browsing. (Ravinder & Venugopal, 2016) presented a novel algorithm for content-based video indexing and retrieval using key-frames texture, edge, and motion features. The algorithm extracts key frames from a video using k-means clustering based method, followed by extraction of texture, edge, and motion features to represent a video with the feature vector.

Traditional video retrieval methods fail to meet technical challenges due to large and rapid growth of multimedia data, demanding effective retrieval systems. In the last decade Content Based Video Retrieval (CBVR) has become more and more popular. The amount of lecture video data on the World Wide Web (WWW) is growing rapidly. Therefore, a

more efficient method for video retrieval in WWW or within large video archives is urgently needed.

The advancement in the web technologies has increased the lecture video contents tremendously. The lecture video retrieval for the e-learning process is a challenging task since the videos are unstructured and have a large size. Since many video lectures have less information, the video retrieval system needs to be built with the enhanced features to improve the efficiency of the retrieval process.

With the vast growth and progress of multimedia on the Internet, particularly videos, which caused a growing demand for video retrieval, organizing, and automate systems, as many users mandate content-based retrieval systems over text-based retrieval systems? The process of extracting and selecting features, plays a very important role in content-based video retrieval regardless the significance of video characteristics. The reduction of time and space costs of the retrieval process is dependent on effective features selection.

(Chivadshetti et al., 2016) presented an implementation of automated video indexing and video search in large video database and also present personalized results. Proposed system works in three different phases, in the first phase video segmentation and key frame detection is performed to extract meaningful key frames. Secondly, OCR, HOG and ASR algorithms are applied over the keyframe to extract textual keyword. In the third phase, Color, Texture and Edge features are also extracted. Finally, search similarity measure is performed on the extracted features that are saved in SQL database and the output with personalized re-ranking results as per interest is presented to the users.

(Daga, B. S. et al., 2017) proposed model uses the feature mixture database with the more relevant features such as text, semantic word, and the Local Gabor Pattern vectors. The video retrieval from the feature mixture database is done by using the hybrid K-Nearest Neighbour Naive Bayes (KNB) classifier. This classifier uses the techniques of both the Naive Bayes (NB) classifier and the K-Nearest Neighbour (K-NN) classifier. The performance metrics such as precision, recall and the f-measure analyze the efficiency of the proposed model. Simulation is done by giving the text query and the video query to the video database.

(Silber-Varod et al., 2017) compared ASR with Optical Character Recognition (OCR) as facilitating access to textual and speech content and show their current performance in

under-resourced languages. (Furini, 2018) proposed Video Lecture Browsing, a system designed to facilitate both the retrieval of video lectures within video archives and the finding of the most appropriate segment of a video lecture that covers a searched topic by automatically producing a general picture of the contents of a video lecture. To achieve these goals, the system introduces the idea of timed tag-clouds, which are produced with a combination of aural and visual analysis. Results of a MOS evaluation show that users highly appreciate the timed tag-clouds.

(Araujo & Girod, 2018) introduced a new retrieval architecture, in which the image query can be compared directly with database videos - significantly improving retrieval scalability compared with a baseline system that searches the database on a video frame level. Matching an image to a video is an inherently asymmetric problem. They proposed an asymmetric comparison technique for Fisher vectors and systematically explore query or database items with varying amounts of clutter, showing the benefits of the proposed technique. Large-scale experiments using three datasets show that this technique enables faster and more memory-efficient retrieval, compared with a frame-based method, with similar accuracy. The proposed techniques are further compared against pre-trained convolution neural network features, outperforming them on three datasets by a substantial margin.

(Wang et al., 2018) presented a key frame based method that employs shot boundary detection and "bag-of visual-words (BoW)" based on local keypoints for key frame extraction and semantic concept detection. The performance of BoW features is optimized by choosing appropriate representation choices. The experimental results demonstrate that the proposed approach is capable of retrieving videos. (Medida & Ramani, 2019) presented a methodology for efficient search and retrieval of lecture videos based on Machine Learning (ML) text classification algorithm. The text transcript is generated exclusively from the audio content extracted from the video lectures. This content is utilized for the summary and keyword extraction which is used for training the ML text classification model. An optimized search is achieved based on the trained ML model. The performance of the system is compared by training the system using Naive Bayes, Support Vector Machine and Logistic Regression algorithms. Performance evaluation was done by precision, recall, F-score and accuracy of the search for each of the classifiers. It is observed that the system trained on Naive Bayes classification

algorithm achieved better performance both in terms of time and also with respect to relevancy of the search results.

(Waykar & Bharathi, 2019) proposed the intent aware optimization based on grey Wolf optimizer for retrieving the lecture video. The grey Wolf optimizer is applied to the input database where the clustering task acquires the optimal solution. Finally, the user selected video is matched with the optimal solution to retrieve the more relevant video for the input query. The experimental results are validated, and the parameters used to analyze the performance are F-measure, Recall, and Precision. The performance is compared with the existing systems using MATLAB implementation. (Bhute & Meshram, 2019) discussed the different techniques for text extraction from images and videos. Secondly, they reviewed the techniques for indexing and retrieval of image and videos by using extracted text.

(Jacob et al., 2019) proposed wormhole algorithm to generate video captions are generated by the deep learning network model by combining global that ensure minimum worst-case time for searching a key. Hence, searcher can get and download a video clip instead of downloading entire video from the video storage. This reduces the bandwidth requirement and time taken to download the videos. (Shamsi et al., 2019) proposed a mechanism to understand the contents of video and categorize it as Music Video, Talk Show, Movie/Drama, Animation and Sports. For video classification, the proposed system uses audio and visual features like audio signal energy, zero crossing rate, spectral flux from audio and shot boundary, scene count and actor motion from video.

(Tamil Priya & Divya Udayan, 2019) discussed about the various techniques used to retrieve the multimedia content from the web engines with the semantic knowledge awareness is provided. The various research methods that intend to perform the annotation on multimedia contents is discussed in detailed. And also, this analysis work provides the discussion about the techniques in terms of their merits and demerits. Finally, this work also provides the numerical evaluation of each technique based on their accuracy in the video retrieval outcome. The overall evaluation of the surveyed methodologies is performed and provides discussion about the outcome obtained from each technique.

(Spolaôr et al., 2020) Content-based video retrieval and indexing have been associated with intelligent methods in many applications such as education, medicine and

agriculture. By applying a research protocol proposed by us, 153 papers published from 2011 to 2018 were selected. As a result, it was found that strategies for cut-based segmentation, color-based indexing, k-means based dimensionality reduction and data clustering have been the most frequent choices in recent papers. All the information extracted from these papers can be found in a publicly available spreadsheet. This work also indicates additional findings and future research directions.

(Kalaivani, 2020) proposed video summarization system for instructional videos initially separates videos into audio track and then converted into text transcript. The text transcripts are preprocessed and extracted the important relevant keywords which can be used for indexing the video file. The automatic annotation of instructional videos reduced the summary length and also preserved the semantic of the videos. The research indicates that the students found these summarized content is very helpful for preparing and reviewing the lecture material and also for the preparation during their examination.

(Sedky Adly et al., 2020) discussed content-based video indexing and retrieval, in the application of video search engines CBVSE, which is the problem of searching for digital videos over the Internet. Addressing diverse approaches regarding video indexing and retrieval with brief outline and classification. Furthermore, proposed a framework for a CBVSE that will offer a less complex solution for content-based video retrieval for video search engines over the web using reduced features vector with better accuracy to decrease the complexity trade-off. These features include temporal boundaries patterns combination and key-object concept features, as well as classifying the index list with a multi key-object concepts arrangement built on the statistical redundancy of the detected key-object concepts for each video shot in the sequence.

(Jacob et al., 2020) proposed a novel methodology for the analysis of video content and using video storytelling and indexing techniques for the retrieval of the intended video clip from a long duration video. Video storytelling technique is used for video content analysis and to produce a description of the video. The video description thus created is used for preparation of an index using wormhole algorithm, guarantying the search of a keyword of definite length L, within the minimum worst-case time. This video index can be used by video searching algorithm to retrieve the relevant part of the video by virtue of the frequency of the word in the keyword search of the video index. Instead of downloading and transferring a whole video, the user can download or transfer the

specifically necessary video clip. The network constraints associated with the transfer of videos are considerably addressed.

(Gayathri N & Mahesh K, 2020) analyzed with multimedia content may enhance the accuracy of the indexing and retrieval system, which is performed with two different phases. Former is content-based, and the latter is context-based annotation. The previous annotation includes both higher level and lower level features, which is acquired from pixel intensity; similarly, the latter is based on semantic video details. Here, an extensive analysis is done with three different methods known as Fuzzy based SVM classifier, logit boost ensemble classifier, and Naïve Bayes classifier, where feature extraction is concentrated on various factors. These diverse methods pretend to carry of retrieval of the video effectually with specific advantages and disadvantages. Here, numerical analysis is done with different factors like accuracy, sensitivity, specificity, F-measure, Recall, and so on. The total amount of frames considered for computation is also determined. The videos are attained from the available online dataset, and initial pre-processing is carried out to make the quality of image frames to be superior. The overall analysis shows the performance of classifiers with improved prediction accuracy for video retrieval.

(Janwe & Bhoyar, 2020) Retrieval of videos efficiently and effectively has become a challenging issue nowadays and dealing with multi-concept videos is the center of focus. The aim of the work presented here is to propose an improved semantic concept-based video retrieval method using a novel ranked intersection filtering technique and a foreground driven concept co-occurrence matrix. In the proposed ranked intersection filtering technique, an intersection of ranked concept probability scores is taken from key-frames associated with a query shot to identify concepts to be used in retrieval. Convolutional neural network is used as a baseline. The proposed method is implemented using a classifier built with a fusion of asymmetrically trained deep CNNs to deal with data imbalance problem, a novel foreground driven concept co-occurrence matrix to exploit concept co-occurrence information and a ranked intersection filtering approach. Performance is evaluated by a measure, mean average precision on TRECVID multi-label dataset. The results are compared with state-of-the-art other existing methods in its class and shown its superiority.

(Ashok Kumar et al., 2021) Recently, the popularity for massive online open course (MOOC) learning among student community is increasing day by day. In most of the

MOOC learning Web sites, indexing and retrieval techniques are based on the keywords like names, titles and Web addresses of the videos. As a result, the response for content-based queries is not good in most of the scenarios. In this paper, we propose an efficient scene content-based indexing and retrieval (ESCIR) framework for lecture videos. The proposed ESCIR framework consists of two phases: offline phase and online phase. In offline phase, we apply novel block-level key frame extraction (BLKFE) technique to segment video frames into shots and choose the right frame. The optical character recognition (OCR) tool is applied to the generated key frames in the videos for extracting the text. The jacquard similarity coefficient measure is used to eliminate the duplicate text frames, and then use of stop word removal and stemming algorithms are applied to get meaningful keywords from the scene. We used single-pass in-memory indexing (SPIMI) technique for building the indexing with the help of extracted keywords. In online phase, search and matching algorithms will map input queries given by user to the corresponding videos. We applied the okapi BestMatch25 (BM25) ranking function to rank those matched videos for best relevance results. The proposed ESCIR framework has been validated on standard video lectures and found to give better results than the existing state of art algorithms in terms of precision and accuracy. We compared our results with existing methods of lecture video techniques reflects improvising robust performance on lecture videos.

(Saoudi & Jai-Andaloussi, 2021) 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.

(Tsubasa Yoshizawa, 2021) Recent studies have been proposed to extract speech from the captured video of objects vibrating by sound waves. Among them, from the viewpoint of equipment cost, the method of extracting speech from the video captured by rolling-shutter cameras, which are widely used in consumer digital single-lens reflex cameras, has been attracting attention. The conventional method with the rolling-shutter video uses a grayscale video for processing based on phase images. However, a grayscale video has a smaller dynamic range than an RGB video, and thus the speech extraction accuracy of the conventional method degrades. Therefore, this paper proposes a speech extraction method based on RGB-intensity gradients on an RGB video to improve speech extraction accuracy. The proposed method extracts the speech by calculating the similarity of R, G, and B intensity gradients, and using these three intensity gradients expands the dynamic range. The experimental results on the quality and intelligibility of the extracted speech show our proposed method outperforms the conventional method.

(Spolaôr et al., 2021) Using the voice to interact with systems is attractive in medicine and other areas due to its friendliness and flexibility. Video indexing and retrieval have benefited from this resource. However, few initiatives use speech recognition to support both tasks. This work aims to develop and evaluate a prototype system to index and retrieve videos from speech transcription. In particular, the user can narrate each video's content, generating the utterance that is captured, transformed into text and timestamped by the computational system. Simple text processing techniques are then applied to the obtained transcript before indexing. Afterward, the user can also query by speech or text to find relevant videos previously indexed. We conducted an experimental evaluation of the prototype in sets of 50 and 10 public videos. As part of this process, one collaborator manually narrated the 50 videos, while four others narrated a subset of 13 videos. An automatic narration scheme was also applied to this subset and the set of 10 videos. The evaluation showed promising results regarding Brazilian Portuguese speech recognition and retrieval performance. For example, the average word error rate reached down to 0.03 and the mean average precision achieved up to 1.00. Besides performing well, the computational tool is flexible since few changes are required to support other languages.

4. RESEARCH GAP

This section deals with the research gaps and challenges identified from the existing literature survey.

Most of the lecture videos are annotated with random keywords which are totally inappropriate and leads to irrelevant results. Users can fetch the materials only based on limited options. The search can be made mainly based on occurrence or on the basis of tags attached. Video retrieval purely depends on the metadata of the videos such as title, author details and annotations or user navigation from generic to specific topics. Irrelevant and random annotations, navigations and manual annotations without considering the video contents are the major missing gaps identified in the retrieval of lecture videos. This may lead to irrelevant retrievals and more time consumption for the users.

CONCLUSION

Contents of lecture video consists of different modes of multimedia information namely, text, images and speech. From the exhaustive literature review, for the retrieval of lecture videos, only text and speech transcripts are considered for content based retrieval. Other low level features namely color, texture takes place in lecture video may also affect the retrieval. Our research work considers both low level and high level features for the retrieval. Text and image features in the lecture videos are extracted using OCR and image processing filters respectively. Here, text is the high level feature and image feature is the low level feature. In the literature, either clustering or classification techniques are used for the retrieval process. To enhance the retrieval results, videos are clustered based on the similarities of the extracted text and image features in our research. Classification techniques are used to retrieve the relevant videos from the clustered videos. To further improve the precision of the retrieval, semantic relationship between the videos is identified using WordNet and 26 domain ontology. Based on the similarity between the videos, relevant videos are extracted using semantic web technologies.

This section has dealt with the existing techniques employed in lecture video retrieval. The distinct methodologies adopted for lecture video retrieval have been categorized on the basis of the techniques utilized, like machine learning, content based techniques, indexing based techniques, semantic based techniques and so on. A brief description of these works has been provided along with their limitations.

Araujo, A., & Girod, B. (2018). Large-Scale Video Retrieval Using Image Queries. *IEEE Transactions on Circuits and Systems for Video Technology*, *28*(6). https://doi.org/10.1109/TCSVT.2017.2667710

Ardizzone, E., & La Cascia, M. (1997). Automatic Video Database Indexing and Retrieval. *Multimedia Tools and Applications*, *4*(1). https://doi.org/10.1023/A:1009630331620

Ashok Kumar, P. M., Ambati, R. R., & Arun Raj, L. (2021). An Efficient Scene Content-Based Indexing and Retrieval on Video Lectures. In *Advances in Intelligent Systems and Computing* (Vol. 1171, pp. 521–534). https://doi.org/10.1007/978-981-15-5400-1_53

Balagopalan, A., Balasubramanian, L. L., Balasubramanian, V., Chandrasekharan, N., & Damodar, A. (2012). Automatic keyphrase extraction and segmentation of video lectures. *Proceedings - 2012 IEEE International Conference on Technology Enhanced Education, ICTEE 2012*. https://doi.org/10.1109/ICTEE.2012.6208622

Balasubramanian, V., Doraisamy, S. G., & Kanakarajan, N. K. (2016). A multimodal approach for extracting content descriptive metadata from lecture videos. *Journal of Intelligent Information Systems*, *46*(1). https://doi.org/10.1007/s10844-015-0356-5

Bhute, A. N., & Meshram, B. B. (2019). Text Based Approach For Indexing And Retrieval Of Image And Video: A Review. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3430312

Bolettieri, P., Falchi, F., Gennaro, C., & Rabitti, F. (2007). Automatic metadata extraction and indexing for reusing e-learning multimedia objects. *Workshop on Multimedia Information Retrieval on The Many Faces of Multimedia Semantics - MS '07*, 21. https://doi.org/10.1145/1290067.1290072

Cagliero, L., Canale, L., & Farinetti, L. (2019). VISA: A supervised approach to indexing video lectures with semantic annotations. *Proceedings - International Computer Software and Applications Conference*, *1*. https://doi.org/10.1109/COMPSAC.2019.00041

Carlos A. F. Pimentel Filho, & C. A. Saibel Santos. (2010). A new approach for video indexing and retrieval based on visual features. *Journal of Information and Data Management*, *2*(1).

Chang, H. S., Sull, S., & Lee, S. U. (1999). Efficient video indexing scheme for content-

based retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, *9*(8). https://doi.org/10.1109/76.809161

Chivadshetti, P., Sadafale, K., & Thakare, K. (2016). Content based video retrieval using integrated feature extraction and personalization of results. *Proceedings - IEEE International Conference on Information Processing, ICIP 2015*. https://doi.org/10.1109/INFOP.2015.7489372

Daga B. S., A. A, G., & V.M., T. (2017). Semantic Enriched Lecture Video Retrieval System Using Feature Mixture and Hybrid Classification. *Advances in Image and Video Processing*, *5*(3). https://doi.org/10.14738/aivp.53.2852

Djeraba, C., Sebe, N., & Lew, M. S. (2005). Systems and architectures for multimedia information retrieval. *Multimedia Systems*, *10*(6). https://doi.org/10.1007/s00530-005-0174-7

DUC NGO, T., DINH LE, D., & SATOH, S. (2014). Scalable approaches for content based video retrieval. *Progress in Informatics*, *11*, 31. https://doi.org/10.2201/NiiPi.2014.11.5

Furini, M. (2018). On introducing timed tag-clouds in video lectures indexing. *Multimedia Tools and Applications*, *77*(1), 967–984. https://doi.org/10.1007/s11042-016-4282-5

Gayathri N, & Mahesh K. (2020). An Effectual Video Indexing and Retrieval Model: A Comparative Study. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3734782

Goularte, R., Trojahn, T. H., & Kishi, R. M. (2019). Multimedia information retrieval in big data using OpenCV python. *Proceedings of the 25th Brazillian Symposium on Multimedia and the Web, WebMedia 2019*. https://doi.org/10.1145/3323503.3345030

Hanjalic, A. (2012). New grand challenge for multimedia information retrieval: bridging the utility gap. *International Journal of Multimedia Information Retrieval*, *1*(3). https://doi.org/10.1007/s13735-012-0019-z

Haojin Yang, & Meinel, C. (2014). Content Based Lecture Video Retrieval Using Speech and Video Text Information. *IEEE Transactions on Learning Technologies*, *7*(2), 142–154. https://doi.org/10.1109/TLT.2014.2307305

Jacob, J., Elayidom, M. S., & Devassia, V. P. (2020). Video content analysis and retrieval system using video storytelling and indexing techniques. *International Journal of Electrical and Computer Engineering*, *10*(6). https://doi.org/10.11591/ijece.v10i6.pp6019-6025

Jacob, J., Ilayidom, S., & Devassia, V. P. (2019). Content Based Video Retrieval System Using Video Indexing. *International Journal of Computer Sciences and Engineering*, *7*(4), 478–782. https://doi.org/10.26438/ijcse/v7i4.478782

Janwe, N., & Bhoyar, K. (2020). Semantic concept based video retrieval using convolutional neural network. *SN Applied Sciences*, *2*(1). https://doi.org/10.1007/s42452-019-1870-9

Jawahar, C., & Chennupati, B. (2005). Video retrieval based on textual queries.

Proceedings of the …, 1–6. https://doi.org/10.1.1.61.9857

Kalaivani, B. A. (2020). Automatic Annotation of Instructional Videos. *International Journal of Recent Technology and Engineering*, *8*(6), 1683–1687. https://doi.org/10.35940/ijrte.F7414.038620

Kambau, R. A., & Hasibuan, Z. A. (2018). Evolution of information retrieval system: Critical review of multimedia information retrieval system based on content, context, and concept. *Proceedings of the 11th International Conference on Information and Communication Technology and System, ICTS 2017*, *2018-January*. https://doi.org/10.1109/ICTS.2017.8265652

Kambau, R. A., & Hasibuan, Z. A. (2017). Unified concept-based multimedia information retrieval technique. *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, *4*. https://doi.org/10.11591/eecsi.4.1007

Kankanhalli, M. S., & Rui, Y. (2008). Application potential of multimedia information retrieval. *Proceedings of the IEEE*, *96*(4). https://doi.org/10.1109/JPROC.2008.916383

Karpathy, A., & Li, F.-F. (2015). Private Content Based Multimedia Information Retrieval Using Map-Reduce. *Scientific World Journal*, *3*(01).

Kate, L. S., Waghmare, M. M., & Priyadarshi, A. (2015). An approach for automated video indexing and video search in large lecture video archives. *2015 International Conference on Pervasive Computing: Advance Communication Technology and Application for Society, ICPC 2015*. https://doi.org/10.1109/PERVASIVE.2015.7087169

Kotecha, K. S., & Rajurkar, A. M. (2009). Content based video retrieval using ordinal measures, motion and color distribution. *Proceedings of the 2009 International Conference on Image Processing, Computer Vision, and Pattern Recognition, IPCV 2009*, *2*.

Kothawade, A. Y., & Patil, D. R. (2015). A Survey on Automatic Video Lecture Indexing. *International Research Journal of Engineering and Technology*, *02*(02).

Lew, M. S. M. S., Sebe, N., Djeraba, C., & Jain, R. (2006). Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, *2*(1), 1–19. https://doi.org/10.1145/1126004.1126005

Lin, L. (2011). Multimedia data mining and retrieval for multimedia databases using associations and correlations. *ACM SIGMultimedia Records*, *3*(1). https://doi.org/10.1145/2069196.2069200

M.Kamde, P., Shiravale, S., & P. Algur, S. (2013). Entropy Supported Video Indexing for Content based Video Retrieval. *International Journal of Computer Applications*, *62*(17). https://doi.org/10.5120/10169-9974

MacFarlane, A. (2016). Knowledge organisation and its role in multimedia information retrieval. In *Knowledge Organization* (Vol. 43, Issue 3). https://doi.org/10.5771/0943-7444-2016-3-180

MacFarlane, A., Missaoui, S., & Frankowska-Takhari, S. (2020). On Machine Learning and Knowledge Organization in Multimedia Information Retrieval. *KNOWLEDGE ORGANIZATION*, *47*(1). https://doi.org/10.5771/0943-7444-2020-1-45

Masihi, Z. G., & Charkari, N. M. (2005). Content based Video Retrieval based on Approximate String Matching. *EUROCON 2005 - The International Conference on "Computer as a Tool," II*, 1300–1303. https://doi.org/10.1109/EURCON.2005.1630196

Medida, L. H., & Ramani, K. (2019). An optimized E-lecture video retrieval based on machine learning classification. *International Journal of Engineering and Advanced Technology*, *8*(6). https://doi.org/10.35940/ijeat.F9114.088619

Meghini, C., Sebastiani, F., & Straccia, U. (2001). A model of multimedia information retrieval. *Journal of the ACM*, *48*(5). https://doi.org/10.1145/502102.502103

Mittal, A., & Gupta, S. (2006). Automatic content-based retrieval and semantic classification of video content. *International Journal on Digital Libraries*, *6*(1). https://doi.org/10.1007/s00799-005-0119-y

Patel, B. V. (2012). Content Based Video Retrieval Systems. *International Journal of UbiComp*, *3*(2), 13–30. https://doi.org/10.5121/iju.2012.3202

Priscilla, C. V., & Rajeshwari, D. (2020). Perspective study on content based video retrieval. *International Journal on Emerging Technologies*, *11*(2).

Radha, N. (2016). Video retrieval using speech and text in video. *2016 International Conference on Inventive Computation Technologies (ICICT)*, *2*, 1–6. https://doi.org/10.1109/INVENTIVE.2016.7824801

Raieli, R., & Raieli, R. (2016). Introducing Multimedia Information Retrieval to libraries. *JLIS.It*, *7*(3).

Ravinder, M., & Venugopal, T. (2016). Content-Based Video Indexing and Retrieval using Key frames Texture, Edge and Motion Features. *International Journal of Current Engineering and Technology*, *6*(2).

Repp, S., & Meinel, C. (2006). Semantic indexing for recorded educational lecture videos. *Proceedings - Fourth Annual IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2006*, *2006*. https://doi.org/10.1109/PERCOMW.2006.122

Rho, S., & Hwang, E. (2009). Content-based scene segmentation scheme for efficient multimedia information retrieval. *International Journal of Wireless and Mobile Computing*, *3*(4). https://doi.org/10.1504/IJWMC.2009.029350

Rinaldi, A. M., Tommasino, C., & Russo, C. (2020). A knowledge-driven multimedia retrieval system based on semantics and deep features. *Future Internet*, *12*(11). https://doi.org/10.3390/fi12110183

Roshdi, A., & Roohparvar, A. (2015). Review : Information Retrieval Techniques and Applications. *International Journal of Computer Networks and Communications Security*, *3*(9), 373–377.

Rui, Y., Huang, T. S., Mehrotra, S., & Ortega, M. (1997). A relevance feedback architecture for content-based multimedia information retrieval systems. *Proceedings - IEEE Workshop on Content-Based Access of Image and Video Libraries, CBAIVL 1997*. https://doi.org/10.1109/IVL.1997.629724

Saoudi, E. M., & Jai-Andaloussi, S. (2021). A distributed Content-Based Video Retrieval system for large datasets. *Journal of Big Data*, *8*(1). https://doi.org/10.1186/s40537-021-00479-x

Saxena, P., Singh, S. K., & Srivastava, M. (2020). Content-Based Retrieval of Multimedia Information Using Multiple Similarity Indexes. *Advances in Intelligent Systems and Computing*, *1053*. https://doi.org/10.1007/978-981-15-0751-9_113

Sedky Adly, A., Abdelwahab, M. S., Hegazy, I., & Elarif, T. (2020). Issues and Challenges for Content-Based Video Search Engines A Survey. *2020 21st International Arab Conference on Information Technology (ACIT)*, 1–18. https://doi.org/10.1109/ACIT50332.2020.9300062

Shamsi, F., Daudpota, S. M., & Shaikh, S. (2019). Content-based automatic video genre identification. *International Journal of Advanced Computer Science and Applications*, *10*(6). https://doi.org/10.14569/ijacsa.2019.0100677

Silber-Varod, V., Winer, A., & Geri, N. (2017). Opening the knowledge dam: Speech recognition for video search. *Journal of Computer Information Systems*, *57*(2). https://doi.org/10.1080/08874417.2016.1183423

Smeaton, A. F. (2007). Techniques used and open challenges to the analysis, indexing and retrieval of digital video. *Information Systems*, *32*(4). https://doi.org/10.1016/j.is.2006.09.001

Spolaôr, N., Lee, H. D., Takaki, W. S. R., Ensina, L. A., Coy, C. S. R., & Wu, F. C. (2020). A systematic review on content-based video retrieval. *Engineering Applications of Artificial Intelligence*, *90*. https://doi.org/10.1016/j.engappai.2020.103557

Spolaôr, N., Lee, H. D., Takaki, W. S. R., Ensina, L. A., Parmezan, A. R. S., Oliva, J. T., Coy, C. S. R., & Wu, F. C. (2021). A video indexing and retrieval computational prototype based on transcribed speech. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-021-11401-1

Szulc, J. (2019). The features and functions of an effective multimedia information retrieval system (MMIR). *Advances in Intelligent Systems and Computing*, *833*. https://doi.org/10.1007/978-3-319-98678-4_14

Tamil Priya, D., & Divya Udayan, J. (2019). A comprehensive survey on various semantic based video/image retrieval techniques. *International Journal of Innovative Technology and Exploring Engineering*, *8*(6).

Tsubasa Yoshizawa, A. Y. I. T. N. (2021). Design and Implementation of Video Speech Extraction Text System Based on Deep Learning. *Software Engineering and Applications*, *10*(04). https://doi.org/10.12677/sea.2021.104057

Uke, N. (2012). Segmentation and Organization of Lecture Video Based On Visual

Contents. *International Journal of E-Education, e-Business, e-Management and e-Learning*. https://doi.org/10.7763/ijeeee.2012.v2.95

Uma, R., & Muneeswaran, K. (2017). OMIR: Ontology-Based Multimedia Information Retrieval System for Web Usage Mining. *Cybernetics and Systems*, *48*(4). https://doi.org/10.1080/01969722.2017.1285163

Wagenpfeil, S., Vu, B., McKevitt, P., & Hemmje, M. (2021). Fast and effective retrieval for large multimedia collections. *Big Data and Cognitive Computing*, *5*(3). https://doi.org/10.3390/bdcc5030033

Wan, G., & Liu, Z. (2008). Content-based information retrieval and digital libraries. *Information Technology and Libraries*, *27*(1). https://doi.org/10.6017/ital.v27i1.3262

Wang, W., Huang, Z., Wang, W., Zhang, S., & Tian, R. (2018). Video indexing and retrieval based on key frame extraction. *International Journal of Performability Engineering*, *14*(8). https://doi.org/10.23940/ijpe.18.08.p19.18241832

Waykar, S. B., & Bharathi, C. R. (2019). Intent aware optimization for content-based lecture video retrieval using grey Wolf optimizer. *Journal of Engineering Research (Kuwait)*, *7*(3).

Wilkins, P. (2009). An Investigation Into Weighted Data Fusion for Content-Based Multimedia Information Retrieval. *Computing*.

Yang, H., Oehlke, C., & Meinel, C. (2012). An Automated Analysis and Indexing Framework for Lecture Video Portal. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 7558 LNCS* (pp. 285–294). https://doi.org/10.1007/978-3-642-33642-3_31

Yang, H., Siebert, M., L??hne, P., Sack, H., & Meinel, C. (2011). Automatic lecture video indexing using video OCR technology. *Proceedings - 2011 IEEE InternationalSymposium on Multimedia, ISM 2011*, 111–116. https://doi.org/10.1109/ISM.2011.26

Yin, J., Yu, X., & Dong, M. (2010). Application of approximate string matching in video retrieval. *ICACTE 2010 - 2010 3rd International Conference on Advanced Computer Theory and Engineering, Proceedings*, *4*. https://doi.org/10.1109/ICACTE.2010.5579482

Yu, H., Su, B., Lu, H., & Xue, X. (2007). News Video Retrieval by Learning Multimodal Semantic Information. In *Advances in Visual Information Systems* (pp. 403–414). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-76414-4_39

Zhang, D., & Nunamaker, J. F. (2004). A natural language approach to content-based video indexing and retrieval for interactive e-Learning. *IEEE Transactions on Multimedia*, *6*(3). https://doi.org/10.1109/TMM.2004.827505

Zhang, P., Xia, J., & Li, R. (2013). Personalized multimedia information retrieval based on user profile mining. *Journal of Networks*, *8*(10). https://doi.org/10.4304/jnw.8.10.2340-2346

Zhao, J. (2011). Research on content-based multimedia information retrieval.

Proceedings - 2011 International Conference on Computational and Information Sciences, ICCIS 2011. https://doi.org/10.1109/ICCIS.2011.224

Zhou, G. T., Ting, K. M., Liu, F. T., & Yin, Y. (2012). Relevance feature mapping for content-based multimedia information retrieval. *Pattern Recognition*, *45*(4). https://doi.org/10.1016/j.patcog.2011.09.016

Zhu, X., Elmagarmid, A. K., Xue, X., Wu, L., & Catlin, A. C. (2005). InsightVideo: Toward hierarchical video content organization for efficient browsing, summarization and retrieval. *IEEE Transactions on Multimedia*, *7*(4), 648–666. https://doi.org/10.1109/TMM.2005.850977