AIVP ADVANCES IN IMAGE AND VIDEO PROCESSING
SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM
VOLUME 5, ISSUE 3
ISSN 2054-7412

Semantic Enriched Lecture Video Retrieval System Using Feature Mixture and Hybrid Classification

[1]B. S. Daga, [2]Dr A. A. Ghatol and [3]V.M.Thakare
[1]Associate Professor
Fr. Conceicao Rodrigues College of Engineering, Mumbai, India
[2]Former Vice-Chancellor
Dr Babasaheb Ambedkar Technological University,
Lonere, India
[3] Professor&Head, Department of Computer Science & IT,
Sant Gadge Baba Amravati University
bsdaga0177@gmail.com

ABSTRACT

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. In this paper, a semantic enriched lecture video retrieval system has been proposed. The key frames from the video are extracted through the pre-processing. The proposed model uses the feature mixture database with the more relevant features such as text, semantic word, and the Local Gabor Pattern (LGP) 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. The simulation results show that the proposed model has better precision and recall value of 1.0 and 0.7500 respectively. The f-measure of the proposed model has a better value of 0.8519 than the existing K-NN system.

Keywords: Keyframes, Semantic word, LGP vector, and feature mixture database

1 Introduction

Video recordings in the classroom and the conference halls make the learning process of the student to be easier. It also helps the students with a visual challenge to recollect the lecture notes from their classes [1]. The e-learning makes the learning process to be simple. The e-learning allows the digitization of the classroom lectures and the study materials. The classroom lectures mainly compiled of the presentations with the video contents. The presentation video of these lectures has a digital copy and can be stored for the further reference [8]. The era of the modern technology has tremendously increased the video files. Since these video files have large storage and high complexity, video retrieval from the large video database has evolved as a challenging task [5]. Video retrieval defines the process of obtaining the required video content from the large database. The video

DOI: 10.14738/aivp.53.2852
Publication Date: 18th June, 2017
URL: http://dx.doi.org/10.14738/aivp.53.2852

recordings from the universities across the world have a long and unstructured character. This increases the difficulty of browsing, navigation, and location of the video contents from the internet [4]. The video retrieval from the large database should be more related to the query of the user. The video files from the lectures and the conferences contain both the text-based and the content-based image frames. The lecture video file has two common retrieval methods. They are,

- Text-based video retrieval,
- Content-based video retrieval.

The video file has both the text and the images flowing at a rapid rate [30-33]. The text based video retrieval method retrieves the text video files present in the database. The content-based video retrieval method allows the retrieval of the video files with small amount of the text data [6]. The content-based video retrieval method finds a wide application since most of the video files have a small amount of the text data. The lecture video files have insufficient text data, and hence content based methods allow retrieval of these video files. Few research works have focused on the retrieval of the video files combining two video streams [10]. The combined video files combine the video streams from the video camera and the frame grabber. The content-based video retrieval system retrieves the video by extracting the key phrases from the video contents [2]. The key phrases from the video make the retrieval process simpler. The content based retrieval system also allows topic based video segmentation. The video retrieval process uses the following three components [5],

- Query
- Database
- Ranking function

The user given query to the retrieval system represents the type of the video with the semantic concept which the user wants from the internet. The database represents the video collection, and it allows the extraction of the required video for the user. The ranking function rearranges the database with respect to the query of the user. When the user gives the query to the video retrieval system, the system provides the videos based on the query. But, due to the complexity of the video database, the video retrieval system will not produce the desired results [21]. This inclusion of the semantic theme level in the video retrieval system improves the video retrieval process. The semantic theme represents the general topic of the required video file. The semantic theme relates the query and the videos of the database through the general terms other than text and the content. Few examples of the topical queries with the semantic themes are scientific papers, journals, economics, sports events, and conference, etc. [7]. The semantic themes have different levels of abstraction and degrees which make the retrieval process less complicated [14]. Various research allows finding of the suitable semantic label from the video related to the user query. The use of semantic labels in the video retrieval system eliminates the conventional way of assigning the labels for the video retrieval [19].

In this research work, a video retrieval system with the semantic enriched features has been proposed. This research designs and develops a semantic-enriched lecture video retrieval system using feature mixture and hybrid classification. The proposed semantic-enriched lecture video retrieval system utilizes feature mixture which includes textual features, semantic information and content features for the retrieval purpose. At first, input videos will be read out, and frames will be extracted from the input videos. Then, key frames will be identified from input frames. Once key frames are identified, feature mixture will be computed using three level of information. The first level of the feature is the textual contents which will be extracted using Optical Character Recognition (OCR) methods. The

second set of the feature is based on the semantic information which includes the hypernyms and hyponyms of every keyword which will be computed based on WordNet ontology. The third level of the feature is based on the visual content which will be extracted based on the texture strength. Texture consistency will be effectively estimated using Local Gabor Pattern (LGP) which is one of the recent and effective techniques for texture descriptors of images. These three set of information will be extracted from every video, and it will be stored in the indexed database. When query frame or text information is given as input, the proposed system will extract these feature mixture from the input, and it will be matched with the database using the proposed hybrid classifier which will be newly developed by combining the K-Nearest Neighbour (K-NN) classification and Naive Bayes (NB) classification models.

The major contributions of this paper are listed as follows:

- The primary contribution of this paper allows the video retrieval through the feature mixture of the text, contents and the semantic features.
- The secondary contribution of this paper introduces the novel hybrid classifier which combines the properties of the K-NN and the NB classifier models.

The organization is done as follows: Section 1 discusses the introduction of the paper. The Section 2 reviews the existing technologies in the lecture video retrieval system. Section 3 briefs the proposed model along with the newly developed hybrid KNB classifier. The Section 4 discusses the implementation of the proposed model, results obtained and its comparison with the existing systems. The Section 5 summarizes the paper with the conclusion and the future work.

2 Motivation

2.1 Review and Comparison

This section compares the various video retrieval systems used for retrieving the lecture video contents.

Manish Kanadje *et al.* [1] have proposed an unsupervised model for feature keyword extraction from the speaker systems. This keyword spotting system provides results using Segmental Dynamic Time Warping method without the use of the training data from the video. This method has better robustness for the varying environmental conditions, and the retrieval precision of the model has a higher value due to feature normalization of the keywords. The MFCC model suffers from a disadvantage when multiple speakers are used. Vidhya Balasubramanian *et al.* [2] have proposed the ASR-rule based algorithm for the lecture video retrieval. This model combines the Naive Bayes classifier and a rule-based refiner for retrieving the metadata from the lecture videos. This model has an improved recall and F-measure in the retrieval process. The precision of getting the required video content from the database has low results during the occurrence of the transcription error in the searching process. Haojin Yang and Christoph Meinel [3] have proposed a retrieval method based on the OCR and ASR technology. This method has a better keyword extraction due to the use of the ASR method. This model is based on the content-based video retrieval method. The OSR method makes the retrieval to achieve low rejection rate and performs the retrieval process faster. The inclusion of the salt and pepper noise in the video reduces the accuracy of the retrieval process. Kai Li *et al.* [4] have proposed a keyword-based video retrieval technique which concentrates more on the semantic-based approaches. The model detects the semantic labels through the screen detection and localization approach. This method has a better semantic structure for the query. This model does not detect the slide progressions in the videos.

Ruben Fernandez-Beltran and Filiberto Pla [5] have compared the four topic models with the various video retrieval systems. The simulation results show that the pLSA model better retrieves the video files at the sparse conditions. The LDA techniques outperform the pSLA and the IpSLA in crowded video database, and IpLSA shows an average performance in both the video databases. This model allows the retrieval of the new topic from the video database. The use of the multi-modal data in the system makes the video retrieval less accurate. Sara Memar *et al.* [6] have proposed an approach which combines the knowledge-based and corpus-based semantic word similarity measures. This method allows the effective retrieval of the video files through the finding of the semantic similarity. This method serves as an advantage when the system has fewer video annotations. This method improves the degree of relevancy between video query and shot in the database. Stevan *Rudinac et al.* [7] have proposed a semantic theme based video retrieval. When a query is given to the database, the system automatically performs the search process without the use of the training data from the database. It reduces the noise output from the concept detectors. This model has less accuracy. Nhu Van Nguyen *et al.* [8] have proposed a Multi-modal and cross-modal way of video retrieval. This proposed model uses the Bag of Subjects model to retrieve the lecture video files effectively. Accuracy gets reduced due to the error generation from the text recognition phase of the model.

2.2 Challenges

The various challenges involved in the lecture video retrieval process are explained as follows,

- The retrieval of the lecture video contents from the large video sources often has major challenges. The efficiency of the retrieval process gets affected when the source database has a larger size. The popular video search engines such as YouTube and Bing etc. allow the search process with query title, genre, description, etc. [3].
- The search process uses low-level image features such as various color, shape, and texture in the video retrieval [18]. These techniques have produced a low-quality video retrieval since the videos have many common features.
- The lecture video retrieval from the wealth of information database requires efficient video retrieval techniques with the indexing feature, analysis of the database, better search environment, and organization of multimedia data from the video content [15].
- The raw lecture videos from the internet database have a long and random video files, and hence the retrieval model has to be made more efficient with the features of content-based indexing, browsing, video location and searching [4].
- The major difference gap between the visual features and the query of the user allows the video retrieval complicated [13]. The semantic gap should be improved for the better performance.
- The maximum percentage of video lectures from the video database has no tag and thus provides less information for the user to retrieve [24].

3 Proposed Methodology: Semantic Enriched Lecture Video Retrieval System using Feature Mixture and Hybrid Classification

This paper introduces a novel method for the video retrieval for the large source of the video database. This paper has proposed a semantic enriched lecture video retrieval system using feature mixture and hybrid classification. The proposed semantic-enriched lecture video retrieval system utilizes feature mixture which includes textual features, semantic information and content features for the retrieval purpose. The block diagram of the Semantic enriched lecture video retrieval system is shown in figure

1. At first, input videos from the video database will be read out, and frames will be extracted from the input videos. The pre-processing allows the extraction of the key frames from the input frames. Once key frames are identified, feature mixture will be computed using three level of information.

The first level of the feature is the textual contents which will be extracted using OCR methods. The second set of the feature is based on the semantic information which includes the hypernyms and hyponyms of every keyword which will be computed based on WordNet ontology. The third level of the feature is based on the visual content which will be extracted based on the texture strength. LGP estimates the texture consistency of the key frames. These three set of information will be extracted from every video, and it will be stored in the indexed database or the feature mixture database. When query frame or text information is given as input, the proposed system will extract the feature mixture. The query is matched with each and every window, and the distance measure is found. The feature mixture database is split as two groups based on the query. The proposed hybrid K-NN Naive Bayes (KNB) classifier performs the classification of the videos in the groups and retrieves the essential lecture videos from the database. The hybrid KNB classifier uses the models of both the K-NN classification and naive Bayes classification methods.

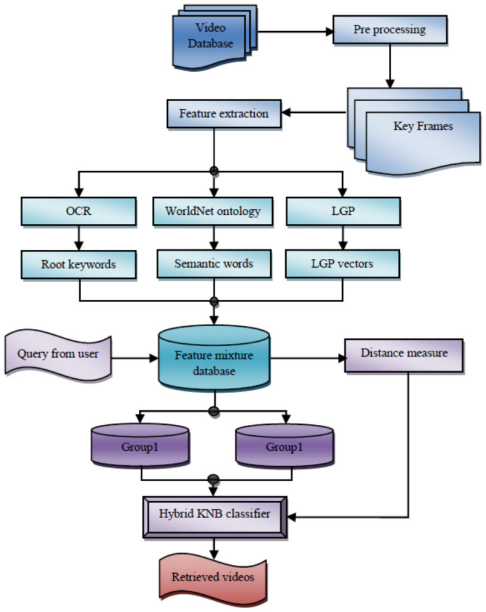


Figure 1 Semantic enriched lecture video retrieval system

3.1 Pre-processing

The initial process in the video retrieval is the pre-processing. The frames in the video files have the noise and distortion contents. To remove these noise factors from the video files preprocessing has to be done. The pre-processing process allows extraction of the key frames from the frames of the video files. The m frames in the video files make the feature extraction process to be difficult. The pre-processing step extracts the p key frames from the video with frames. The key frame extraction from the video frames reduces the computational complexity and improves the efficiency of the video retrieval process. The keyframes have the maximum information contents of the video. Consider a database D containing n video contents. The database can be mathematically expressed by the equation 3.1.

$$D = \{V_i ; \quad 1 \le i \le n\} \tag{1}$$

Where V_i represents the video files. The videos in the database contain image frames with the rapid arrival rate. For the video V_i with the m frames, the mathematical expression can be given in the equation 3.2.

$$V_i = \left\{ f_j^i ; \quad 1 \le j \le m \right\} \tag{2}$$

where f_j^i is the j[th] frame of the i[th] video content in the database D. The p key frames from the m frames are taken for the feature extraction process. The frame in the video is a two-dimensional image. The equation 3.3 expresses the j[th] frame of the i[th] video content with frame size X × Y.

$$f_j^i = \begin{Bmatrix} f_j^i(x,y); & 1 \le x \le X \\ & 1 \le y \le Y \end{Bmatrix} \tag{3}$$

where f_j^i is the j[th] frame of the i[th] video content with the size X × Y.

3.2 Construction of feature mixture database

The videos on the internet have various image sizes, objects, random colors and the change in the video motion. However, the lecture videos in the video database have a constant motion and the objects [4]. The lecture videos have the series of the frames as the videos with the image slides and the texts. The various features in the lecture videos are text words, semantic words, Local Gabor vectors.

3.2.1 Features of the Lecture video

This paper introduces the features such as root keyword, Semantic words and the Local Gabor Pattern vectors for the lecture video retrieval process. The feature extraction process of the video content provides root keyword, Semantic words from the Word set Ontology and the Local Gabor Pattern vectors from the Local Gabor filters.

a) Root keyword

The root keyword contains the main information contents such as title, name, etc. This feature reduces the complexity of the lecture video retrieval process by finding the topic and the subtopics of the video in the database based on the query. The Optical Character Recognition (OCR) allows the extraction of the root keywords from the video frames. The OCR search engine used in this paper is Tesseract OCR engine [23]. The Tesseract OCR model has the improved accuracy. This OCR is one of the widely used tools for the text extraction from the images in the commercial applications.

The Tesseract OCR model extracts the keywords from the video frames in four steps. The Tesseract OCR engine uses the binary image as the input since it does not have a defined layout. The binary images have the defined text regions which form as the keywords for the feature extraction process. The primary step in the OCR involves the analysis of the connected components in the image frame. This step detects the inverse text as the black on white text. This simplifies the process of the child and grandchild outline detection. In the next step, the outlines from the images are nested together to form as a blob. The proportional text from the blob is obtained from well-organized text lines and regions. The image frames have a different character spaced text lines since the splitting of the text lines is done. The character cells in the fixed pitch text allow the splitting of the text lines. In the third step, the definite spaces and fuzzy spaces in the OCR engine split the proportional text in the image frame into words. The word from the image frames needs to be identified for the further processing.

The two-pass process in the Tesseract OCR recognizes the words. The two pass process is done as follows,

- In the first pass, the splitted proportional text word is recognized in the alternative turns. The word acts as a training data for the adaptive classifier of the Tesseract OCR. The adaptive classifier gets updated based on the incoming word and recognizes the word from the bottom of the image.
- Then classifier learns the nature of the data from the first pass, and thus recognizes the data from the top of the image frame in the second pass.

The words which were not recognized in the first pass are recognized in the second pass. In the final step, the fuzzy spaces in the image frames are resolved, and the root keywords from the image get extracted. The root keyword from the feature extraction of the video content V_i can be mathematically expressed as follows,

$$R\left(f_j^i\right)=\left\{r_j^i(a);\quad 1\le a\le A\right\} \tag{4}$$

Where $r_j^i(a)$ indicates the a[th] root keyword in the j[th] frame of the i[th] video content. The frames contain 'A' root keywords.

b) Semantic words:

The semantic word set features are obtained from the root keywords. The semantic word set finds the relation between the root keywords. The WordNet Ontology finds semantic word extraction from the root keyword of the video frames. The WordNet groups the identical words under a common category from the word collection. The synsets define the common word collection. The WordNet contains a group of synsets [25]. The words in the synsets are related in two ways,

- Conceptual Semantic relation
- Lexical relation

The semantic word contains the hypernym and the hyponym relation of the root keyword. The words in the WordNet have been linked with these relations. The WordNet forms the tree structure with the superior and the inferior relation between the words. The hypernym defines the name of the group, for ex: vehicles whereas the hyponym defines the words that can be included in the group, for ex: car, bus, van, etc. This defines the hyperonymy as the synset relation [26].

The WordNet Ontology finds the semantic word set from the root keyword. The WordNet Ontology is a computer-based lexical reference system for finding the relation between the synsets [22]. The synsets in the WordNet has lexical memory reference, Nouns, Adverbs, Adjectives and Verbs as a common set. The WordNet Ontology allows the WordNet as an online dictionary with the relations between the words. It also provides the automatic word analysis and the machine intelligence for finding the relation between the words. The WordNet allows the 'is the synonym of' relation between the word sets. The WordNet Ontology is an open ware and thus can be used as free for the researches. Since it is a generic ontology, the LSI concepts can be used for finding the semantic words [27].

The semantic word set from the root keywords is found by relating the words with the conceptual relation. The linguistic meaning of the word invokes the relation between the root keywords. The WordNet Ontology allows the building of the relationship tree between the word set based on the concept or the inherency. The common feature between the root words is found by enabling the linguistic domain. The semantic features find the difference between the concepts in the root

keywords. The semantic property in the WordNet Ontology finds the relation between the root words and classifies it accordingly.

The semantic word set obtained from the WordNet Ontology allows the video retrieval process more accurate. It relates the query and the video database to obtain the required video contents. The intersection of the semantic word set provides the class features of the query needed for the video retrieval process. The equation 3.5 indicates the semantic word set from the root keyword. The semantic word set has B words.

$$W\left(f_j^i\right)=\left\{w_j^i(b);\quad 1\le b\le B\right\} \tag{5}$$

Where $w_j^i(b)$ indicates the semantic word set from the root keyword $r_j^i(a)$ in the j[th] frame of the i[th] video.

c) Local Gabor Pattern (LGP)

The LGP vectors define the commonly repeating patterns in the video frame. The lecture video files have the repeating patterns of the frames, and hence this feature plays an important role in the video retrieval. This feature allows finds the repeating patterns in the image through the Gabor filters. The LGP vectors allow the texture analysis of the image frames. The Gabor filters find major application in the pattern analysis. The Gabor filters are used commonly because it has a better quality of extraction of the text patterns from the large complex images [28]. The Gabor filters find the text of the both 2D and the 3D images and hence can be used in the lecture video retrieval system. The Gabor filter in the 2D format has two representations. The equation 3.6 and 3.7 represents the 2D Gabor filter.

$$G_c[x,y]=De^{-\frac{\left(x^2+y^2\right)}{2\sigma^2}}\cos\left[2\pi f\left(x\cos\theta+y\sin\theta\right)\right] \tag{6}$$

$$G_s[x,y]=Ee^{-\frac{\left(x^2+y^2\right)}{2\sigma^2}}\cos\left[2\pi f\left(x\cos\theta+y\sin\theta\right)\right] \tag{7}$$

where D and E indicate the constants for the normalization.

f indicates the frequency at which the texture analysis is done.

θ indicates the direction of the texture in the image.

σ indicates the size of the image region of the frame.

(x, y) indicates the pixel location of the image.

The LGP vectors can be obtained from the video by applying the image frames to the Gabor filter. The LGP vector for the j[th] frame of the i[th] video is mathematically expressed by equation 3.8. The equation 3.9 represents the application of the j[th] frame of the i[th] video to the Gabor filter.

$$L\left(f_j^i\right)=\left\{l_j^i(c);\quad 1\le c\le C\right\} \tag{8}$$

$$l_j^i(c)=G\left(f_j^i,(x,y)\right) \tag{9}$$

The application of the Gabor filter to the video gives the LGP vectors of the corresponding frame. The LGP vectors from the frame are found by analyzing the center pixel and the surrounding pixels of the frame. The application of the Gabor filter to the j[th] frame of the i[th] video is expressed by equation 3.10.

$$G\left(f_j^i,(x,y)\right)=\sum_{p=0}^{7}s\left(f_p-f_c\right)2^p \tag{10}$$

Where, $s\left(f_p-f_c\right)=\begin{Bmatrix}1; & f_p\geq f_c\\ 0; & f_p<f_c\end{Bmatrix}$

f_p indicates the pixels surrounding the center pixel of the image. The value of the p varies from 0 to 7.

f_c indicates the center pixel of the j^{th} frame.

3.2.2 Feature mixture:

The Feature mixture database collects the features of the video frames. The feature mixture database contains the root keywords, semantic word set and the LGP vectors of the each and every video in the video database. The feature mixture database makes the classification process to be simpler and reduces the storage for the video retrieval. Figure 2 explains the feature mixture database.

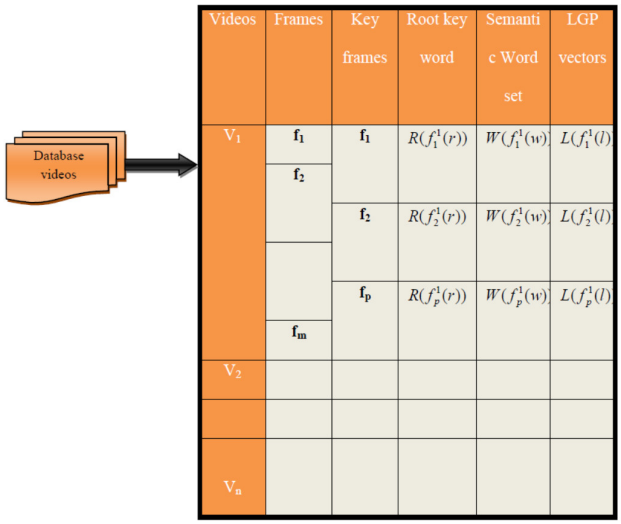


Figure 2 Feature Mixture Database

The Feature mixture database can be mathematically expressed by the equation 3.11. The term $F\left(V_i\right)$ in the equation 3.11 is given by the equation 3.12.

$$F_D=\left\{F\left(V_i\right);\quad 1\leq i\leq n\right\} \tag{11}$$

$$F\left(V_i\right)=\left\{R\left(f_j^i\right),W\left(f_j^i\right),L\left(f_j^i\right)\right\} \tag{12}$$

where $R\left(f_j^i\right)$ indicates the Root keyword of the j^{th} frame of the i^{th} video.

$W\left(f_j^i\right)$ indicates the Semantic word set of the j^{th} frame of the i^{th} video.

$L\left(f_j^i\right)$ indicates the LGP vectors of the j^{th} frame of the i^{th} video.

3.3 Hybrid KNB classifier for video retrieval

The Feature mixture database F_D provides the essential features required for the video retrieval process. Now, the database can be classified using the hybrid KNB classifier. The hybrid KNB classifier has the properties of both the Naive Bayes classifier and the K-NN classifier. The Naive Bayes classifier has the strong naive assumptions, and hence it is easy to build the classifier with the Bayesian statistics. The naive Bayes classifier has the probabilistic approach for the data classification and requires less training data for the classification process. The K-NN classifier performs the classification of the larger data set by finding the similarity between the data set.

3.3.1 Construction of the Hybrid KNB classifier:

The proposed hybrid KNB classifier performs the classification of the video files in the database based on the query given by the user. The query may be a text, image or the part of the video file. The equation 3.13 expresses the query from the user. The query when subjected to the feature extraction the root keywords, semantic words and the LGP vectors can be found.

$$Q = \{Q^r, Q^w, Q^l\} \tag{13}$$

where Q^r indicates the root keyword of the query.

Q^w indicates the semantic word of the query.

Q^l indicates the LGP vector of the query.

The query from the user is compared with the every video of the database to find the distance. The distance d is the measure of the distance between the query and the video V_i in the feature extraction database. The equation 3.14 provides the expression for the distance measure.

$$d = \{d_i; \quad 1 \le i \le n\} \tag{14}$$

where d_i = Distance $[Q, V_i]$

$$d_i = \frac{1}{K*3}\sum_{i=1}^{K}\left\{\left[\frac{Q^r \cap R(f_j^i)}{Q^r \cup R(f_j^i)}\right] + \left[\frac{Q^w \cap W(f_j^i)}{Q^w \cup W(f_j^i)}\right] + \left[\frac{Q^l - L(f_j^i)}{\kappa}\right]\right\} \tag{15}$$

where K indicates the constant whose value depends on the query. The distance measure is done by comparing every feature of the database with the query. The calculation of the distance measure between the query and the feature mixture database simplifies the retrieval process. Now, define a threshold value T. The threshold T depends on the query of the user. This threshold splits the database into two groups based on the distance measure d. The Feature mixture database is now expressed as follows,

$$F_D = \begin{Bmatrix} G_1 << \{V_i\}; & if\ (d_i > T), & \forall i \\ G_2 << \{V_i\}; & else. & \end{Bmatrix} \tag{16}$$

$$F_D = \{G_1, G_2\} \tag{17}$$

When the value of the distance is greater than the threshold T, then the Feature mixture database contains the videos of the G_1 otherwise, it has the videos of the G_2. To perform the video retrieval the mean and the variance value of the F_D has to be found. The expression of the mean and the variance

depends on the expression of the Naive Bayes classifier. The mean of the group G_1 and G_2 is expressed by equation 3.18 and 3.20. The mean of the group G_1 and G_2 is expressed by equation 3.19 and 3.21. The i value depends only on the G_1 if the mean is calculated for the group 1.

$$\mu_{G1} = \frac{1}{|G1|} \sum_{\substack{i=1, \\ i \in G1}}^{|G1|} d_i \tag{18}$$

$$\sigma_{G1} = \frac{1}{|G1|} \sum_{\substack{i=1, \\ i \in G1}}^{|G1|} (d_i - \mu_{G1})^2 \tag{19}$$

$$\mu_{G2} = \frac{1}{|G2|} \sum_{\substack{i=1, \\ i \in G2}}^{|G2|} d_i \tag{20}$$

$$\sigma_{G2} = \frac{1}{|G2|} \sum_{\substack{i=1, \\ i \in G2}}^{|G2|} (d_i - \mu_{G2})^2 \tag{21}$$

where μ_{G1} indicates the mean of the group 1

μ_{G2} indicates the mean of the group 2

σ_{G1} indicates the variance of the group 1

σ_{G2} indicates the variance of the group 2

$|G1|$ indicates the average distance of the group 1

$|G2|$ indicates the average distance of the group 2

The video retrieval from the group is done by calculating the posterior probability. The classification of the query is done finding the posterior probability of the group 1 and group 2. The posterior probability of the query with respect to the group 1 is given by the equation 3.22 and the equation 3.23 explains the posterior probability of the group 2.

$$posterior\,(Q\,with\,G_1) = \frac{1}{2\pi\sigma_{G1}^{\,2}} e^{-\frac{(d_i-\mu_{G1})^2}{2\sigma_{G1}^{\,2}}} \tag{22}$$

$$posterior\,(Q\,with\,G_2) = \frac{1}{2\pi\sigma_{G2}^{\,2}} e^{-\frac{(d_i-\mu_{G2})^2}{2\sigma_{G2}^{\,2}}} \tag{23}$$

The posterior probability value from the group 1 and group 2 are compared. The group with the highest posterior probability is more related to the query Q. The NB classifier retrieves the video contents from the group with the highest posterior probability. The retrieved video contents from the Naive Bayes Classifier V_R^{NB} is represented by equation 3.24,

$$V_R^{NB} = \left\{ V_i \in \min_{t=1}^{2} G_t \; with\, highest\; posterior\; probability \right\} \tag{24}$$

The K-NN classifier performs the classification of the videos of the database by assigning a weight 1/k to its nearest neighbors. The nearest neighbor videos in the group are found by using the distance measure d. The K-NN classifier retrieves the video from the database F_D which has the weight of 1/k. The value k is the weight of the K-NN classifier, and its value depends on the query. The retrieved video content from the K-NN classifier is given by the expression 3.25.

$$V_R^{KNN} = \left\{ V_i ; i \in K \quad K \text{ represents the nearest neighbours with weight } \frac{1}{k} \right\} \tag{25}$$

The proposed hybrid KNB classifier finds the retrieved lecture videos from the Naive Bayes and K-NN classifier. The final retrieved video content from the hybrid KNB classifier is the common video contents from the NB classifier and the K-NN classifier, and it is given by the equation 3.26.

$$V^R = V_R^{NB} \cap V_R^{KNN} \tag{26}$$

3.4 Pseudo code

Figure 3 explains the pseudo code for the proposed semantic-enriched lecture video retrieval system.

1	**Input**
2	Q = Query from the user;
3	V_i = Videos in the database D;
4	T=Threshold;
5	K= Weight of the K-NN classifier
6	**Begin**
7	Calculate distance d_i from Q to each V_i using equation
8	Calculate Feature Mixture Database F_D from equation
9	**If** (d_i > T)
10	$F_D = G_1$;
11	Calculate μ_{G1} and σ_{G1} from equation
12	Calculate $posterior(Q\ with\ G_1)$ from equation
13	**Else**
14	$F_D = G_2$;
15	Calculate μ_{G2} and σ_{G2} from equation
16	Calculate $posterior(Q\ with\ G_2)$ from equation
17	**End if**
18	**If**[$posterior(Q\ with\ G_1) > posterior(Q\ with\ G_2)$]
19	$V_R^{NB} = V_i ; \quad i \in G_1$
20	**Else**
21	$V_R^{NB} = V_i ; \quad i \in G_2$
22	**Endif**
23	**Call K-NN classifier**
24	Calculate V_R^{KNN} from the equation
25	Calculate V_R from the equation
26	**Return** V_R
27	End

Figure 3 Pseudo code

4 Results and discussion

This section discusses the results obtained through the semantic enriched video retrieval system using the hybrid KNB classifier. The performance metrics such as Precision, Recall and the F-measure determines the efficiency of the proposed model. The performance of the model is compared with the existing K-NN classifier model.

4.1 Experimental set up

The experimentation of the lecture video retrieval is done in the Personal Computer with the Intel I3 processor containing the 4GB Ram and the Windows 8 operating system. The entire work is implemented using the MATLAB tool.

a) Dataset description

This research work includes the various video samples for the experimentation. The video from the fields of Agriculture, Image Processing, India 20-20, Quantum optics and Networking were taken as the samples for the lecture video retrieval. The total sum of 50 video samples was taken. The experimentation of the proposed model was done by giving the text query and the video query.

b) Evaluation metrics

The performance parametric such as Precision, Recall, and F-measure analyzes the efficiency of the proposed model. The performance metrics in the proposed system is explained as follows,

i) Precision:

Precision defines the ratio of the total number of the retrieved video contents which matches the query to the sum of the relevant videos matching and not matching the query from the video database. The precision can be mathematically expressed as follows,

$$\Pr ecision = \frac{V_R^m \cap V^T}{V^T} \tag{27}$$

where, V_R^m represents the retrieved videos from the database

V^T represents the relevant videos present in the database

ii) Recall:

Recall defines the ratio of the total number of the retrieved video contents which matches the query to the sum of the retrieved videos matching the query in the video database. Equation 4.2 defines the mathematical expression for the Recall parameter.

$$\mathrm{Re} call = \frac{V_R^m \cap V^T}{V_R^m} \tag{28}$$

iii) F-measure:

F-measure defines the harmonic mean of the precision and the recall parameters. The F-measure is expressed as follows,

$$F - measure = \frac{2 * \Pr ecision * \mathrm{Re} call}{\Pr ecision + \mathrm{Re} call} \tag{29}$$

c) Methods taken for Comparison

The proposed hybrid KNB classifier model has been compared with the conventional model such as the K-NN classifier for the video retrieval. The performance of the model was analyzed by calculating the performance metrics such as Precision, Recall and the F-measure. The proposed model performs the video retrieval with the combined properties of the K-NN and the NB classifier.

4.2 Experimental results

The simulation of the proposed model uses the 50 video samples from the various fields. The video samples have various frames are subjected to the feature extraction and given to hybrid KNB classifier. The performance of the model is analyzed by giving the text query and the video query. Figure 4 and 5 explains the retrieved video contents from the text query and the video query.

a) Text Query b) Retrieved videos

Figure 4 Video retrieval with the Text Query

(a) Query **(b) Retrieved output**

Figure 5 Video retrieval with the Video Query

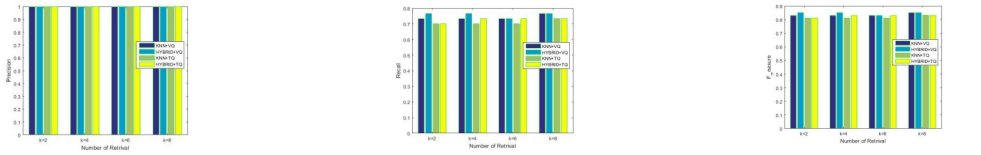
4.3 Performance evaluation

This section discusses the performance evaluation of the proposed model by comparing it the existing K-NN classifier model for the video retrieval.

4.3.1 Analysis of 60% of database

In this set, the performance evaluation is done by considering any three categories of the videos from the database. Figure 6 explains the change in the performance metric values for the various K (number of retrievals) values. The graph is drawn between the performance metrics and the number of retrievals. For the K value as 2, when the text query is given by the user, the precision value of the

Hybrid KNB classifier and the K-NN classifier remain same. The precision is nearly perfect with the value of 1. The recall measure for the hybrid KNB classifier is 0.7333, and the F-measure value is 0.8301. When the video query is used, the proposed model outperforms the K-NN classifier. The recall value is improved by 0.7667, and the F-measure is improved by 0.8519. Thus for the video retrieval from the database with the three categories of the video, the proposed model has better performance than the K-NN model when the video query is used.

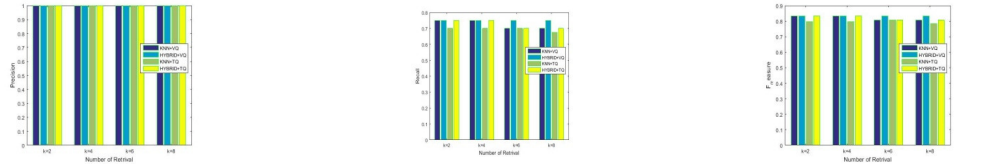


(a) Precision metric **(b) Recall metric** **(c) F-measure metric**

Figure 6 Performance of the Hybrid KNB classifier when analyzing 60% of database videos

4.3.2 Analysis of 80% of database

In this set, the performance evaluation is done by considering any four categories of the videos from the database. Figure 7 explains the change in the performance metric values for the various K (number of retrievals) values. For the K value as 2, when the text query is given by the user, the precision value of the Hybrid KNB classifier and the K-NN classifier remain same. The recall measure for the hybrid KNB classifier is 0.7500 which is better than the K-NN model. The F-measure value of the hybrid KNB has the higher value of 0.8333 than the K-NN. When the text query is used, the proposed model outperforms the K-NN classifier. For the video query, the recall value remains the same as the K-NN classifier, and the F-measure value is 0.8056. Thus for the video retrieval from the database with the four categories of the video, the proposed model has better performance than the K-NN model when the text query is used.

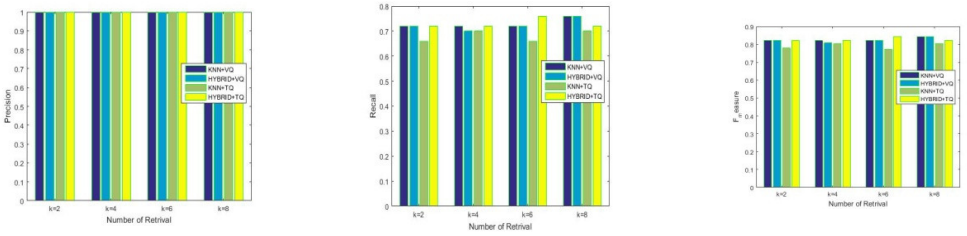


a) Precision metric **b) Recall metric** **c) F-measure metric**

Figure 7 Performance of the Hybrid KNB classifier when analyzing 80% of database videos

4.3.3 Analysis of 100% of database

In this set, the performance evaluation is done by considering any five categories of the videos from the database. Figure 8 explains the change in the performance metric values for the various K (number of retrievals) values. For the K value as 2, when the text query is given by the user, the precision value of the Hybrid KNB classifier and the K-NN classifier remain same. The recall measure for the hybrid KNB classifier is 0.7200 which is better than the K-NN model. The F-measure value of the hybrid KNB has the higher value of 0.8222 than the K-NN. When the text query is used, the proposed model outperforms the K-NN classifier. For the video query, the recall value remains the same as the K-NN classifier, and the F-measure value is 0.8222. Thus for the video retrieval from the database with the four categories of the video, the proposed model has better performance than the K-NN model when the text query is used.



a) Precision metric **b) Recall metric** **c) F-measure metric**

Figure 8 Performance of the Hybrid KNB classifier when analyzing 100% of database videos

4.4 Discussion

Table 1 compares the performance of the proposed system with the existing K-NN model. The performance metrics precision, recall, and the f-measure analyze the efficiency of the proposed model. From the Table 4.1, the precision of the proposed model and the K-NN classifier remains the same. When k=2 and the text query is used, the proposed hybrid KNB classifier has the better performance with higher recall value of 0.7500 than the K-NN model. The f-measure value of the proposed system is higher than the K-NN model for the various k values when the text query is used. The proposed system shows a better performance when the video query is used. For various the k values, the performance metrics have achieved better values than the existing methods. Thus, the proposed model has clearly outperformed the existing system with the better retrieval of the video lectures from the database.

Table 1 Comparison between the Hybrid KNB and the K-NN with the text query and the video query

	Performance metric	Text Query		Video Query	
		K-NN	Hybrid KNB	K-NN	Hybrid KNB
Analysis of 60% of database	Precision	1	1	1	1
	Recall	0.7333	0.7333	0.7333	0.7667
	F-measure	0.8301	0.8301	0.8301	0.8519
Analysis of 80% of database	Precision	1	1	1	1
	Recall	0.7000	0.7500	0.7500	0.7500
	F-measure	0.8000	0.8333	0.8056	0.8056
Analysis of 100% of database	Precision	1	1	1	1
	Recall	0.6667	0.7200	0.7200	0.7200
	F-measure	0.8000	0.8222	0.8222	0.8222

5 Conclusion

In this paper, a novel method to address the problems in the lecture video retrieval has been discussed. The proposed semantic enriched video retrieval system uses the feature mixture database and the hybrid KNB classifier. In this paper, a new classifier called Hybrid KNB has been introduced for the video retrieval. The hybrid KNB classifier has the properties of the NB classifier and the K-NN classifier. The proposed classifier has the better accuracy and less complexity since it depends on the Naive Bayes assumption. The performance of the proposed system is analyzed by implementing the system in the database containing 50 videos of different categories. The performance metrics such as precision, recall and the f-measure analyze the efficiency of the system. Simulation is done by giving the text query and the video query. The simulation results show that the proposed model has better precision value of 1.0 and the recall value of 0.7500. The f-measure of the proposed model has a better value of 0.8519 than the existing system. Thus the proposed model with the improved metrics has better performance than the other conventional systems for video retrieval. In the future, this work can be extended by adding other features such as local span, cue words, etc.

REFERENCES

[1]. Manish Kanadje, Zachary Miller, Anurag Agarwal, Roger Gaborski, Richard Zanibbi and StephanieLudi, "Assisted keyword indexing for lecture videos using unsupervised keyword spotting," Pattern Recognition Letters, vol. 71, pp. 8-15, 2016.

[2]. Vidhya Balasubramanian, Sooryanarayan Gobu Doraisamy and Navaneeth Kumar Kanakarajan, "A multimodal approach for extracting content descriptive metadata from lecture videos," Journal of Intelligent Information Systems, vol. 46, no. 1, pp. 121-145, 2016.

[3]. Haojin Yang and Christoph Meinel, "Content Based Lecture Video Retrieval Using Speech and Video Text Information," IEEE Transactions on Learning Technologies, vol. 7, no. 2, pp. 142-154, 2014.

[4]. Kai Li, Jue Wang, Haoqian Wang and Qionghai Dai, "Structuring Lecture Videos by Automatic Projection Screen Localization and Analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 6, pp. 1233-1246, 2015.

[5]. Ruben Fernandez-Beltran and Filiberto Pla, "Incremental probabilistic Latent Semantic Analysis for video retrieval", Image and Vision Computing, vol. 38, pp. 1-12, 2015.

[6]. Sara Memar, Lilly Suriani Affendey, Norwati Mustapha, Shyamala C. Doraisamy and Mohammadreza Ektefa, "An integrated semantic-based approach in concept based video retrieval", Multimedia Tools and Applications, vol. 64, no. 1, pp. 77-95, 2013.

[7]. Stevan Rudinac, Martha Larson and Alan Hanjalic, "Leveraging visual concepts and query performance prediction for semantic-theme-based video retrieval", International Journal of Multimedia Information Retrieval, vol. 1, no. 4, pp. 263-280, 2012.

[8]. Nhu Van Nguyen, Mickal Coustaty and Jean-Marc Ogier, "Multi-modal and cross-modal for lecture videos retrieval," In Proceedings of IEEE 22nd International Conference on Pattern Recognition, pp. 2667-2672, 2014.

[9]. Hyun Ji Jeong, Tak-Eun Kim, Hyeon Gyu Kim, and Myoung Ho Kim, "Automatic detection of slide transitions in lecture videos," Multimedia Tools and Applications, vol. 74, no. 18, pp. 7537-7554, 2015.

[10]. Haojin Yang, Maria Siebert, Patrick Lühne, Harald Sack and Christoph Meinel, "Lecture Video Indexing and Analysis Using Video OCR Technology," In Proceedings of IEEE Signal-Image Technology and Internet-Based Systems (SITIS), pp. 54-61, 2011.

[11]. Lecia Barker, Christopher Lynnly Hovey and Jaspal Subhlok, Tayfun Tuna, "Student Perceptions of Indexed, Searchable Videos of Faculty Lectures," In Proceedings of IEEE Frontiers in Education Conference (FIE), pp. 1-8, 2014.

[12]. Tayfun Tuna, Jaspal Subhlok and Shishir Shah, "Indexing and Keyword Search to Ease Navigation in Lecture Videos," In proceedings of IEEE Applied Imagery Pattern Recognition Workshop, pp. 1-8, 2011.

[13]. Bailan Feng, Juan Cao, Xiuguo Bao, Lei Bao, Yongdong Zhang, Shouxun Lin and Xiaochun Yun, "Graph-based multi-space semantic correlation propagation for video retrieval," The Visual Computer, vol. 27, no. 1, pp. 21-34, 2011.

[14]. Vasconcelos N, Lippman A, "Towards semantically meaningful feature spaces for the characterization of video content," In Proceedings of IEEE International conference on image processing, Computer Society, 1997.

[15]. Ali Shariq Imran, Laksmita Rahadianti, Faouzi Alaya Cheikh and Sule Yildirim Yayilgan, "Objective Keyword Selection for Lecture Video Annotation," In Proceedings of European Workshop on Visual Information Processing (EUVIP), pp. 1-6, 2014.

[16]. Karl K. Szpunar, Helen G. Jing, Daniel L. Schacter, "Overcoming overconfidence in learning from video-recorded lectures: Implications of interpolated testing for online education," Journal of Applied Research in Memory and Cognition, vol. 3, no. 3, pp. 161-164, 2014.

[17]. Ankush Mittal and Sumit Gupta, "Automatic content-based retrieval and semantic classification of video content," International Journal on Digital Libraries, vol. 6, no. 1, pp. 30-38, 2006.

[18]. Sara Memar, Lilly Suriani Affendey, Norwati Mustapha and Mohamamdreza Ektefa, "Concept-based Video Retrieval Model Based on The Combination of Semantic Similarity Measures," In Proceedings of IEEE International Conference on Intelligent Systems Design and Applications, pp. 64-68, 2013.

[19]. Dianting Liu and Mei-Ling Shyu, "Semantic Retrieval for Videos in Non-Static Background Using Motion Saliency and Global Features," In Proceedings of IEEE International Conference on Semantic Computing, pp. 294-301, 2013.

[20]. Ali Shariq Imran, Alejandro Moreno and Faouzi Alaya Cheikh, "Exploiting Visual Cues in Non-Scripted Lecture Videos for Multi-modal Action Recognition," In Proceedings of IEEE International Conference on Signal Image Technology and Internet Based System, pp. 8-14, 2012.

[21]. Poonam Yadav, "Annotation of web pages using semantic tagging and ranking model to effective information retrieval," International Journal of Computer Science & Engineering Technology, Volume 5, Issue 12, pp 1094-1098, 2014.

[22]. SKR P. Vijaya, G. Raju, “An Ontology-Based Meta-Search Engine for Effective Web Page Retrieval,” International Review of Computers and Software(IRECOS), Volume 8, Issue 2, pp 533-541, 2013.

[23]. Ray Smith, "An Overview of the Tesseract OCR Engine," Ninth Int. Conference on Document Analysis and Recognition (ICDAR), IEEE Computer Society, pp. 629-633, 2007.

[24]. Arun Balagopalan, Lalitha Lakshmi Balasubramanian, Vidhya Balasubramanian, Nithin Chandrasekharan, and Aswin Damodar, “Automatic keyphrase extraction and segmentation of video lectures,” IEEE International Conference on Technology Enhanced Education (ICTEE), pp. 1 – 10, 2012.

[25]. Che-Yu Yang and Hua-Yi Lin, “An automated semantic annotation based on WordNet ontology,” International Conference on Networked Computing and Advanced Information Management, pp. 682 – 687, 2010.

[26]. Gangemi A, Navigli R, and Velardi P, “The OntoWordNet Project: Extension and Axiomatization of Conceptual Relations in WordNet,” International Conference on Ontologies Databases and Applications of Semantics (ODBASE 2003), Catania, Sicily, Italy, pp. 820–838, 2003.

[27]. V. Snasel, P. Moravec, and J. Pokorny, “WordNet Ontology-Based Model for Web Retrieval,” International Workshop on Challenges in Web Information Retrieval and Integration, pp. 220 – 225, 2005.

[28]. Kaveh Samiee, Peter Kovács, and Moncef Gabbouja, “Epileptic seizure detection in long-term EEG records using sparse rational decomposition and local Gabor binary patterns feature extraction,” Knowledge-Based Systems, Volume 118, pp. 228–240, 2016.

[29]. A. Sharmila and P. Geethanjali, “DWT Based Detection of Epileptic Seizure from EEG Signals Using Naive Bayes and k-NN Classifiers,” Volume 4, pp. 7716 – 7727, 2016.

[30]. B. S. Daga and A. A. Ghatol, "Multicue Optimized Object Detection for Automatic Video Event Extraction", Indian Journal of Science and Technology, Vol. 9, no. 47, December 2016.

[31]. B. V. Patel, B. S. Daga, B. B. Meshram, "Building Multimedia Applications", B. V. Patel, B. S. Daga, B. B. Meshram, International journal on computer engineering and information technology, vol. 14, no. 19, pp. 10-15, 2010.

[32]. Brijmohan Daga, "Content based video retrieval using color feature: an integration approach", in proceedings of the International Conference on Advances in Computing, Communication, and Control, pp 609-625, 2013.

[33]. Brijmohan Daga, Avinash Bhute, Ashok Ghatol, "Implementation of Parallel Image Processing using NVIDIA GPU Framework", in proceedings of the International Conference on Advances in Computing, Communication and Control, Vol. 125, pp 457-464, 2011.

.