RETRIEVING OF VIDEO SCENES USING ARABIC CLOSED-CAPTION

H. M. Nassar[1], A. Taha[1], T. M. Nazmy[2] & K. A. Nagaty[2]

[1] *Faculty of Computesr and Informatics, Suez Canal University*
[2]*Faculty of Computer and Information Sciences, Ain Shams University*
*email:*ah_twins@hotmail.com

Abstract

The increased use of video documents for multimedia-based applications has created a demand for strong video database support, including efficient methods for browsing and retrieving video data. Most solutions to video browsing and retrieval of video data rely on visual information only, ignoring the rich source of the accompanying audio signal and texts. Speech is the significant information that has a close connection to video contents. The closed-caption text facilitates the acquisition of the video transcript. This paper proposes a novel scheme for retrieving video scenes by exploring the Arabic closed-caption text. The new approach classifies the input query and retrieves all the video scenes with the same category of the input query using a text-based video scenes classification technique. Then, a caption-based filtering technique is utilized to determine the relevant scenes. The proposed system is implemented and tested using a self collected and prepared dataset. Experimental results show that the proposed approach achieves an average precision equal to 88.94% and average recall equal to 93.93% for video scenes retrieval.

Keywords

Content-Based Video Retrieval, Video Data Retrieval, Arabic Closed-Caption, Video Content Analysis.

1. Introduction

Video data is a rich and expressive medium and its availability has been recently increased by digital technology. The developments in visual information technology have enabled users to view large amount of video data. The effective usage of these videos depends on the development of systems supporting effective retrieval, thus preventing users from being confused with videos that do not match their needs. However, tools for facilitating search and retrieval of these video data are still limited. Retrieval of relevant videos from large video databases has immense applications in various domains such as large distributed digital libraries, broadcasting or production archives and video databases [3].

Content-based video retrieval is an active area of research with continuing attributions from several domain including image processing, computer vision, database system and artificial intelligence [5]. Video indexing is the basic step for preparing video collections for retrieval. Traditional approaches for video indexing are based on representing the video content using a set of low-level features such as color, texture and shape. Such strategy, however, is ineffective since the low-level features do not reflect well the semantic content of the video data [1]. Also, these techniques often do not utilize the commonly available high-level information conveyed by textual captions [2]. The presence of textual captions and audio can provide complementary semantic cues that are very useful in retrieving video databases [3].

Many ways of querying a video retrieval system have been used. One way is to use example images as it is popular for content-based image retrieval. However, this way has a limitation that it does not utilize the motion information of the video and employs only the appearance information. Another way is to use example video clip but it is quite complex for many applications to find example video clips for the concept of interest [9]. Textual query is a promising approach for querying in video databases, since it offers a more natural interface [3].

Also, when querying a video archive, the users are interested in retrieving precise answers that best answers the query. However, current video retrieval systems, including the search engines on the web, are designed to retrieve either a pre-defined summary, or the whole video sequence. This is unsatisfactory as it often takes a long time for the users to locate where the information is in the video stream. Ideally, the system should return a reasonably short video segment to provide the requested information [4].

In this paper, a new approach for video scene retrieval is proposed. The proposed approach bridges the semantic gap for multimedia content analysis and retrieval by utilizing the Arabic closed-caption text. Obviously, the speech of the video contains more detailed semantic information than the visual information of the video image. The closed-caption is the video transcript that represents the spoken dialogue or narration and it includes also sound effects, speaker identification, music, and other non-speech information. Our new approach is also designed to retrieve video scenes not the whole video document hence; the users will have a precise answer to their queries. Moreover, Arabic language has a complex morphology and needs much processing to be convenient for retrieval [18].

This paper is organized as follows: In Section 2 some related work is presented. Section 3 discusses some issues related to Arabic text retrieval. In Section 4, the proposed approach for retrieving video scenes is presented. The experimental results to evaluate the proposed approach are reported in Section 5. Finally, the conclusion is presented in Section 6.

2. Related Work

Many methods have been developed for the content-based video retrieval task. These methods are based on the different design aspects [7]. In the literature, there are three main approaches in indexing and retrieval of videos in multimedia databases. The first one is the content-based indexing in which videos are indexed and retrieved using spatial and temporal characteristics of the video. In spatial domain, feature vectors are calculated from different parts of the frames and their relationship is encoded as a descriptor. In temporal domain, the video is partitioned into its basic elements like frames, shots or scenes. Each video segment is then characterized by the appearance of the video content [3].

Many works were proposed using the first approach. Lin et al presented in [2] a scheme for content-based video retrieval by exploring the spatio-temporal information. They developed two content descriptors to measure video content variations and represent subshots: dominant color histogram (DCH) and spatial structure histograms (SSH). Also, DeMenthon and Doermann presented a methodology for a compact representation of a video's spatio-temporal structure in which videos are analyzed to produce spatio-temporal descriptors that summarize the location, color and dynamics of independently moving regions with only a small number of bytes. The similarities of sequences are defined using these descriptors [10]. Nakamura et al also presented a video scene retrieval system using the layerization of images and the stochastic attributed relational graph (ARG) matching method [12].

The second approach is the keyword-based indexing which uses descriptive text or keywords to index video documents. Retrieval in this approach is performed by matching the query, given in the form of keywords, with the stored keywords. One way is to manually annotate the video by textual information describing its content. However, this approach is not satisfactory, because the text-based description tends to be incomplete, imprecise, and inconsistent in specifying visual information. Another way is to analyze video frames using OCR systems to extract text string contained in the video image and use this text in the indexing process [8]. However, this way is not good enough because not all videos contain text suitable for retrieval. Finally, the speech recognition systems are used to obtain the audio content of the video and use it in indexing.

Jawahar et al presented an approach that enables search based on the textual information present in the video [3]. In their approach, regions of textual information are first identified within the frames of the video. Video is then annotated with the textual content present in the images. Also, Lai et al proposed a relevance ranking approach for Web-based video search using both video metadata and the rich content contained in the videos such as semantic descriptions and speech [11].

The third approach is the multimodal-based indexing which is a combination of the above two approaches. Nevertheless, the use of different components to represent the video still has its problems. Like many other data fusion problems, video retrieval using this approach depends to a large extent on the weights assigned to each feature under consideration [8].

Zhu and Zhou developed a system for content-based video browsing and retrieval [8]. The system integrates the audio-visual as well as text information and natural language understanding technique to extract scenes and content information of video documents and to organize and classify video scenes. Also, Zhai et al presented an integrated system for news video Retrieval [7]. Their system incorporates both speech and visual information in the search mechanisms.

3. Arabic Text Retrieval

Arabic is one of the six official languages of the United Nations and it is the mother tongue of 300 million people [13]. It belongs to the Semitic language group which also includes Hebrew and Aramaic. Unlike Latin-based alphabets, Arabic is written from right-to-left. The Arabic alphabet consists of 28 letters and it can be extended to ninety by additional shapes, marks, and vowels [19].

The grammatical system of Arabic language is based on a root-and-pattern structure. Therefore, it is considered as a root-based language with not more than 10000 roots and 900 patterns [14]. The root is the bare verb form and it can be tri-literal, which is the majority of words (85%), and to a lesser extent, quad-literal, pen-literal, or hexa-literal, each of which generates different forms of verbs and nouns by the addition of derivational affixes (i.e. prefixes and suffixes) [14].

Arabic is a challenging language for information retrieval because of many reasons. In what follows, some Arabic characteristics are listed that may affect the Arabic information retrieval [14, 15, 17]:

1. Some Arabic letters have different orthographic forms. For example, the Arabic letter Alef (ا) can be written in different ways (آ - إ - أ).

2. Irregular plurals are more common in Arabic than English. The plural form might be produced by the addition of suffixes, prefixes or infixes, or by a complete reformulation of the word depending on the root and the singular form of the word. For example, the word (نساء) which means women is an irregular plural for the word (امرأة) which means woman.
3. A given Arabic word can be found in huge number of different forms. Many definite articles, conjunctions, particles and other prefixes can attach to the beginning of a word, and large numbers of suffixes can attach to the end. However, in English they appear in separable form. For example, the Arabic word (وشربتها) means "and she drank it".
4. Unlike English, Arabic has a form called “dual” that indicates two of something. For example, a student is called (طالب), two students are (طالبان), and students are (طلاب).
5. Arabic nouns are either feminine or masculine as in French, Spanish, and many other languages, and the verbs and adjectives that refer to them must agree in gender. For example, (معلم) and (معلمة) are the masculine and feminine form of the English word teacher.

Those problems make exact keyword match insufficient for Arabic retrieval [15]. This profusion of forms means a greater likelihood of mismatch between the form of a word in a query and the forms found in the searched documents. Spelling normalization and stemming are well-known techniques that can significantly improve retrieval. Many algorithms are proposed for Arabic stemming and normalization [13, 14, 17, 18].

4. Proposed Approach

The proposed approach is built on the idea that the embedded text in the video document is very useful in video analysis, especially for high-level semantic content analysis. Text information can be obtained by speech recognition if noise can be negligible and it can be extracted directly from key frames with video Optical Character Recognition (OCR). Currently, the availability of closed-caption text facilitates the acquisition of the video transcript. The closed-caption text is simply the speech transcript of the video document. Hence, it contains more semantic and rich information than other video components and it is more beneficial than extracting text strings that may appear in the video image. Thus, the proposed approach takes the advantage of semantically rich information found in the closed-caption text to retrieve video scenes. More specifically, it works on the Arabic closed-caption. Arabic is not lucky as English and other European languages in the research. There are relatively few studies on the retrieval of Arabic documents although it needs more efforts. Also, one of the important strengths of the proposed approach is that it can be used to search for a video scene in either a single video document or a large digital video collection.

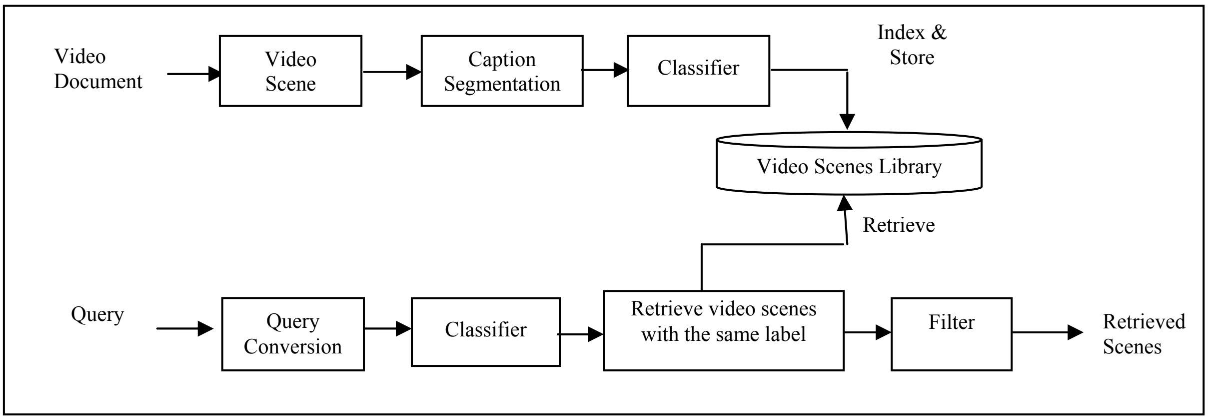

Figure 1. The general block diagram of the proposed approach

Figure 1 shows the proposed approach and its individual components interacting together. The proposed approach consists of two processes: the indexing process and the retrieving process. In the indexing process, the video document is first segmented into a set of scenes using a video scene detection algorithm. Then, the closed caption text of the video document is segmented to extract the caption of each detected scene. After that, each video scene extracted in the scene segmentation process is classified into predefined semantic categories. This will provide users a table of content similar to that of printed books to facilitate fast navigation and retrieval. For this purpose, a text-based video classification algorithm using Arabic closed-caption has been developed and integrated into our system [20]. It is used to automatically classify video scenes on the basis of the texts obtained from closed caption. Finally, each video scene is indexed and stored in the video scenes library with its class to be ready for retrieval.

Figure 2 shows the block diagram of the retrieval process of the proposed approach. The approach is a highly modularized system with four primary functional modules: the query conversion module, the classifier module, scenes retrieving module, and closed-caption based filtering module.

When the user enters a textual query either by writing some keywords or by inputting the closed-caption text of an example video scene, the query conversion module is responsible for processing the query and extracting the feature vector. Then, the feature vector is passed to the classifier to determine its semantic category like politics, economics, religions…etc. After the category is determined, the video scenes having the same category are retrieved from the video scenes library. Those scenes need some filtering process to determine the more relevant scenes to the input query and narrow the result space (search focusing). It is the task of the filtering module which utilizes the closed-caption text of the retrieved scenes to determine the final matched scenes. The four modules of the proposed approach are presented in more details in the following subsections:

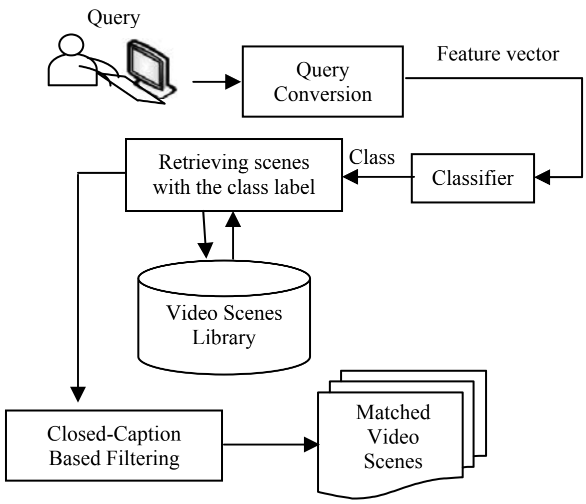

Figure 2. The block diagram of video retrieval process

4.1 Query Conversion Module

This module is in charge of processing the user-query and converting it to a feature vector that passes to the classifier. Figure 3 shows a diagram for the query conversion module and its components.

The first step is the keyword extraction in which keywords are extracted from the query by removing the numbers and words less than three letters where these words In Arabic language have no effect in the classification process [20].

The second step is the spelling normalization of the extracted keywords. As mentioned before, Some Arabic letters have different orthographic forms which can affect the retrieval results. Therefore, there is a need to unify different forms of the same letter. Some minor replacements of certain Arabic characters are performed on the Arabic words to facilitate the retrieval process. For example, the Arabic word (القاضى) which means the judge to (القاضي) and the word (الحرية) which means the freedom to (الحريه). The proposed approach performs the spelling normalization as follows:

- Replace أ , إ ,and آ with ا
- Replace ـة with ـه
- Replace ى with ي
- Replace ىء , and يء with ئ

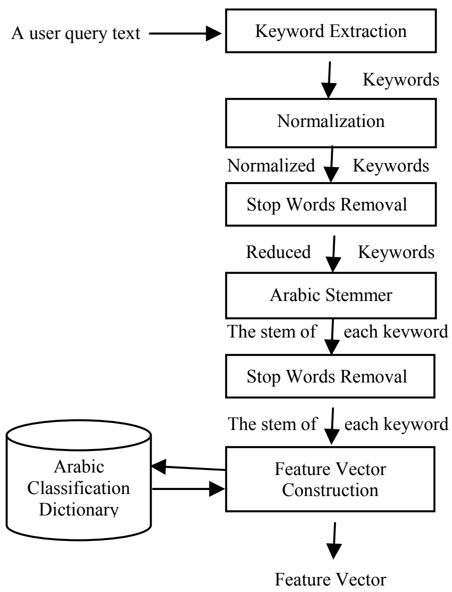

Figure 3. The query conversion module

The third step is the removal of stop words which are those words that do not carry a particular and useful meaning for the retrieval process. These words are sometimes called functional words and removing them has a significantly positive effect in reducing the time needed for classification and retrieval. We set up a stop list, which contains 312 Arabic words (including pronouns, prepositions, adverbs, articles, question words…etc). Some examples of these Arabic stop words are shown in Table1.

The fourth step in this module is the stemming which is the process of removing the affixes from the word and extracting the word root. In our proposed approach, the light stemming is used since many researchers proved that it is the best effective stemming method for languages with more complex morphology like Arabic [13, 14, 17, 18]. The aim of this technique is not to produce the linguistic root of a given Arabic surface form, rather is to strip off the most frequent suffixes and prefixes without trying to deal with infixes or recognize patterns and find roots [14]. Table 2 shows the list of the prefixes and suffixes that the Arabic stemmer removes them while Table 3 shows some examples of Arabic words and their light stems.

Table 1. Some examples of Arabic stop words

Arabic Word	English Meaning	Type
أين	where	Question word
منذ	since	Adverb
أنت	you	Pronoun
ثم	then	Article
على	on	Preposition

Table 2. The list of the prefixes and suffixes removed by the Arabic stemmer of the proposed approach

Arabic Prefixes	والـ - فالـ - كالـ - بالـ - الـ - للـ - ل – ف – ك – ب – و
Arabic Suffixes	هما - كما - كن - كم - وا - نا - ها - هن - هم - ين - ون - ان – تى

Table 3. Examples of light stems of Arabic words

Arabic Word	English Meaning	Prefix	Suffix	Light Stem
المخترعين	the inventors	الـ	ين	مخترع
كالطير	like the bird	كالـ	-	طير
طالبات	students	-	ات	طالب
بكتابهم	in their book	بـ	هم	كتاب
قلمكم	your pen	-	كم	قلم

After word stemming is performed, some stop words that were not detected in the third step may appear and hence they are removed in the fifth step. For example, the Arabic word (كأنها) which means like has the prefix (كـ) and the suffix (ـها). The existence of these affixes makes the word not detected as a stop word in the third step but after applying the light stemming, the produced stem will be (أن) which is one of the stop words. Hence, stop words removal is required after the stemming.

The last step in this module is the construction of the feature vector. The number of components of the feature vector is equal to the number of the predefined semantic categories used in the classification process. In our experiments, eight semantic categories are identified: politics, economics, sports, religion, social, tourism, weather, and health. Each component in the feature vector is the percentage of the words found in the query belonging to the corresponding category. In other words, it represents the membership probability for each category. An Arabic classification dictionary is used to compute this percentage [20]. Each Arabic word is stored in the dictionary with its classification after applying a light stemmer. The word is classified to one or more category according to its meaning. Moreover, the Arabic classification dictionary is incrementally updated.

4.2 The Classifier Module

There are many classifiers used in text categorization such as Naïve Bayes probabilistic classifiers, Decision Tree classifiers, Regression methods, Neural Networks, K-Nearest Neighbor (k-NN) classifiers, Support Vector Machine (SVMs), and Rocchio classifiers. The k-NN and Rocchio classifiers are the two most frequent classifiers used in text categorization and they are both similarity-based [21].

The Rocchio classifier forms a simple representation of each class called the centroid or the prototype vector. This vector is usually calculated by averaging the training samples for each class during the learning process. Classification is based on similarity to or distance from centroid or prototype which can be determined using Euclidean distance, cosine similarity, dot/scalar product, or any other distance metric. When given an unknown sample, the Rocchio classifier calculates the distance between the unknown sample and the prototype vector for each category. The unknown sample is assigned the category label that has the minimum distance to its prototype vector. Figure 4 shows the Rocchio classification algorithm.

Training:

For each class c_j

Compute centroid vector $\vec{\mu}(c_j)$

Where $\vec{\mu}(c) = \frac{1}{|c|} \sum_{\vec{x} \in C} \vec{X}$

Testing:

Assign document $\vec{d}$ to class c

Where $c = \arg\min_{c_j} \| \vec{\mu}(c_j) - \vec{d} \|$

Figure 4. The Rocchio classifier

In our proposed approach, the Rocchio classifier is used because in contrast to other classifiers like K-nearest neighbor, Rocchio classifier does not require much storage for storing the training samples because it replaces the whole training samples by generalized prototype vectors. Also, those generalized prototype vectors make the classifier robust towards any noise that may occur in the training samples. Moreover, it is a fast classifier because the computational classification cost is low and the learning is done offline before receiving any unknown sample.

4.3 Scenes Retrieving Module

After the feature vector of a given query is constructed and classified into a semantic category. The proposed system retrieves all the video scenes that belong to the same category of the classified feature vector from the video scenes library. This library archives video scenes according to their category. A text-based video classification system using Arabic closed-caption has been developed and integrated into our system to classify video scenes [20]. When adding a new video document to the library, it is first segmented into scenes then each scene is classified with our system to be stored in the video scenes library. This design of the proposed approach allows the system to support category-based browsing that enables users to browse and retrieve video scenes using their categories. It should be mentioned that the proposed approach can be used to search for a video scene not only in a large digital video collection but also in a single video document. In this case, the scene retrieving module retrieves all the video scenes from the same video document instead of video scenes library.

4.4 Closed-Caption Based Filtering Module

In order to focus the search result, the proposed approach utilizes the closed-caption text for each retrieved video scene to determine the relevant scenes. It makes some kind of word matching but after stemming. It compares the light stems of the query keywords with the light stems of closed-caption keywords of each retrieved scene. For each retrieved video scene, it counts the number of stems in the closed caption of the scene that match the stems of the query keywords and compute the following Relevancy Factor (RF):

$$RF = \frac{\text{No of matched stems between the query \& the scene}}{\text{No of stems found in the query}}$$

As the value of the factor tends to one, the relevancy of the scene to the query is increased. After the factor is computed for each scene, the proposed approach returns the most relevant scenes which have high relevancy factor.

4.5 An Illustrative Example

Consider the Arabic text-query "مباريات كرة القدم في دورة الألعاب الآسيوية" which means "Football matches in Asian sports tournament". When this query is submitted to the proposed system, it is first passed to the query conversion module which extracts the keywords and removes the useless words. Here, the word (في) is removed because its letters are less than three. Then, the extracted keywords are normalized and become "مباريات كره القدم دوره الالعاب الاسيويه". Note that the query does not contain any stop word to be removed so the stemming is performed directly and the classes are identified for each produced stem using the Arabic classification dictionary. Each Arabic word may belong to more than one category depending on its meaning that is why some words are counted in more than one category.

Table 4. The construction of feature vector of the query

Category	# of words belonging to each category	Component value
Politics	1	1/5 = 0.2
Economics	1	1/5 = 0.2
Sports	5	5/5 = 1.0
Religion	0	0/5 = 0.0
Social	2	2/5 = 0.4
Tourism	0	0/5 = 0.0
Weather	0	0/5 = 0.0
Health	0	0/5 = 0.0

Table 4 shows the computations of each component in the feature vector of the user query. In order to calculate the first component, the number of words belonging to the politics category (1) is divided by the total number of words found in the Arabic classification dictionary (5) which gives (0.2). The rest of the feature vector's components are calculated using the same way and the vector becomes (0.2, 0.2, 1, 0, 0.4, 0, 0, 0). Then the classifier module uses this vector and classifies the query. The proposed approach classifies this query as sports. The retrieval module retrieves all the video scenes stored in the library with class sports (about 87 scenes). After that, the filtering module begins to determine the most relevant scenes to the query by computing the relevancy factor for each scene. In this example, the number of retrieved scenes is 46 scenes (1 scene with relevancy factor equal to 1, 12 scenes with relevancy factor from 0.5 to less than 1 and 33 scenes with relevancy factor greater than 0 and less than 0.5).

5. Experimental Results

One of the major limitations for Arabic information retrieval is the lack of adequate resources that could help in testing the system to get good evaluation of the system performance [19]. The only large scale resources known and available for researchers for Arabic datasets are the Linguistic Data Consortium (LDC) collection. However, it provides only speech and text databases not video databases. On the other side, since 2001, the "TRECVID" has been a benchmark for evaluating video retrieval systems [19] however most of the videos are in English not in Arabic thus it is not suitable to test our system. For these reasons, a self collected and prepared dataset is used to evaluate the proposed approach. Arabic closed captioning tool was developed to caption the collected videos. The dataset is collected from Arabic news videos and Arabic documentary films from different Arabic channels. It consists of about 456 scenes, which include 221 politics scenes, 30 economics scenes, 87 sports scenes, 4 religions scenes, 73 social scenes, 22 tourism scenes, 18 health scenes and one weather scene.

In order to evaluate the performance of our proposed approach, the following performance measures: precision and recall are employed [6, 8]:

$$\text{Precision (P)} = \frac{\text{Number of relevant scenes being retrieved}}{\text{Number of scenes being retrieved}}$$

$$\text{Recall (R)} = \frac{\text{Number of relevant scenes being retrieved}}{\text{Number of relevant scenes}}$$

Precision measures the proportion of results returned that are relevant, while Recall measures the proportion of relevant results returned. Moreover, it is important to evaluate precision and recall in conjunction, because it is easy to optimize either one separately. The F-Measure consists of a weighted combination of precision and recall and it is sometimes called harmonic mean. The general form of F-Measure is:

$$F_{\alpha} = \frac{(\alpha^2 + 1) * P * R}{\alpha^2 * P + R}$$

Where α is a weighting factor that determines the relative importance of precision and recall. For example, if α=0.5 means that recall is half as important as precision; and α=2 means that recall is twice as important as precision. However, in most experiments, there is no particular reason to favor precision or recall, so most researchers use a balanced weighting measure between precision and recall with α=1 as follows:

$$F_1 = \frac{2 * P * R}{P + R}$$

A number of queries were formulated for testing purposes. For each of these queries, a set of relevant scenes is predetermined manually. The experimental results based on some of these queries are presented in Table 5. The precision and the recall are computed for each query but the average precision and the average recall of the proposed system is computed for all the queries. For example, when the query "حادث تصادم قطارين" which means "an accident of two trains" is submitted to the proposed system, it retrieves four scenes. Three of them are relevant to the query while the other one is irrelevant so the precision of this query will be 75% (3/4). On the other side, the number of the relevant scenes supposed to be retrieved is three so the recall of this query will be 100% (3/3). Our experimental results are satisfactory; the system achieves an average precision equal to 88.94% and average recall equal to 93.93% which leads to a harmonic mean equal to 90.9%. The system can achieve better rates by enhancing the Arabic classification dictionary with more words which would also enhance the video scenes classification process.

Table 5. Experimental results of retrieving video scenes using the proposed approach

Query	# of retrieved scenes			Actual number of relevant scenes to be retrieved	Precision (%)	Recall (%)
	Relevant	Irrelevant	Total			
درجات الحرارة المتوقعة	1	0	1	1	100	100
انتخاب مجالس البلديات في ايران	١٣	٤	١٧	١٤	٧٦.٤٧	٩٢.٨٦
زيارة أعضاء مجلس الأمة المصانع المصرية	37	5	42	37	88.1	100
مهرجان دبي السينمائي	4	0	4	4	100	100
سرطان الثدي	7	1	8	٩	87.5	7٧.٧٨
مسابقة اختيار ملكات الجمال	10	3	13	11	76.92	90.9١
تصوير الهرم الأكبر	6	1	7	7	85.71	85.71
حادث تصادم قطارين	٣	١	٤	٣	٧٥	١٠٠
تمويل إنشاء المتحف المصري الجديد	٩	١	١٠	٩	٩٠	١٠٠
افتتاح نادي ضباط البوليس	٧	١	٨	٨	٨٧.٥	٨٧.٥

Unfortunately, there is no previous work proposed for retrieving video scenes using Arabic closed-caption so comparing our work with the others will not be fair-minded. This work is only the beginning; more research needs to be carried out. In general, when users input keywords to query video scenes, the proposed system can provide the exact answer and play corresponding scenes in different categories.

Figure 5 shows the user interface of our experiments. In the browsing and retrieval interface, users may select the categories of video document of interest from the scene database or input keywords in textbox to retrieve matched scenes in some or all categories. The filtered search results are those scenes with some degree of matching with the user input query while the general search results are those scenes that have the same semantic category of the query but have no matching with query keywords.

Figure 5. The user interface for the scene retrieval system

6. Conclusion

In this paper, a new approach for retrieving video scenes is presented. It combines the method of content-based video browsing based on category and the method of querying by inputting keywords. The proposed approach exploits the Arabic closed-caption text in retrieving video scenes to cross the semantic gab of multimedia content analysis and retrieval. The new approach classifies the input query and retrieves all the video scenes with the same category of the input query using a text-based video scenes classification technique. Then, a caption-based filtering technique is utilized to determine the relevant scenes. The evaluation of the developed system has indicated that the proposed approach is efficient in terms of both precision and recall. It achieves an average precision equal to 88.94%, an average recall equal to 93.93% and the F-Measure equal to 90.9%. These rates can be enhanced by adding more words in the Arabic classification dictionary used in the video scenes classification step.

7. References

1. Y. Li, W. Ming and C.-C. J. Kuo, "Semantic Video Content Abstraction Based on Multiple Cues," Proceedings of the IEEE International Conference on Multimedia and Expo (ICME'01), pp. 159-162, August 2001.
2. Tong Lin, Chong-Wah Ngo, Hong-Jiang Zhang, and Qing-Yun Shi, "Integrating Color and Spatial Features for Content-Based Video Retrieval," Proceedings of IEEE International Conference on Image Processing (ICIP01), vol.3, Thessaloniki, GREECE, pp. 592 - 595, October 2001.
3. C. V. Jawahar, Balakrishna Chennupati, Balamanohar Paluri, and Nataraj Jammalamadaka, "Video Retrieval Based on Textual Queries," Proceedings of the Thirteenth International Conference on Advanced Computing and Communications, Coimbatore, December 2005.

4. Hui Yang, Lekha Chaisorn, Yunlong Zhao, Shi-Yong Neo, and Tat-Seng Chua, "VideoQA: Question Answering on News Video," Proceedings of the Eleventh ACM International Conference on Multimedia, Berkeley, CA, USA, pp. 632 - 641, 2003.
5. Nevenka Dimitrova, Hong-Jiang Zhang, Behzad Shahraray, Ibrahim Sezan, Thomas Huang, and Avideh Zakhor, "Applications of Video-Content Analysis and Retrieval," IEEE Multimedia Magazine, Volume: 9, Issue: 3, pp. 42 – 55, Jul-Sep 2002.
6. Kieran Mc Donald, "Discrete Language Models for Video Retrieval," Ph.D. thesis, Centre for Digital Video Processing, Dublin City University, Dublin, Ireland, 2005.
7. Yun Zhai, Jingen Liu, and Mubarak Shah, "Automatic Query Expansion for News Video Retrieval," Proceedings of the 2006 IEEE International Conference on Multimedia and Expo, Toronto, ON, Canada, pp. 965-968, July 2006.
8. Yingying Zhu, and Dongru Zhou, "Video Browsing and Retrieval Based on Multimodal Integration," Proceedings of the IEEE/WIC International Conference on Web Intelligence (WI'03), Halifax, Canada, pp. 650-653, October 2003.
9. Hangzai Luo, Jianping Fan, Shin'ichi Satoh, and William Ribarsky, "Large Scale News Video Database Browsing and Retrieval via Information Visualization," Proceedings of the 2007 ACM Symposium on Applied Computing, Seoul, Korea, pp. 1086 – 1087, March 2007.
10. Daniel DeMenthon, and David Doermann, "Video Retrieval Using Spatio-Temporal Descriptors," Proceedings of the Eleventh ACM International Conference on Multimedia, Berkeley, California, USA, pp. 508 – 517, November 2003.
11. Wei Lai, Xian-Sheng Hua, and Wei-Ying Ma, "Towards Content-Based Relevance Ranking for Video Search," Proceedings of the 14th Annual ACM International Conference on Multimedia, Santa Barbara, CA, USA, pp. 627 – 630, October 2006.
12. N. Nakamura, S. Takano, and K. Niijima, "Video Scene Retrieval Based on the Layerization of Images and the Matching of Layer-Trees," Proceedings of the Image and Vision Computing 2005 (IVCNZ2005), New Zealand, pp. 449-454, 2005.
13. Mohammed Aljlayl, and Ophir Frieder, "On Arabic Search: Improving the Retrieval Effectiveness via a Light Stemming Approach," Proceedings of the 11^{th} International Conference on Information and Knowledge Management, McLean, Virginia, USA, pp. 340 - 347, 2002.
14. Hayder K. Al Ameed, Shaikha O. Al Ketbi, Amna A. Al Kaabi, Khadija S. Al Shebli, Naila F. Al Shamsi, Noura H. Al Nuaimi, and Shaikha S. Al Muhairi, " Arabic Light Stemmer: A New Enhanced Approach," Proceedings of the Second International Conference on Innovations in Information Technology (IIT'05), Dubai, UAE, September 2005.
15. Jinxi Xu, Alexander Fraser, and Ralph Weischedel, "Empirical Studies in Strategies for Arabic Retrieval," Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Tampere, Finland, pp. 269 - 274, August 2002.
16. Haidar Moukdad, and Andrew Large, "Information Retrieval from Full-Text Arabic Databases: Can Search Engines Designed for English Do the Job?," The International Journal of Libraries and Information Services (Libri), Vol. 51, No. 2, pp. 63 – 74, JUNE 2001.

17. L. Larkey, L. Ballesteros and M. E. Connell. "Improving Stemming for Arabic Information Retrieval: Light Stemming and Co-occurrence Analysis," In Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval SIGIR '02, Tampere, Finland, pp. 275–282, August 2002.
18. Youssef Kadri, and Jian-Yun Nie, "Effective Stemming for Arabic Information Retrieval," The Challenge of Arabic for NLP/MT, International Conference at the British Computer Society (BCS), London, pp. 68-74, October 2006.
19. Ahmed Abdelali, Jim Cowie, and Hamdy Soliman, "Arabic Information Retrieval Perspectives," Proceedings of the 11th Conference on Natural Language Processing, Journes d'Etude sur la Parole - Traitement Automatique des Langues Naturelles (JEP-TALN 2004 Arabic Language Processing), Fez, Morocco, April 2004.
20. Hamed Nassar, Ahmed Taha, Taymoor Nazmy and Khaled Nagaty, "Classification of Video Scenes Using Arabic Closed-Caption," Proceedings of the Third International Conference on Intelligent Computing and Information Systems, Cairo, Egypt, March 2007.
21. Gongde Guo, Hui Wang, David A. Bell, Yaxin Bi, Kieran Greer, "An kNN Model-Based Approach and Its Application in Text Categorization," Proceedings of the Fifth International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2004), CAU, Seoul, Korea, pp. 559-570, February 2004.