VIDEO CLASSIFICATION AND RETRIEVAL USING ARABIC CLOSED CAPTION

A.Anwar, Gouda Ismail Salama, and M.B. Abdelhalim

Arab Academy for Science, technology & Maritime Transport

ahmedelasfr@hotmail.com, dr_gouda80@yahoo.com, mbakr@ieee.org

Abstract

Vast volumes of digital video data are generated recently in our daily life. One of the most challenging problems is classifying and retrieving the desired information from huge collections of digital video. Consequently, the closed caption text has been utilized as an alternative to enhance the video retrieval and classification. Some systems are designed based on English closed caption however results have shown that Arabic is not lucky as English and other European languages in the research. This paper adopts an approach that enables video scenes classification and retrieving based on the Arabic closed-caption text that is present in the video. Experiments are performed over prepared dataset collected from Arabic news videos and Arabic documentary films across different Arabic channels. The results show that the proposed framework is efficient for retrieving Arabic videos and also for classifying Arabic video scenes into a set of eight predefined semantic categories including politics, economics, sports, religion, social, tourism, weather, and health.

***Index Terms*:** Arabic Closed Caption (ACC), Video Retrieval, Video classification.

1 INTRODUCTION

Arabic is the mother language of more than 300 million people [1]. Unlike Latin-based alphabets, the orientation of writing in Arabic is from right to left; the Arabic alphabet consists of 28 letters plus a special character called Hamza (ء). little research works have been conducted on Arabic corpuses mainly since Arabic language is highly rich and requires special treatments such as order verbs, morphological analysis, etc . Particularly, in Arabic morphology, words have affluent meanings and contain a great deal of grammatical and lexical information [2], Arabic has a more complex morphology than English, where the style of writing letters in a word varies depending on the position of the letter within the word. So, if the letters comes at the beginning, middle or at the end of the word, the letter shapes changes. 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. At a deeper level, most noun, adjective, and verb stems are derived from a few thousand roots by infixing, for example, creating words like (مكتب) pronounce as maktab which means (office), (كتاب) pronounce as kitaab which means (book),(كُتُب) pronounce as kutub which means (books), (كَتَبَ) pronounce as kataba which means(he wrote), and (نكتُب) pronounce as naktubu which means (we write), from the root(كتب) ktb [3].

Current rapid advances of the Internet and Multimedia technology enable huge amount of digital videos to be produced or downloaded, stored and organized, before they are retrieved, or shared on a daily basis. This phenomenon triggers a fast growth of video data, and thus, we need to have a content-based Video Classification and Retrieval System (VCRS), which can index the video according to its content and category, to allow users browse the video effectively. Many approaches of querying a video retrieval system have been used. One approach 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 approach is to use example video clip but it is quite complex for many applications to find example video clips for the concept of interest [4]. But Textual query is a promising approach for querying in video databases, since it offers a more natural interface [5].

The task of video classification is to find rules or knowledge from videos using extracted feature or mined results and then assign the video into predefined categories, one of the extracted features is

transcript which content ACC text, start time, end time and caption id. M.Al-diabat [6] applied different classification data mining algorithms (C4.5, PART, RIPPER, OneRule) on the problem of Arabic text classification. The result reveals that the most applicable algorithm to the Arabic data set is PART in which it derives higher results in all evaluation criteria than RIPPER, and PART. R. Al-Shalabi et al. [7] evaluate the use of the machine learning algorithm k Nearest Neighbor (kNN) to Arabic text, the results illustrated that kNN is applicable to Arabic text; kNN can work good with small number of training patterns provided that there are sufficient number of examples for each category.

Bacher [8] design the Monologue Dissector, which uses closed captions to identify jokes within a monologue. Certain words and phrases are hard-coded into his system in order to identify where jokes began and ended. This allows a user to search for jokes containing words of interest and then playing the video of that joke.

W. Zhu et al. [9] have used closed-captioned text in their research to categorize news videos. Authors have used 425 news stories from CNN and compared the categorizing performance with different classification methods. Those authors have defined eight categories: Politics, Daily Events, Sports, Weather, Entertainment, Health, Business and Science & Technology. The system is trained with a randomly selected ten to eighty percent of the stories and tested with the remaining ones. The authors have achieved a high precision and recall rates for the categories with salient language feature such as Sports, Weather, Health and Business. However, as expected, categories such as Daily Events fail to perform well since it is such a broad category and has a very limited number of unique language features. Their approach has yield poor results for the Entertainment and Science & Technology categories due to the insufficient training examples where no unique feature items could be found.

Darin Brezeale et al. [10] investigated closed captions and discrete cosine transform coefficients individually as features for classifying movies by genre and learning user preferences using a support vector machine as the classifier. They find that these features work very well for classification by genre but the results are less satisfactory when learning user preferences.

Nevenka Dimitrova et al.[11] develops two different video classification methods: domain knowledge based and HMM based. The former method applies classification based on face and text trajectory patterns within different categories of programs, which are extracted through observations and statistical approaches. The HMM method, represents the video clip as a series of frame labels, and uses the label strings as observation sequences to train HMM's and evaluate the probabilities of the given clip being one of the four categories of TV programs. The accuracy of the domain knowledge based method is 75% as opposed to 85% for the HMM method

In [12], the authors proposed a framework for text-based video-content classification for online-video-sharing sites. Different types of user-generated data (e.g., titles, descriptions, and comments) were used as proxies for online videos, and three types of text features (lexical, syntactic, and content-specific features) were extracted. They also adopted feature selection to improve accuracy and identify key features for online video classification. In addition, three feature-based classification techniques (C4.5, Naïve Bayes, and the SVM) were compared. Experiments conducted on extremist videos on YouTube demonstrated the good performance of their proposed framework.

H.Nassar et al. [13] presented a framework for classifying video scenes, focusing on its semantic features. The proposed framework utilizes the Arabic closed-caption text that contains the speech transcript of the video. Experiments were performed over self collected and prepared dataset. The results showed that the suggested framework is efficient for classifying video scenes into a set of predefined semantic categories. The classifier achieved an average recall rate of 92.38% and an average precision rate of 93.34% for 507 video Scenes.

This paper proposes an approach that enables video scenes classification which categorizes the video scene into one of a few predetermined scene classes based on the Arabic closed-caption text present in the video. Query about the scene and retrieving it can be made from at least one class after applying pre-processing, features and classification phases

The paper is organized in 4 sections entitled as follows. Section 2 Introduces the schematic process diagram for the video classification and retrieval model based on Arabic closed caption. Section 3 reports some experiments that are performed over self collected and prepared dataset collected from Arabic news videos and Arabic documentary films across different Arabic channels. Section 4 provides conclusions.

2 PROPOSED FRAMEWORK

Figure.1 presents the schematic process diagram for the video classification and retrieval model based on Arabic closed caption. Here, the video classification model is composed of two phases: learning phase and testing phase. The learning phase can be divided into four main processes. The first process is the acquisition of a video that represents the real world domain. The second process is the building of the SRT file that contains the caption ID, Scene start time, Scene end time and the caption text. The third process is the preprocessing (normalization, tokenization, stop word removal and stemmer) of the Arabic text. The fourth process is the feature extraction based on word dictionary and building the SRT dictionary that contains old sentence before applying preprocessing function, new sentence after preprocessing and features for each category. The testing phase can be divided into four main processes. The first process is the preprocessing of the input text query as the same like the learning phase. The second process is the feature extraction. The third process is the classification of the input query to a certain category using KNN classifier. The fourth process is the retrieval of the best 10 matching scenes using Jaro-Winkler distance metric.

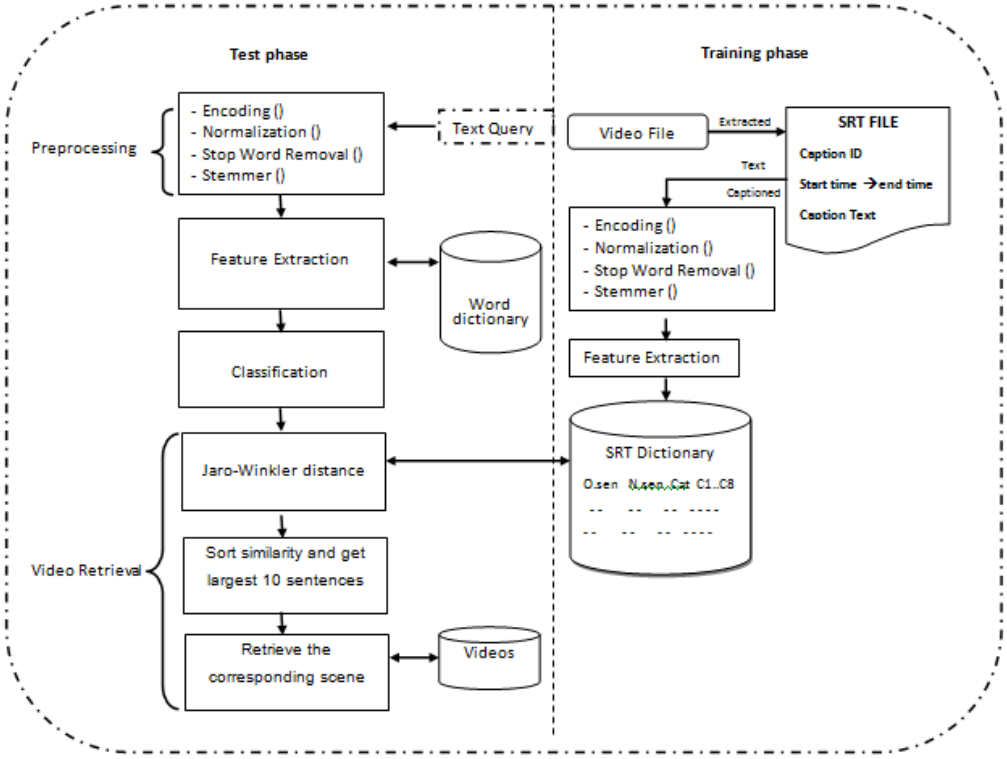

Figure1. The schematic process diagram for the video classification and retrieval model based on Arabic closed caption

2.1 ENCODING

Encoding of the texts in one standard format is used to represent texts without any deformation of character during the reading. All our corpus texts are represented with an UTF-8 encoding, the encoding supported by JAVA language. The collection of text and queries can be encoded differently, making them in comparable for example: documents are represented in Unicode (UTF-8) and requested in ISO-8859.6 or any other coding system. In order to standardize the documents with the queries, we must re-use converting tools between different encodings systems. Thus, everything would be converted into Unicode in our case

2.2 THE TEXT PRE-PROCESSING

This is an important process in our proposed system. It is used to obtain caption text symbols, numbers and words less than three letters are removed because these words in Arabic language will not affect the classification results. Then there are some necessary phases in caption text such as: Stop word removing [articles, determinates, auxiliaries, common words, etc], Normalization and Stemming

A. Normalization

Orthographic variations for Arabic language writers cause inaccuracies in natural language processing (NLP) application. So normalization is important process because it converts equivalent string to a unique format. Table 1 show some functions used in the normalization process.

Table 1. The functions of normalization

Function	Arabic latter Before normalization	Arabic latter After normalization
Strip_TATWEEL	العـــــــــربية	العربية
Normalize_HAMZA_above	أهؤلاء	اهءلاء
Normalize_TEh_Marbuta	دمية	دميه
Normalize_YEH	سلمي	سلمى
Normalize_ALEF_HAMZA_BELOW	إدارة	ادارة
Normalize_ALEF_MADDA	آدم	ادم
Normalize_FATHATAN	ولدً	ولد

For Example, after applying the normalization process for the Arabic sentence

"أستشتري دمي آلية لأبنائك قبل الإغلاق

It becomes

استشرى دمى اليه لابناءك قبل الاغلاق

B. Tokenization

Is the isolation of words to get text stream into segments. This isolation is based on white spacing and punctuations. This segmentation is necessary for NLP, Table 2 shows the tokenization process after applying on sentence.

Table 2: Example of Tokenizer

Phrase Before	لا تتم الأعمال العظيمة بالقوة ولكن بالصبر
Phrase After	لا | تتم | الأعمال | العظيمة | بالقوة | ولكن | بالصبر

C. Removing stop words

Stop words are common words that usually modify other words and carry no inherent meaning. There are techniques to eliminate stop words [14]. The first technique is based on using a list of stop words. The second technique is based on the elimination of words over a certain number of occurrences in the collection. In [14], the first technique is used to remove stop words by using a list of stop words. There is no general standard stop word list to use in an Information Retrieval (IR) experiments for Arabic language. in our study, the stop list constructed on the basis of [14] and [15] and completed by our care where we add missing common words approximately 13000 word. Fig. 2 presents a sample of Arabic stop words sets.

Common status	Arabic word	Arabic pronunciation	English meaning
Pronouns	هو	howwa	He
	هي	heyya	Shy
	هما	humaa	They
Adverbs	اليوم	El-youm	Today
	الامس	Al-ams	yesterday
	دائما	daiman	Always
	هنا	hona	Here
Prepositions	من	Men	From
	عن	an	About
	على	ala	On
Months	ديسمبر	December	December
	أكتوبر	October	October
Days	السبت	El-sabt	Saturday
	الاربعاء	El-arbaa	Wednesday

Figure 2. Example of Arabic stop words list

D. Stemming

Stemming is normalizing word variations by removing prefixes and suffixes to get the affix-free word. There are more kinds of stemmers but researchers proved that light-10 stemming based on morphological analysis is the best for Arabic language [16]. The affix removal technique uses a list of prefixes and suffixes to convert words to their base form (root or stem). In our approach, we use the light-10. Table 2 shows the stemming list of Light10. It does not produce the root of a given Arabic word; rather it removes the most frequent suffixes and prefixes without trying to deal with infixes.

Table 2. Light 10 stemming list

Remove Prefixes	ال – وال – كال – بال – فال – لل – و
Remove Suffixes	ها – ان – ات – ون – ين – يه – ية – ه – ة – ي

2.3 FEATURE VECTOR COMPOSITION

Consider the Arabic text-query "الذهبية لقطر والفضية للعراق في مسابقة كرة القدم" which means “Golden for Qatar and silver to Iraq in soccer competition”. We will take this text as input query in our proposed system, the first stage (preprocessing) will do over it 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 "ذهب لقطر فضي عراق مسابق كر القدم" " 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 3. Calculate feature vectors

Religion	Politics	Sports	Tourism	Weather	Social	Health	Economics
0	2	3	0	0	1	0	1
0/7= **0**	2/7= **0.29**	3/7= **0.43**	0/7= **0**	0/7= **0**	1/7= **0.14**	0/7= **0**	1/7= **0.14**

To calculate the first component, the number of words belonging to the religion category (0) is divided by the total number of words found in the Arabic classification dictionary (7) which gives (0). The rest of the feature vector's components are calculated using the same way and the vector becomes (0, 0.29, 0.43, 0, 0, 0.14, 0, 0.14). Then the classifier module uses this vector and classifies the text query.

2.4 VIDEO CLASSIFICATION

A number of statistical classification and machine learning techniques have been applied to text classification, including regression models, Bayesian classifier, decision tree, nearest neighbor classifiers, neural network and support vector machines. In pattern recognition the k-nearest neighbor algorithm (K-NN) is a method for classifying objects based on closest training example in the feature space. K-NN is a type of instance-based learning or lazy learning where the function is only approximated locally on all computation is deferred until classification. The object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its K-NNs (k is a positive number, typically small). If k=1, then the object is simply assigned to the class of its nearest neighbor.

In the previous section we are trying to build a feature vector for each data point (text to be classified) $\mathbf{p}$ and use this feature vector to compute distance between this data point and all feature vectors of stored (training) data points $\mathbf{q}$. Then getting the majority votes to determine the associated category for the test data point. The distance is usually measured in terms of Euclidean distance, represented in the following equation.

$$d(p,q) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \qquad (1)$$

The similarity function that described in the following section will retrieve the corresponding Scene.

2.5 VIDEO RETRIEVAL

In this phase, we use a Jaro-Winkler distance similarity between two strings. It is a variant measure proposed in [17], from the distance of Jaro [18] which is mainly used in the detection of duplicates. The resultant distance measure between two strings is normalized to have a measurement between 0 and 1, where zero representing the absence of similarity. The Jaro distance score d_j of two strings S1 and S2 can be calculated as follows [19]:

$$d_j = \frac{1}{3}\left(\frac{m}{S1} + \frac{m}{S2} + \frac{m-t}{m}\right) \qquad (2)$$

Where $\mathbf{m}$: the number of matching character

$\boldsymbol{t}$: Half number of transpositions. Where transposition is number of match character but different sequence order

It's made up of the average of three sub-calculations.

1. The ratio of matching characters to the length of the first string.
2. The ratio of matching characters to the length of the second string.
3. The ratio of non-transpositions to the number matching of characters.

Example: consider two Strings المرجع and المعرج Table 4 shows number of matched character flagged by ones :

Table 4. Matching of two string "المرجع" and "المعرج"

	ا	ل	م	ر	ج	ع
ا	**1**	0	0	0	0	0
ل	0	**1**	0	0	0	0
م	0	0	**1**	0	0	0
ع	0	0	0	0	**1**	0
ر	0	0	0	**1**	0	0
ج	0	0	0	0	0	**1**

The Jaro-distance is: $d_j = \frac{1}{3}\left(\frac{6}{6}+\frac{6}{6}+\frac{6-1}{6}\right)$=0.944

had an additional insight: these kinds of errors are much more likely to occur later in the string. By simply wrapping Jaro's algorithm to forgive some penalty according to the similarity of the first few characters, Winkler was able to get significantly better results with little additional overhead.

$$d_w = d_j + \left(lp(1-d_j)\right) \quad (3)$$

Here we see that the Jaro-Winkler distance (d_w) is equal to the result of the Jaro distance (d_j) plus one minus that same value times some weighted metric (lp) where $\mathbf{l}$ is the length of common prefix at the start of the string up to a maximum of 4 characters Meanwhile $\mathbf{p}$ is a weight which can be up to 1/4, or one over the maximum possible value of $\mathbf{l}$. After much experimentation, Winkler recommends using a $\mathbf{p}$ value of 0.1, which is also what we use.

$$d_w = 0.944 + \left(3 * 0.1(1-0.944)\right)=0.96$$

Queries are usually generated from the user perspective and the more the information available the more is the documents retrieved. In general, any query consists of a word or a collection of words. Search engines retrieve the documents related to each word in comparison with the content and show the results on the interface.

3 EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

3.1 EXPERIMENTAL SETUP

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 [20]. 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 [20]. However most of the videos are in English not in Arabic thus it is not suitable to test the system. For these reasons, a self collected and prepared dataset is used to evaluate the proposed approach. Our proposed system is developed by JAVA desktop application and Java Media Framework (JMF) to caption, search and classification across the collected videos. We have collected dataset from Arabic news videos and Arabic documentary films across different Arabic channels. The database consists of approximately 3 hours and 30-minute long MPEG formatted videos. Also, about 2500 Arabic words are entered in the dictionary and classified into eight predefined semantic categories including politics, economics, sports, religion, social, tourism, weather, and health. Several experiments are conducted in order to measure the performance of the proposed classifier.

3.2 EVALUATION MEASURES:

Recall (R), Precision (P) and F-Measure (F) are standard information retrieval measures, and they form the basis of most measures of effectiveness in video retrieval. Recall and Precision are set-based evaluation measures, but that do not take into account the ranking of the search results. We can compute the precision as how many classifications Scenes did you get right based upon the number of retrieved Scenes as:

$$Precision\ (P) = \frac{Number\ \ of\ relevant\ \ scenes\ \ being\ \ retrieved}{Number\ \ of\ scenes\ \ being\ \ retrieved} \quad (4)$$

Also, we can compute recall as how many classifications Scenes did you get right compared to the total number of correct classifications Scenes as:

$$Recall\ (R) = \frac{Number\ \ of\ relevant\ \ scenes\ \ being\ \ retrieved}{Number\ \ of\ relevant\ \ scenes} \quad (5)$$

Precision is easier to assess correctly than recall and the effort involved is related to the result set length. Recall is more difficult to calculate than precision because it requires knowledge of all relevant documents within the test collection. For large collections, it is prohibitive expensive to evaluate the relevance of each Scene.

The F-Measure consists of a weighted combination of precision and recall and it is sometimes called harmonic mean [21]. The general form of F-Measure is given by:

$$F = \frac{2 * P * R}{P + R} \quad (6)$$

3.3 EXPERIMENTAL RESULTS

The user interface of the recent experiments is shown in Figure (3). Across browsing and retrieval interface, users input keywords in query textbox to retrieve the matched scenes in either some or all categories. Scenes with some degree of matching with user's input query are the filtered search results while the general search results are those scenes that have the same semantic category of the query, but have no matching with query keywords. In the recent experiment the text query " الذهبية لقطر والفضية للعراق" is input which means here "The golden for Qatar and the silver for Iraq". This text query is used in the feature vector section where the result shows that it is sports category through which the filter results show zero scene matches with degree 100%, 664 matches with 50% to less than 100% and 348 matches with 0% to less than 50%. For the best retrieving we put threshold value 65% so there are 3 scenes only in the top 10 list.

Figure 3: The user interface of the video retrieving system for the text query "الذهبية لقطر والفضية للعراق", "The golden for Qatari and silver for Iraq"

Figure (3) shows another experiment which contains new input query " إنفجار منجم للفحم" which means "A coal mine explosion". The result shows that this query is classified (Economic, Politics and Social) category which filter results show zero scene matches with degree 100%, 755 matches with 50% to less than 100% and 330 matches with 0% to less than 50%. For the best retrieval we put threshold value 65% so there are 5 Scene only in the top 10 list.

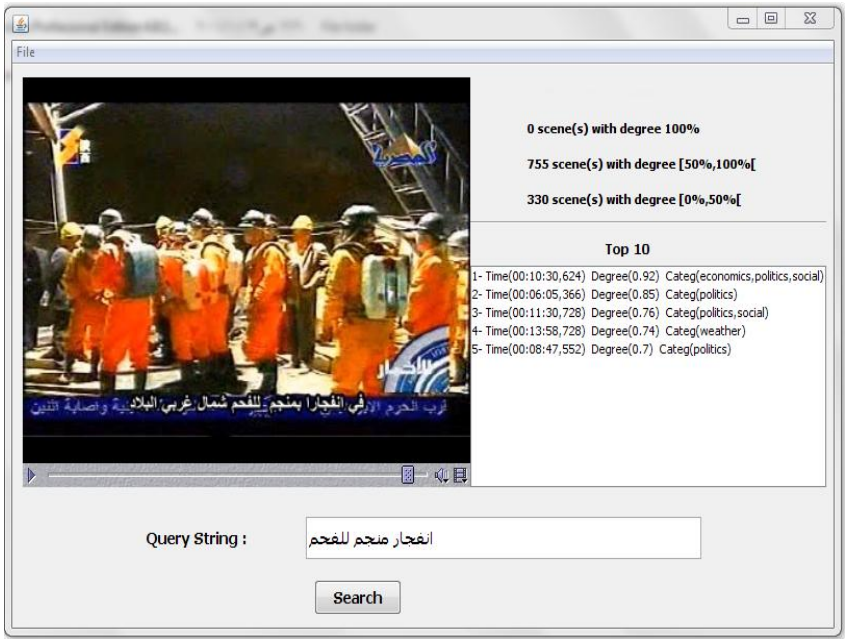

Figure 4. the user interface of the video retrieving system for the text query "إنفجار منجم للفحم", "A coal mine explosion"

Confusion matrix results are represented in table (5) where the classifying 624 Scenes are classified using the proposed approach. To start with, in the politics category, 128 political scenes are entered to the proposed approach. The recent application classifies 124 scenes correctly while it classifies 4 scenes incorrectly to other categories so the recall will be 124/128= 96.87%. In the other side, it classifies 4 scenes are classified as political scenes incorrectly so the precision will be 124/ 128 = 96.87%. Following the same steps, 40 economics scenes are passed to the classifier. 38 scenes of them are classified correctly so the recall of the social category will be 38/40=95%. Moreover, the classifier misclassifies 2 scenes as politics scenes so in this stage the precision will be 38/ (38+9) = 80.85%. From the results shown in table 4, the accuracy of all categories are good while economics and social categories got the lowest accuracy. we have applied the same methodology across the rest of the eight predefined categories. The proposed approach achieves an average recall rate equal to 94.03%, an average precision 93.5% and the F-measure equal to 93.7%.

Table 5 Confusion matrix of classifying video scenes using the proposed approach

	Politics	Economics	Sports	Religion	Social	Tourism	Weather	Health	Total	Precision %
# of scene	128	40	201	20	35	45	35	120	624	
Politics	124	2	0	0	1	0	0	1	128	0.96875
Economics	3	38	3	0	2	1	0	0	47	0.808511
Sports	0	0	197	0	0	2	0	6	205	0.960976
Religion	0	0	0	19	0	0	0	0	19	1
Social	1	0	1	1	32	0	0	7	42	0.761905
Tourism	0	0	0	0	0	42	0	0	42	1
Weather	0	0	0	0	0	0	33	0	33	1
Health	0	0	0	0	0	0	2	106	108	0.981481
Recall	0.96875	0.95	0.9801	0.95	0.91429	0.93333	0.94286	0.88333		

3.4 COMPARISON OF THE PROPOSED RETRIEVAL APPROACH AND SIMILAR APPROACHES

In this section, we build a comparison across different stages used in the recent proposal and the one presented by H. Nassar et al. [13], which in some way similar to our recent proposal in some stages, but different in the other ways that leads to the desired results depending on recommendations in future works listed in [13]. These differences start at the user's behavior. First, when he uses the search tool and classification he does not care much for entering the enquiry text without format and demarcation. In a trial to solve such problem, the researcher depends on Unicode (UTF-8) system in order to standardize the document with queries. Then, the preprocessing stage starts which passes through many different stages than those of H.Nassar et al. [13] across the tokenization process through which words can be separated from each others to support classification accuracy. After that, the stop word removing process is applied in which H.Nassar et al. [13] has made a list of 312 words, which the researcher - among other researchers - have extended it up to about 13000 words. This enlarged list helps to increase the classification accuracy as the new list is able to delete all words that do not affect the Scene classification

Moreover, H.Nassar et al. [13] uses the most important process of word stemming to extract the word root hence the Arabic language accepts writing words with many different syntax inputs. Consequently, H.Nassar et al. [13] uses a list that does not depend on a special kind of stemming that actually has different types and editions across which the researcher has chosen the best (Light-10 Stemming), which has proven its accuracy in text classifying. Across the feature vector stage, H.Nassar et al. [13] has extracted features based on a predefined word dictionary that includes more than 1150 words. This dictionary is expanded to reach about 2500 words in most categories. In addition, the data set video is enlarged to about 3 hours which also improves classification performance. Furthermore, H.Nassar et al. [13] uses the Rocchio classifier which represents each category in a prototype vector. This vector is usually calculated by averaging the training samples for each category. The unknown sample is assigned a category label that has the minimum distance to its prototype vector. The distance is usually measured in terms of Euclidean distance.

In the current proposal, the researcher uses K-NN classifier through which the object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its K-NNs. accordingly, the closest k-category is determined through the distance that is usually measured in terms of Euclidean distance. Hence, more than one class is determined to search in it, which is different from H.Nassar et al. [13] , who searches in one class only. After determining class it is required to retrieve the Scene that wants to search about. In this stage, it is time for Jaro- Winkler distance that leads to similarity between two strings, which retrieves all Scenes in which the inquiry word has been mentioned. These results are arranged, then the highest 10 Scenes that proves more than 65% similarity in more than one class are selected.

Across many experiments, the researcher has achieved an average recall rate equal to 94.03%, an average precision 93.5% and the F-measure equal to 93.7% for 624 scenes represent 8 categories. On the other hand, H.Nassar et al. [13] has achieved average recall rate equal to 92.38%, average precision 93.34% and F-measure of 92.86% for 507 scenes represent the same categories.

Table 5 Comparison of our proposed system and other similar systems

Method	Average precision	Average Recall	F-Measure
H.Nassar et al	93.34%	92.38%	92.86%
Proposed Approach	93.50%	94.03%	93.7%.

4 CONCLUSIONS

This paper addressed the general structure of video scenes classification and retrieving based on the Arabic closed-caption text that is present in the video. The new approach classifies the input query and retrieves all the video scenes with the same category of the input query using KNN classifier. Experiments were performed over self collected and prepared dataset about 2500 Arabic words are entered in the dictionary and classified into eight predefined semantic categories including politics, economics, sports, religion, social, tourism, weather, and health. As a result of the experiments, we were able to classify and retrieve Arabic videos into the set of eight predefined semantic categories with an average recall rate equal to 94.03%, an average precision 93.5% and the F-measure equal to 93.7%.

REFERENCES

[1] Mohammed.Aljlayl and Ophir.Frieder, "On Arabic Search: Improving the Retrieval Effectiveness via a Light Stemming Approach," proceedings of the ACM International Conference on Information and Knowledge Management (CIKM 2002), McLean, Virginia, USA, pp. 340-347, November 2002

[2] Fadi Thabtah, Omar Gharaibeh and Rashid Al-Zubaidy "Arabic text mining using rule based classification," Journal of Information and Knowledge Management,vol.11, No.1 pp.1-10 December 2011.

[3] Wightwick,J. and Gaafar, M.Arabic verbs and essentials of grammar. Chicago: Passport Books, 1998.

[4] Hangzai Luo, Jianping Fan, Shin'ichi Satoh, and William Ribarsky, "Large Scale News Video Database Browsing and Retrieval via Information Visualization," Proceedings of the ACM Symposium on Applied Computing, Seoul, Korea, pp. 1086 – 1087, March 2007.

[5] 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.

[6] M. Al-diabat, "Arabic Text Categorization Using Classification Rule Mining," Applied Mathematical Sciences, Vol. 6, pp. 4033 - 4046, 2012.

[7] R. Al-Shalabi, G. Kanaan, M. Gharaibeh, "Arabic Text Categorization Using kNN Algorithm", Proceedings of The 4th International Multiconference on computer Science and Information Technology (CSIT), vol.4, Amman 2006

[8] D. R. Bacher, "Content-based indexing of captioned video," SB Thesis, Massachusetts Institute of Technology, May 1994.

[9] Weiyu Zhu, Candemir Toklu, Shih-Ping Liou. s.l., "Automatic News Video Segmentation and Categorization Based on Closed-Captioned Text," IEEE International Conference on Multimedia and Expo, 2001

[10] Darin Brezeale, and Diane J. Cook, "Using Closed Captions and Visual Features to Classify Movies by Genre," In proceedings of the 7th International Workshop on Multimedia Data Mining (MDM/KDD2006), Philadelphia, Pennsylvania, USA, pp. 153-157, August 2006.

[11] Nevenka Dimitrova, Lalitha Agnihotri and Gang Wei, "VIDEO CLASSIFICATION BASED ON HMM USING TEXT AND FACES," European Signal Processing Conference, 2000

[12] Chunneng Huang,Tianjun Fu, and Hsinchun Chen, "Text-Based Video Content Classification for Online Video-Sharing Sites," Journal of The American Society For Information Science And Technology, Vol 5, No 61, PP 891–906, 2010

[13] H.Nassar, A. taha, T. Nazmy, KH.Nagaty. "Classification of Video Scenes Using Arabic Closed-caption", Third International Conference on Intelligent Computing and Information Systems, Cairo Egypt, PP.186-192, March 2007.

[14] T. Dilekh, A.Behloul. "Implementation of a New Hybrid Method for Stemming of Arabic Text," International Journal of Computer Applications,Vol.46, No.8,pp 14-19, May 2012

[15] Kadri, Y. & Nie, J."Effective Stemming for Arabic Information Retrieval" in proceedings of the Challenge of Arabic for NLP/ MT Conference, Londres, Royaume-Uni, 2006

[16] L. S. Larkey, L. Ballesteros and M. E. Connell, "Light Stemming for Arabic Information Retrieval," Technical Report, Intelligent Information Retrieval Center, Computer Science Dept., Massachusetts University, 2005 .

[17] Porter, E.H., and Winkler W.E., "Approximate String Comparison and it's Effect in an advanced Record Linkage System," National Academy Press: Washington, D.C, PP 190-199, 1999

[18] Jaro, M.A., "Advances in Record-linkage Methodology as Applied to Matching the 1985 Census of Tampa," Journal of the American Statistical Association, vol.89, PP 414-420, 1989

[19] A.ENAANAI, A.S. DOUKKALI, "An hybrid approach to calculate relevance in the Arabic meta-search engines", The International Journal of Science and Advanced Technology,Vol.2, No.3,pp 127-134 March 2012

[20] Ahmed Abdelali, J. C, and H. Soliman, "Arabic Information Retrieval Perspectives," in the 11th Conference on Natural Language Processing, Fez,Morocco, 2004

[21] Martin Mehlitz, Christian Bauckhage, Jerome Kunegis, and Sahin Albayrak, "A New Evaluation Measure for Information Retrieval Systems," In proceedings of the ccIEEE International Conference on Systems, Man and Cybernetics (SMC2007), Montreal, Quebec, Canada, pp. 1200-1204, October 2007.