Third International Conference on Intelligent Computing and Information Systems

March15-18, 2007. Cairo, Egypt.

Classification of Video Scenes Using Arabic Closed-Caption

Hamed Nassar, Ahmed Taha
Computer Science Department
Faculty of Computers & Informatics
Suez Canal University
ah_twins@hotmail.com

Taymoor Nazmy and Khaled Nagaty
Computer Science Department
Faculty of Computer & Information Sciences
Ain Shams University

ABSTRACT

Classifying video scenes into semantic categories is a research problem of great interest. This process is usually handled using supervised classifiers based on features describing the global distribution of color, texture or edge information of the video image. Obviously, the speech of the video scene contains more detailed semantic information and the video transcript is now available through Closed-caption text. Furthermore, the analysis of the speech transcript of the video scenes is easier than video image analysis. Little work has been done using English closed-caption text but Arabic closed-caption text is not even as fortunate. This paper proposes a framework for the classification of video scenes using Arabic Closed-Caption.

Keywords

Video Scenes Classification, Video Scenes Categorization, Arabic Closed-Caption, Video Content Analysis.

1. INTRODUCTION

The rapid growth of the access to the Internet and developments in image/video processing technology, along with the expansion in networking is facilitating production and consumption of huge amounts of digital video. This continuous increase in the media documents and the wide application of digital video data has led to a significant need for efficient browsing, retrieval and summarization systems. Video content analysis is a basic step towards building these systems. There are many domains that require video content analysis, such as large distributed digital libraries, broadcasting or production archives and video databases.

Video analysis can be performed in the feature level (color, shape, motion…etc) and the semantic level. One of the important ways for performing the semantic video analysis is video classification. It is the first step towards multimedia content understanding [16]. In order to classify the video, it is necessary to segment the video into semantically rich and independent units. The video is represented in the form of cascaded shots, which are segmented based on low level features [19]. A shot is a sequence of images that preserve similar background settings and it is the basic element of the video [18]. But the low-level shot structures do not reflect well the semantic content of the video data [19]. Moreover, the segmentation of the video in this way generates a huge number of shots which may represent a burden when dealing with them. Recently, the researchers consider video segmentation in terms of scenes. A scene consists of a set of continuous and related shots. In this way, the video sequence is first segmented into shots then semantically related and temporally adjoining shots are grouped into scenes [19]. A lot of techniques were developed for the video scene detection [3, 15, 20, 21].

This research is concerned with the video scene classification which is the categorization of a video sequence into one of a few predetermined scene classes [5]. Classifying video scenes into semantic categories is a research problem of great interest. For example, in video analysis there is a need to classify video scenes as containing certain objects (building, car, etc), describing specific scenes (sunset, beach, etc), or showing semantic information (politics, economics, sports, etc). This problem is conventionally solved using supervised classifiers based on features describing the global distribution of color, texture or edge information of the image [6]. Video images provide rich visual information. However, both intuition and experience show that the speech content in digital videos plays a more significant role in video content understanding [5]. Since the video speech is now available through closed-caption text, the video scene classification could be implemented as a text categorization problem using the closed-caption text.

In this paper, a framework for the classification of video scenes using Arabic Closed-Caption is presented. Another contribution is that an Arabic classification dictionary is constructed in which each Arabic word is stored in the dictionary with its classification after applying a light stemmer. This Arabic classification dictionary will be useful for further research. Also, an Arabic closed captioning tool is built to prepare the dataset and finally, Arabic closed captioned videos are collected and captioned to form a dataset for testing the proposed framework.

This paper is organized as follows: related work is discussed in Section 2. Section 3 presents the Arabic text classification techniques. Some Arabic language properties are shown in section 4. Section 5 describes the proposed classification framework. Section 6 presents the experimental results and Section 7 concludes the work.

2. RELATED WORK

Video scene classification is an active research domain that aims to analyze the semantic information of the video content. Three different strategies for video scene classification can be found in the literature. The first strategy is based on the low level features of the video image like color, texture and other properties. The researchers found that the low level features of the video image are not enough in the classification process. This is because it classifies only a small number of scene categories (indoor versus

outdoor, cityscape versus landscape etc...) [2]. Also, the low level features don't contain much semantic information about the scene.

The second strategy doesn't depend only on the video image properties but it uses other clues to enhance the classification like audio, motion and caption that appear on the video image. Zhai et al proposed a framework for video scene classification [18]. The framework utilizes the structural information of the scenes together with the low level features (including motion and audio energy) and a mid level feature (body). Rho et al proposed a scheme for determining video scenes by analyzing both audio and video data [14]. Also, they reported some of the results in the automatic segmentation and classification of audiovisual data for video indexing and retrieval. Four different methods for integrated audio and visual information based on Hidden Markov Models are proposed [5]; they classify TV programs into news reports, weather forecasts, commercials, basketball games, and football games.

In the third strategy, few studies in the video classification have focused on higher-level text features such as transcripts, which are either generated from closed-captions or automatic speech recognition systems [16]. Zhu et al present a statistical approach, called the weighted voting method, for automatic news video story categorization based on the closed captioned text. In this method, News video is segmented into stories using the demarcations in the closed captioned text then a set of keywords is extracted to form a feature vector. The categorization is achieved by computing the likelihood score for each category and the knowledge base is updated incrementally in linear time [17].

Yang et al take a fully text categorization approach to the problem of scene classification based on vector-quantized keypoint features or visual-word features. That is, they treat visual words in images as words in documents, and apply techniques widely used in text categorization to the scene classification problems [6].

3. ARABIC TEXT CLASSIFICATION TECHNIQUES

Text categorization is the assignment of a class from a set of predefined classes or categories to an unknown text or document. Over the years, text categorization was thoroughly studied by many researchers. There are many different methods for text classification including; Bayesian classification, statistical-based algorithms, distance-based algorithms, k-nearest neighbors, decision tree-based methods [8].

Text categorization process can be divided into three stages: preprocessing, document indexing and classification. The preprocessing stage includes the conversion of the document into plain text, stop words removal and stemming process. The second stage includes the construction of the super vector, feature selection and feature weighting. Finally, the third stage includes the construction of the classifier, the classification process and its evaluation [12]. Many algorithms have been developed and tested for each step mentioned above in the text categorization process.

The majority of the research in text categorization is focused on English documents. Little work has been done on text categorization for Arabic language. Different algorithms for Arabic text categorization have been proposed [4, 11, 13]. Also, a model for Arabic text categorization was presented [12]. In this model, the researchers made a comparative study among the different algorithms, measures and techniques in each step of the text categorization process to determine the most suitable algorithms for Arabic.

4. ARABIC LANGUAGE PROPERITIES

Arabic is the official language of 22 countries stretching from eastern Asia to northwest Africa. It belongs to the Semitic family of languages which also includes Hebrew and Aramaic. There are 28 characters in the Arabic alphabet. The language is written in horizontal lines from right to left. It has a more complex morphology than English, where morphological change in Arabic can result from the addition of prefixes and infixes as well as suffixes.

Arabic is a highly productive language. Definite articles, conjunctions, particles and other prefixes can be attached to the beginning of a word. Also large numbers of suffixes can be attached to the end [9]. The major part of words has a tri-letter root. The remaining part has either a quad-letter root, penta-letter root or hexa-letter root. The Arabic language needs much preprocessing in order to be convenient for manipulation. Also, it has a high degree of ambiguity; for example, similarity between affixation letters and stem boundary letters is one of the reasons causing this ambiguity.

5. PROPOSED FRAMEWORK

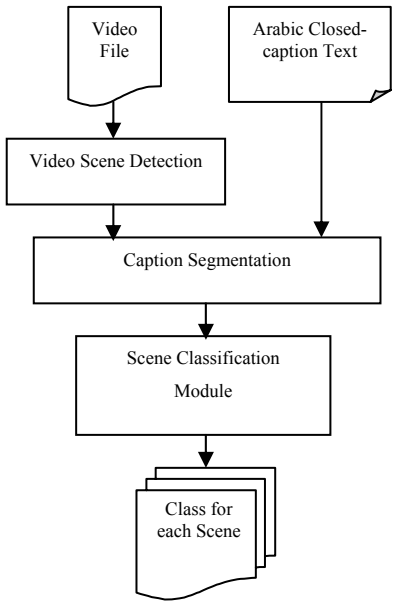

Figure 1. The proposed framework for the video scene classification

Video images are rich with their visual information like color and texture but the video speech contributes more in the semantic analysis of the video scenes. This fact along with the availability of closed caption text makes the video scenes classification problem viewable as a text categorization problem. Figure 1 shows the flowchart of the proposed framework for video scenes

classification based on the Arabic closed-caption text of these video scenes.

First, the video is segmented into a set of scenes using a video scene detection algorithm [3, 15, 20, 21]. Then the closed caption text is segmented to extract the caption for each detected video scene. The caption text consists of several lines of timecode and words. The timecode is used to synchronize the audio in the video with the words in the caption text. Each video scene may contain one or more caption and the caption may appear in one or more consecutive scenes. For example, in Figure 2 scene A contains captions 1, 2 and 3 while scene B contains captions 4 and 5. Note that caption 5 starts at scene B and ends at scene C so it must be included in both scenes during the segmentation process. Also, caption 8 is extended over two consecutive scenes D and E.

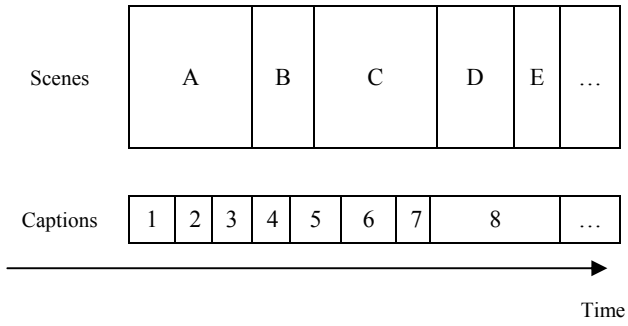

Figure 2. The caption text segmentation

To determine if a given caption will be assigned to a specific scene, the segmentation algorithm in the proposed framework checks if the start time or the end time of the caption lies within the scene time boundaries. If so, the caption will be included (Ex. scenes A, B and C in Figure 2) otherwise, the algorithm checks if the start time of the given caption is less than or equal to the start time of the scene and the end time of the caption is greater than or equal to the end time of the scene (Ex. scenes D and E in Figure 2). If this condition is satisfied, the caption will be assigned to the scene. This segmentation process uses the demarcations found in the closed-caption file and the time boundaries of each detected scene.

After the segmentation process, the closed caption text for each video scene is then passed to the classification module to determine its semantic category like politics, economics, religions…etc. Figure 3 shows the scene classification module. It consists of three stages: the preprocessing stage (including keyword extraction, normalization, stop words removal and stemming), the indexing stage (including feature vector construction) and the classification stage (including the classifier construction and the classification process).

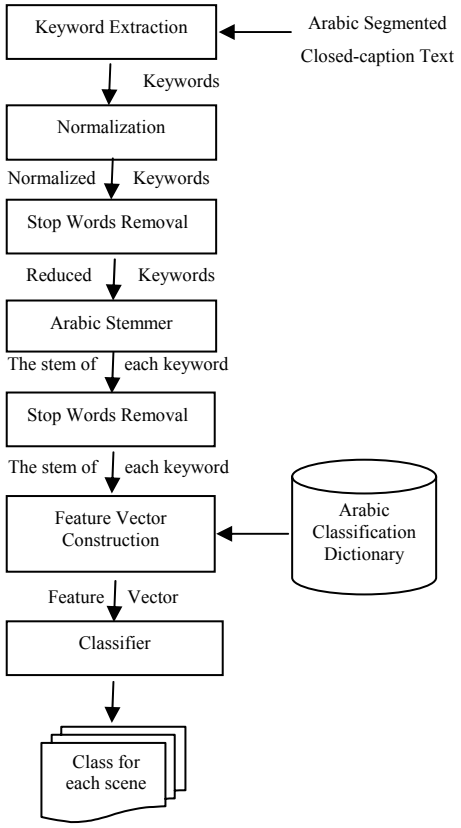

Figure 3. The classification module of the proposed framework

5.1 Keyword Extraction

In this step, closed-caption file is processed to extract caption text. Keywords are extracted from the caption of each scene then numbers and words less than three letters are removed because these words in Arabic language will not affect the classification results [12]. Also the special demarcations of the closed-caption text are removed. The closed-caption text contains several markers that are used to refer to sound effects, speaker identification … etc. For example, the brackets "[" and "]" are usually used to indicate the sound effects in the closed-caption text such as [صوت طلقات رصاص] which means [gun firing]. Also, the symbol ">>" is used to indicate the change of speaker.

5.2 Normalization

Normalization is the unification process of different forms of the same letter. Some minor replacements of certain Arabic characters are performed on the Arabic words to facilitate the categorization process as shown in table 1. More specifically, we map the different forms of Alef to bare Alef (i.e. Alef without Hamza or Madda), Teh Marbuta to Heh Marbuta, Alef Maksura to Yeh and finally the different forms of Yeh with Hamza to Hamza over Yeh. For example, the Arabic word (مكتبة) which means *library* to (مكتبه) and the word (أقلام) which means *pens* to (اقلام).

Table 1. The normalization of the Arabic letters

The Arabic letter before normalization	The Arabic letter after normalization
آ , إ , أ	ا
ـة	ـه
ى	ي
يء , ىء	ئ

5.3 Stop Word Removal

Stop words are the insignificant words that have no effect on the classification process like prepositions and particles. Removing these words is essential to reduce the time needed for classification. We construct a list of stop words containing 312 words (including pronouns, prepositions, adverbs, articles, question words…etc). Table 2 shows some examples of stop words.

Table 2. Examples of Arabic stop words

	Arabic Word		English Meaning
Prepositions	في	Fey	in
	على	Alaa	on
	إلى	Ela	to
Pronouns	أنا	Ana	I
	نحن	Nahno	we
Adverbs	تحت	Tahit	below
	فوق	Fawk	above
	الآن	Alaaan	now
	منذ	Monzo	since
Question	لماذا	Lemaza	what
	متى	Mata	when
Articles	إذا	Eza	if
	ثم	Soma	then
	عدا	Adaa	except

5.4 Stemming

Stemming is the process of removing the affixes from the word and extracting the word root. Many algorithms are proposed for Arabic stemming [1, 7, 9, 10]. The Arabic stemming algorithms can be categorized in three different categories: rule based stemming, light stemming and statistical stemming.

Researchers proved that a hybrid statistical and light stemmer is the most suitable stemming algorithm for Arabic language [12]. Also, Larkey et al compared light stemming with several different stemming approaches based on morphological analysis [9]. Their experimental results proved that the light stemming is the best effective stemming method for languages with more complex morphology like Arabic. In the proposed framework, we use the same strategy. Light stemmer 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, or recognize patterns and find roots. For example, when applying the light stemming to the Arabic word (بمصنعهم) pronounced as *Bemasnaahom* which means *in their factory*, the produced stem will be (مصنع) pronounced as *Masnaa* which means *factory*. This Arabic stem is produced after removing the prefix (بـ) and the suffix (هم). Table 3 shows the list of the prefixes and suffixes that the Arabic stemmer removes them.

Table 3. The list of the prefixes and suffixes removed by the Arabic stemmer

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

5.5 Feature Vector Construction

Each video scene is represented by a feature vector. The number of components in each feature vector is equal to the number of the predefined semantic categories used for the classification process. Each component in the feature vector is the percentage of the words belonging to the corresponding category that appear in the closed-caption text of this scene. In other words, it represents the membership probability for each category. This is analogous to the bag-of-word representation of textual documents that is used in text categorization techniques. An Arabic classification dictionary is used to compute this percentage. 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.

(a)

عقدت الحكومة الفلسطينية جلسة طارئة لدراسة الوضع الأمني المتدهور في الأراضي الفلسطينية بعدما بلغ الأحتقان حدا غير مسبوق بدأ يهدد السلم الأهلي ومن أجل تهدئة الوضع المتوتر

(b)

Politics	Economics	Sports	Religion	Social	Tourism	Weather	Health
9	1	1	0	3	0	0	2

(c)

$$(\frac{9}{11},\frac{1}{11},\frac{1}{11},\frac{0}{11},\frac{3}{11},\frac{0}{11},\frac{0}{11},\frac{2}{11})$$

Figure 4. The feature vector construction for a given video scene

Figure 4.a shows an Arabic closed-caption text for a given video scene. After normalization, stop words removal and stemming stages, classes are identified for each produced stem using the Arabic classification dictionary. Eleven words were found in the classification dictionary. Figure 4.b shows the number of words belonging to each category that appear in the closed-caption text of the given scene. 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. Figure 4.c shows the computations of each component in the feature vector. For example, to compute the first component, the number of words classified as politics (9) is divided by the total number of words found in the dictionary (11) which gives 0.82. Using the same way to compute the rest of the feature vector's components, the

resultant feature vector for this video scene is (0.82, 0.09, 0.09, 0, 0.27, 0, 0, 0.18).

5.6 Classifier

Many classifiers have been used in Arabic text categorization [4, 11, 12, 13]. The proposed framework has used the Rocchio classifier (sometimes known as Relevance feedback classifier). In this classifier, each category is represented by a prototype vector. This vector is constructed during the learning process of the classifier and it is usually calculated by averaging the training samples for each category. 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. The distance is usually measured in terms of Euclidean distance, where the Euclidean distance between two vectors X= (x_1, x_2 ..., x_n) and Y= (y_1, y_2 ..., y_n) is defined by the following equation:

$$d(X,Y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

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.

In the previous example shown in Figure 4, the distances between the feature vector of the given scene and the prototype vector of each category were 0.33 - 1.05 - 1.06 - 1.11 - 0.73 - 1.05 - 1.11 - 0.92 respectively. Since the minimum distance is 0.33 then the scene will be classified as politics.

6. EXPERIMENTAL RESULTS

Up till now, there is no Arabic dataset available for testing. The experiments are performed over a self collected and prepared dataset. We initially built an Arabic closed captioning tool to caption the collected videos. Captioning refers to the addition of text to the video picture, where the spoken words are seen as text. Captions not only display words to indicate spoken dialogue or narration, but also include sound effects, speaker identification, music, and other "non-speech" information.

The experimental data are composed of Arabic news videos and Arabic documentary films that are recorded from different Arabic channels. These videos contain about 500 scenes. Another set of video scenes are prepared and classified manually to be used as a training set. The learning is done offline before receiving any sample.

Also, an Arabic classification dictionary is developed that helps in constructing the feature vector for each video scenes. About 2000 Arabic words are entered in the dictionary and classified to one or more category of a set of eight predefined semantic categories including politics, economics, sports, religion, social, tourism, weather, health. Each word may be classified to one or more category depending on its meaning. For example, the Arabic word (حكم) pronounced as *Hokm* and means *rule* can be used in both politics and religion so it is classified in the dictionary to both of them. Also, the Arabic word (مصاب) pronounced as *Mosaab* and means *wounded person* is classified to health, social and sports.

Several experiments are conducted in order to measure the performance of the proposed classifier. The standard performance measures of a classification method are the recall and the precision. Recall is the proportion of video scenes in the category

Table 4. Experimental results of classifying video scenes using the proposed framework

	Politics	Economics	Sports	Religion	Social	Tourism	Weather	Health	Total	Precision (%)
# of Scenes	225	31	83	54	74	23	1	16	507	
Politics	220	2	4	0	3	2	0	0	231	95.2
Economics	1	28	0	0	0	0	0	0	29	96.6
Sports	0	0	77	0	2	0	0	0	79	97.5
Religion	0	0	0	51	0	0	0	0	51	100
Social	4	1	2	0	66	1	0	2	76	86.8
Tourism	0	0	0	3	1	20	0	0	24	83.3
Weather	0	0	0	0	0	0	1	0	1	100
Health	0	0	0	0	2	0	0	14	16	87.5
Recall (%)	97.8	90.3	92.8	94.4	89.2	87	100	87.5		

that are actually placed in the category while Precision is the proportion of video scenes placed in the category that are really in the category. The two measures can be defined as follows:

$$\text{Recall}(R) = \frac{\#\text{ of correctly classified scenes}}{\#\text{ of the actual scenes of the corresponding category}}$$

$$\text{Precision}(P) = \frac{\#\text{ of correctly classified scenes}}{\#\text{ of scenes cateogarized as the corresponding category}}$$

These two measures in combination define the so called F-measure. It is an average parameter based on precision and recall. F-measure is a standard statistical measure that is used to measure the performance of a classifier system. It is defined as follows:

$$F - Measure = \frac{2.P.R}{P + R}$$

Table 4 shows the experimental results of classifying 507 video scenes using the proposed framework. For the politics category, 225 political scenes have been entered to classifier. The classifier classified 220 scenes correctly while it classified 5 scenes incorrectly to other categories so the recall will be 220/225= 97.8%. In the other side, it classified 11 scenes as political scenes while they don't belong to this category so the precision will be 220/ (220+11) = 95.2%. As it can be noticed from the table, the proposed framework achieves an average recall rate equal to 92.38% and an average pre cision rate equal to 93.34%. By substituting in the above equation, the F-Measure will be equal to 92.86%. These results show that the proposed framework is efficient for classifying video scenes based on Arabic closed-caption text. Unfortunately, there is no previous work proposed for classifying video scenes using Arabic closed-caption so comparing our work with the others will not be fair-minded. It is just a beginning towards further research in this topic.

7. CONCLUSION

In recent years, video scene classification has attracted much attention in digital video libraries and video summarization systems. This is due to its wide applicability in many applications. Classification is a powerful tool to manage a large number of digital videos and it is the basic step in video content analysis. Grouping video scenes into a set of categories enables efficient store and search for the information need.

This paper 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%. Also, an Arabic classification dictionary was built that contains about 2000 Arabic words with their classifications. It can be incrementally updated so it will be very useful for other researchers.

8. REFERENCES

[1] A. Chen and F. Gey, "Building an Arabic Stemmer for Information Retrieval," In Proceedings of the 11th Text Retrieval Conference (TREC 2002), National Institute of Standards and Technology, 2002.

[2] A. Bosch, A. Zisserman and X. Munoz, “Scene Classification via pLSA," Proceedings of the European Conference on Computer Vision, Graz, Austria, pp. 320-324, May 2006.

[3] F. Chung and B. Y. M. Fung, "Fuzzy Color Quantization and its Application to Scene Change Detection," Proceedings of the 5th ACM SIGMM International Workshop on Multimedia Information Retrieval, Berkeley, California, pp. 157-162, 2003.

[4] H. Sawaf, J. Zaplo, and H. Ney, “Statistical Classification Methods for Arabic News Articles,” Workshop on Arabic Natural Language Processing, ACL2001, Toulouse, France, July 2001.

[5] J. Huang, Z. Liu, Y. Wang, Y. Chen and E. K. Wong, “Integration of Multimodal Features for Video Scene Classification Based on HMM,” IEEE Workshop on Multimedia Signal Processing, Copenhagen, Denmark, 1999.

[6] J. Yang and A. G. Hauptmann, “A Text Categorization Approach to Video Scene Classification Using Keypoint Features," Technical Report, School of Computer Science, Carnegie Mellon University, October 2006.

[7] 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.

[8] L. Khreisat, "Arabic Text Classification Using N-Gram Frequency Statistics A Comparative Study," in Proceedings of The 2006 International Conference on Data Mining, Las Vegas, USA, pp 78-82, June 2006.

[9] 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.

[10] M. Aljlayl and O. Frieder, “On Arabic Search: Improving the Retrieval Effectiveness via a Light Stemming Approach,” In ACM CIKM 2002 International Conference on Information and Knowledge Management, McLean, VA, USA, pp. 340-347, 2002.

[11] M. El-Kourd, A. Bensaid, and T. Rachidi, "Automatic Arabic Document Categorization Based on the Naïve-Bayes Algorithm," Workshop on Computational Approaches to Arabic Script-based Languages, COLING-2004, University of Geneva, Geneva, Switzerland, August 2004.

[12] M. M. Syiam, Z. T. Fayed and M. B. Habib, "An Intelligent System for Arabic Text Categorization," The International Journal of Intelligent Computing and Information Sciences, IJICIS, Vol.6, No. 1, JANUARY 2006.

[13] R. M. Duwairi, "Machine Learning for Arabic Text Categorization," Journal of the American Society for Information

Science and Technology, Volume 57, Issue 8, pp 1005 - 1010, 2006.

[14] S. Rho and E. Hwang, "Video Scene Determination Using Audiovisual Data Analysis," icdcsw, 24th International Conference on Distributed Computing Systems Workshops - W1: MNSA (ICDCSW'04), pp. 124-129, March 2004.

[15] S. Chen, M. Shyu, C. Zhang and R. L. Kashyap, "Video Scene Change Detection Method Using Unsupervised Segmentation and Object Training," The IEEE International Conference on Multimedia and Expo, pp. 15-19, August 2001.

[16] W. Lin and A. Hauptmann, "News Video Classification Using SVM-based Multimodal Classifiers and Combination Strategies," In Proceedings of the Tenth ACM International Conference on Multimedia, Juan-les-Pins, France, pp. 323-326, December 2002.

[17] W. Zhu, C. Toklu and S. Liou, "Automatic News Video Segmentation and Categorization Based on Closed-Captioned Text," In Proc. IEEE International Conference on Multimedia and Expo (ICME), ISBN 0-7695-1198-8/01, 2001.

[18] Y. Zhai, Z. Rasheed and M. Shah, "Semantic Classification of Movie Scenes using Finite State Machines," In Proc. IEEE Vis. Image Signal Processing, Vol. 152, No. 6, pp. 896-901, December 2005.

[19] Y. Li, W. Ming and C.-C. J. Kuo, "Semantic Video Content Abstraction Based on Multiple Cues," IEEE International Conference on Multimedia and Expo (ICME'01), pp. 159-162, August 2001.

[20] Y. Zhu and D. Zhou, "Scene Change Detection Based on Audio and Video Content Analysis," Proceedings of the Fifth International Conference on Computational Intelligence and Multimedia Applications (ICCIMA'03), pp. 229-235, September 2003.

[21] Z. Rasheed and M. Shah, "Scene Detection in Hollywood Movies and TV Shows," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'03), pp. 343-349, June 2003.