Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.12785/amis/090117

Utilizing Query by Example for Fast and Accurate Multimedia Retrieval

Saed ALQARALEH[*] *and Omar RAMADAN*

Computer Engineering Department, Eastern Mediterranean University, GaziMagusa, Mersin 10, KKTC, Turkey

Received: 17 Mar. 2014, Revised: 17 Jun. 2014, Accepted: 18 Jun. 2014
Published online: 1 Jan. 2015

Abstract: In this paper, several techniques have been introduced for improving the performance of Query-By-Example (QBE). First, a new crawler technique has been introduced to collect multimedia descriptive information. Second, a clustering technique is used for creating and updating multimedia databases. In addition, a new re-ranking algorithm which depends on the contents of the multimedia files and some user specific actions has been adapted. Furthermore, parallel technique has been used to decrease the query process time. Finally, elimination of duplicated files in multimedia query results has also been done. Several experiments have been studied and it has been observed that the proposed techniques allow the QBE to supports all multimedia types, increase the percent of relevant files and decrease the process time in query process.

Keywords: Multimedia search engines, query by example, clustering algorithm, crawler technique, re-ranking algorithm, information retrieval

1 Introduction

Multimedia files can be considered as one of the most important materials on the internet. In the last few years, the number of multimedia files and applications which are based on accessing the multimedia on the internet has grown considerably. In [1], special search engine working on extracting information from multimedia files has been developed. In [2], a new mechanism which uses a hybrid method that combines ontology and content based methods was presented for effective searching through multimedia contents. In [3,4,5,6,7], the query by sketch has been developed to improve the performance of the QBE. In [8], content-based image retrieval (CBIR) system which uses Intelligent User Interface Agents to improve query-by-sketch and QBE was presented. In [9], a video retrieval method based on QBE has been developed. The strong point in this approach is that they successfully managed to filter irrelevant video shots. In addition, they have introduced a method which selects example shots without user supervision. On the other hand, the main disadvantage of this work is that the query process time is considerably large. In [10], a novel method to represent video data by developing an optical flow tensor (OFT) and incorporating hidden Markov models (HMMs) was introduced. The main advantage of the methodology of [10], is that the query video shots will not be compared with all shots recorded in the database. However, the precision of this approach dose not exceeds 70% . In [11], a graph transformation and matching approach to find similar contents of a short query clip within a long video sequence was presented. The strong points of this approach are as following: First, finding the smaller set of video candidates related to the query. Second, this approach can deal with videos whose shots have been reordered. Although this can increase the number of candidate videos, it can also lead to get files which have no relation with the required ones.

In addition to the above mentioned problems, and even for the state-of-the-art methods which are used for accessing multimedia files on the internet, we still have the following challenges: First, search engines are still using mainly meta-data or keywords to create multimedia databases. As meta-data does not pay attention to the contents of the file itself and it cannot deal with different meaning of words. In addition, in some cases, there may be no relation between the contents of the multimedia files and their names. This lead to have the outcomes of the query in many cases to contain a large number of irrelevant files. Second, QBE is a powerful technique that

[*] Corresponding author e-mail: saed.alqaraleh@emu.edu.tr

can be used to find images that are similar to what the user already has. However, QBE is not successfully used for other type of multimedia such as videos and audio. Third, any new updated multimedia files on the web-sites require some time to appear on the search engines results. Last but not least, it has been found in [12] that around 40% of the pages and files on the web are duplicated.

The main objective of the work presented in this paper is to improve the overall performance of QBE, to adapt QBE to support all multimedia types and to solve the above mentioned challenges. We have introduced a new crawler technique which visits only the updated pages instead of visiting all pages in each web-site and this will reduce the process time of showing the new data in search engine query results. In addition, a clustering technique is used for creating and updating multimedia databases. Furthermore, a new re-ranking algorithm has been developed which works on the result of the query depending on the multimedia contents and some user specific actions such as download, copy or cut a part of a web-page. Finally, we have adapted the approach created in [24], to eliminate the repeated occurrences of multimedia files in any QBE search engines. It has been observed from the experimental work that the proposed techniques allow the QBE to supports all multimedia types, increase the percent of relevant files and decrease the process time in query process.

The paper is organized as follows: Sections 2 present QBE approach and the developed techniques. Section 3 discuses the experimental results. Finally, conclusions are given in Section 4.

2 Query by Example Approach

In this paper, feature extraction and multimedia ontology have been used to adapt QBE to support all multimedia types. We have used ”multimedia content description interface” features (MPEG-7) [13], which is a tool used for describing the multimedia content. The file is segmented, and the extraction of the file main features is done base on the following descriptors information: layout, color structure, dominant color, scalable color, edge histogram, homogeneous texture, contour shape and region shape. In addition, the ontology-based method is used for detecting the objects of the multimedia files. We have used the information of these descriptors to complement each other and to describe the object appearance in detail.

When the developed QBE is required to work with images, the user can select one of the following options:

a. Find exactly similar images to the uploaded one.

b. Find all images which have a specific percentage of similarity to the uploaded image. For example, if the user selects 50% as percentage of similarity, the approach will find all images in the database which have at least half of its pixels similar to the uploaded file.

As the video and audio files have approximately the same structure, we can use the same technique for dealing with these files. In this case the user can select one of the following options:

a. Find exactly similar files to the uploaded one.

b. If the user has a part of a video or audio, then she/he can get the complete file.

c. Find a higher quality video or audio. For example, the user can find similar files but with different extensions.

d. Find all videos which have a specific percentage of similarity to the uploaded one.

2.1 Techniques Used for Enhancing QBE Performance

2.1.1 Crawling Technique

A web crawler is an automatic web object retrieval system that exploits the web’s link structure. It has two primary goals: 1) to find out new web objects, and 2) to observe changes in previously discovered web objects [14]. Based on the fact that there are a huge number of web-sites with considerably large number of new multimedia files, the main problem in the existing crawling techniques is that a period of time is required for the new files to appear on the query results. In this paper, this time delay problem has been solved by introducing a new script file which will increase the number of web-sites that the crawler can deal with. This script file can be added to any web-site, and will monitor the events of updating the web-pages and prepare a report containing the pages that have been modified. The advantages of the script file is that the crawler will visit only the updated pages in each web-site, and this will increase the number of visited web sites and decrease the required time to show the new data of any query results. In addition, the script file that we created is small in size which will not require any specific requirement on the web server and it will not affect its performance. Furthermore, parallel techniques have been used to increase the number of crawlers working at the same time.

2.1.2 Clustering Technique

Document clustering [15,16] is a technique which can be used to find similar files obtained by search engines. Multimedia files can be organized by using clusters which lead to a faster categorization of the data. Then, the relevant multimedia files can be obtained. In general clustering methods can be either hard, where each object belongs to a single cluster only or soft, where each object can belong to several clusters. In this paper, soft clustering has been used in the developed QBE. The approach we have developed uses clustering techniques to improve the performance of QBE. The clustering

techniques will be used during creating the databases. The approach creates and/or adds files to the database as shown in Figure 1, which can be summarized as follows:

1. Extract the features from the files. In case of image files, the descriptors information mentioned before will be extracted and a Gabor wavelet, which is a linear filter, is used for edge detection. In case of video, in addition to the descriptors information the shots and its objects will be extracted.

2. A preprocessing will be done on multimedia files to remove the unwanted and useless part of each file. For example, the shots will be filtered to find irrelevant ones, by assigning a weight to each shot. The weight is assigned based on the percent of similarity between the shots. All the shots that have a weight under a threshold, which will be specified by experimental work, will be considered as irrelevant ones.

3. Normalizing the features. In this case, all features value will be adjusted to be in the same range.

4. Clustering the multimedia files will be done. In this paper, the following clustering algorithms are used: fuzzy c-means [17], Subtractive [18], spectral [19], hierarchical [20], neural network [21]. However, it should be noted that the presented approach can work with any other clustering techniques.

It is important to note that when a query is done, the features and shots of the uploaded file will be extracted and all similar files in the databases will be shown to the user.

2.1.3 Re-ranking Technique

Ranking web-sites is basically ordering the web-sites in a list displayed as the result of a search query [22]. The ranking of a website depends on the number of hits, the keywords, the website meta-tags and its contents. However, this is unfair with multimedia files which are an important part of the web contents. Both text and multimedia files can contain useful information that should be used in retrieving the web. Recently, a group of researchers have implemented an automatic re-ranking process which works on integrating the keyword and visual features for web images retrieval [22]. This approach, however, has been developed for images only. In [23], another re-ranking image approach has been developed. In this approach, it has been assumed that the images clicked in response to a query are relevant one. The original search results in this approach will be re-ranked to promote images that are likely to be clicked to the top of the ranked list. However, when we are looking for a specific file, some unrelated ones may shown up, and in this case the approach of [23], will lead us to get files which are absolutely has no relation to the required ones.

In this paper, a re-ranking technique has been adapted to improve the user chance of getting more relevant files. After the user executes a query, the proposed technique rank query results depending on its contents and on some of the user specific actions summarized below:

1. If the user downloads a file.

2. If the user tries to copy or cut a file or a part of the file in a web-page.

3. If the user spends more than a number of seconds (N) during checking the file, where N will be specified by the system administrator.

We believe that if a user has done one of the above actions on a specific file, it mean that this file is a related one. Then, the developed technique will analyze and extract the features of this file and it will find the most similar ones. One of the advantages of the developed re-ranking technique is that the re-ranking process does not require any user intervention.

2.1.4 Parallel Implementation of the Developed QBE

In this paper, we have divided QBE approach into the following two parts: The first part is the main web-page, which will be used by the user to upload the multimedia file and to show the results to the user. The second part, which will not be shown to the user, designed to get the uploaded multimedia file and compare it with other files in the databases. We have modified the approach to run multiple copies of the second part of our approach to work in parallel. In this case, if we have only one database the number of multimedia files is divided evenly between the parallel copies. In the case that we have more than one database, which is the situation of search engines, a group of the second part of the approach will work for each database.

2.1.5 Duplicated Multimedia Files Elimination

Recently, an approach which eliminates the repeated occurrences of multimedia files has been developed in [24]. In this paper the approach of [24] has been adapted with the proposed QBE in order to eliminate the duplicated files in query results.

3 Simulation Study

The developed QBE was tested by creating a database by taking randomly chosen multimedia files from internet search engines and creating the database. The most frequently searched keywords and the most viewed multimedia files are taken from [25,26] in the last week of December/2012 have been used in the experimental work.

3.1 Performance of Techniques Used in Creating Databases

First of all, we have done an experiment to find out which clustering algorithm is more appropriate for clustering the

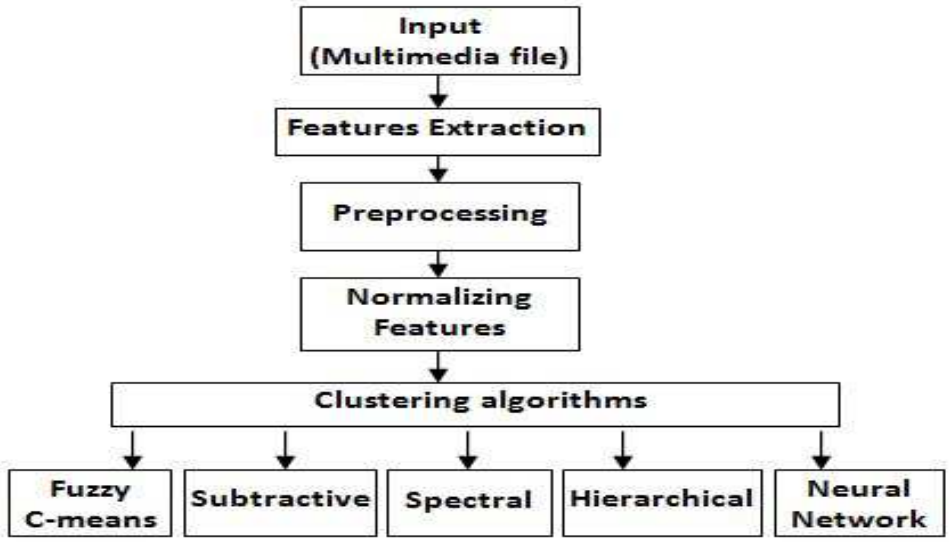


Fig. 1: Steps of creating and or updating the database.

multimedia files. It is important to note that the performance of the algorithms can be effected by the clustered files. Table 1 shows the percentage of relevant files for some queries when the following algorithms are used in creating the database: fuzzy c-means [17], subtractive [18], spectral [19], hierarchical [20], and neural network [21].

It is clear from table 1 that the performances of the clustering algorithms are almost same, and the percentage of relevant files range between 60% and 70% which is not good enough. To improve the quality of the clustering results we have decided to find and use the best combination of the above mentioned clustering algorithms. The Euclidean distance [27], has been used to find the similarity metrics of the clustered files. As it is well known, one of the difficulties in using the clustering techniques is specifying the number of clusters that will be used. We have done an experiment to find the effect of the number of clusters on the percentage of relevant files using the combination of the above algorithms. In this experiment, one hundred million multimedia files have been clustered. In the last part of the experiment, to eliminate the danger of specifying the number of clusters, we have decided to automatically specify the number of clusters based on the percentage of similarity between the files. When a new file is added to the clusters, if its similarity with the previously defined clusters is under a threshold value, then the file will be added to a new cluster. Note that the threshold can be determined by the database administrator. Table 2 shows the effect of clusters number on the percentage of relevant files for some queries. It is clear that the performance of clustering algorithms has been increased, especially when the number of clusters is determined automatically.

In the next experiment, we have studied the performance of the sequential technique where the uploaded file is compared with all files in the database, and the clustering technique with and without the parallel implementation for different number of multimedia files. The result of this experiment is shown in Table 3. It is clear from Table 3 that the clustering techniques have improved the performance of QBE. In addition, a significant improvement has been obtained by using the parallel implementation of clustering. For example, when we have one hundred million files in the database, it takes around 812 seconds using the sequential technique, and 19 seconds is required when clustering techniques are used. On the other hand, by using the developed approach with parallel implementation it takes only 0.89 second. As a result, the process time problem of using QBE applications has been solved.

3.2 Performance of the Crawler, Re-ranking and Elimination of the Duplicated Multimedia Techniques

In the following experiments, the combination of clustering algorithms, where the number of clusters is specified automatically, was used in creating the databases.

3.2.1 Crawler Technique Performance

The performance of the crawler with and without the script file was studied in this section. In this experiment,

Table 1: Percentage of relevant files for some queries using fuzzy c-means, subtractive, spectral, hierarchical, and neural network algorithms.

Video file title	Fuzzy c-means	Subtractive	Spectral	Hierarchical	Neural Network
Gangnam Style	65.5%	68%	68.1%	66.3%	63.6%
Baby	67.5%	67%	68.1%	67.4%	64%
On The Floor	62.3%	64.1%	65.3%	62.9%	60.5%
Love The Way You Lie	61.9%	62.6%	67%	60.6%	61%
Party Rock Anthem	69%	65%	70.2%	62.6%	66.3%
Charlie Bit My Finger - Again!	66.5%	68.7%	69.8%	65.7%	64.9%
Waka Waka (This Time for Africa)	65%	65.9%	70%	69%	65%
Bad Romance	69%	70%	70.5%	69.7%	67.6%
Ai Se Eu Te Pego	67%	63.4%	68.2%	64%	67.1%
Danza Kuduro	64%	65%	70%	63.4%	63%

Table 2: Effect of clusters number on the percentage of relevant files for some queries.

Video file title	Number of clusters				
	2,000	5,000	10,000	25,000	Automatically defined
Gangnam Style	63.1%	65%	67.9%	85.3%	92%
Baby	66%	68%	69.5%	80.1%	91%
On The Floor	70.3%	71%	72.1%	82%	94.2%
Love The Way You Lie	71%	71.9%	72.5%	82.2%	93.1%
Party Rock Anthem	70.2%	71.7%	73%	84.4%	95%
Charlie Bit My Finger - Again!	72.8%	73.6%	74.5%	81%	92.9%
Waka Waka (This Time for Africa)	70%	71.9%	73.1%	82%	94.3%
Bad Romance	68.7%	70%	72.1%	83%	94.1%
Ai Se Eu Te Pego	66%	70.4%	71.7%	82.1%	92.5%
Danzd Kuauro	68%	69.5%	70.5%	80%	91.6%

Table 3: Execution time using the sequential, the clustering technique with and without parallel implementation for different number of multimedia files.

Number of multimedia files (million)	Time(Second)		
	Sequential	Clustering	
		Without parallel implementation	Witn parallel implementation
0.5	109	3	0.12
0.75	147	8	0.19
1	204	11	0.22
10	509	15	0.45
100	812	19	0.89

we recorded the number of web-sites that have been processed in five days. Figure 2 shows the performance of the crawler with and without the script file. It can be seen from Figure 2 that the proposed script file approach increases the number of visited web sites by a factor of approximately 2.5 as compared with the conventional crawler scheme. This leads to decrease the required time to show the new data to approximately less than half of time of the conventional approach.

3.2.2 Re-ranking Technique Performance

The performance of the QBE with and without the developed re-ranking technique has been studied in this section. Figure 3 shows the percentage of the relevant files. It can be shown from Figure 3 that by using the new re-ranking technique, the percentage of relevant files have been increased by approximately 5%, as compared with the convention approach.

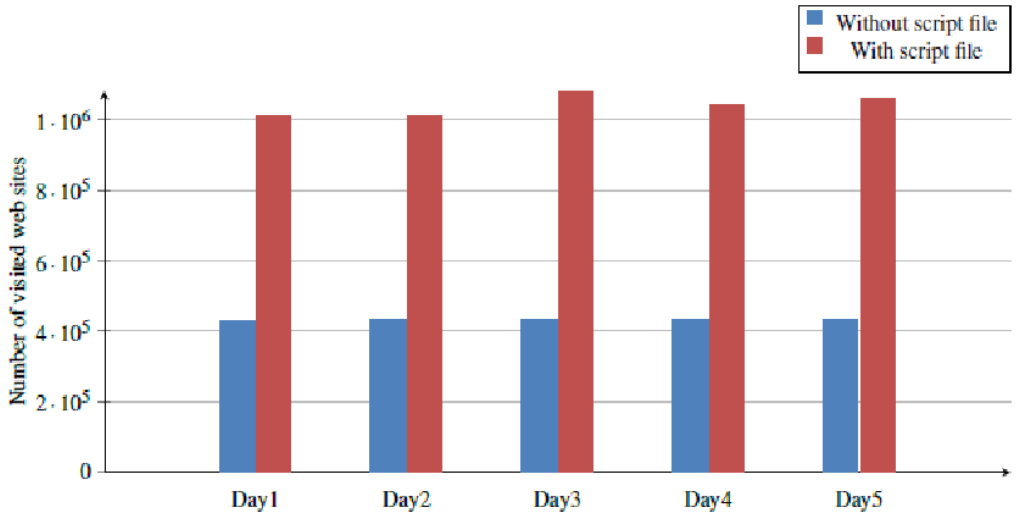


Fig. 2: Crawler performance with and without the script file.

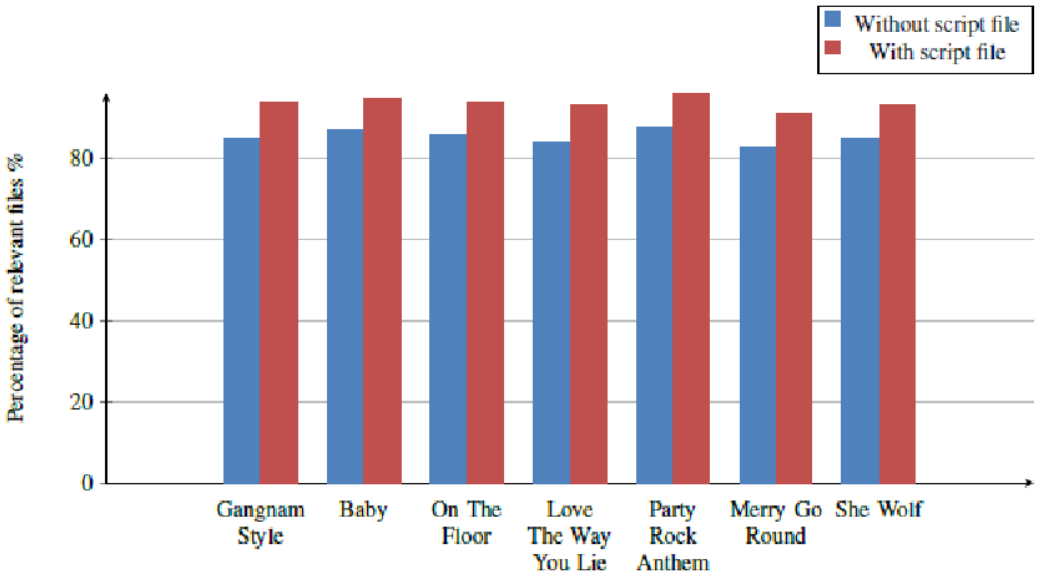


Fig. 3: Percentage of relevant files for specific queries with and without using the developed re-ranking technique.

3.2.3 Elimination the Duplicated Multimedia Files

In this section, we have studied the effect of eliminating the duplicated multimedia files in search engine databases. Table 4 shows the percentage of repeated occurrences and number of result for specific queries for the existing keywords search engines and the developed QBE. It is clear from table 4 that we have successfully managed to eliminate completely the duplicated files.

3.3 Overall Performance of the Developed Query-By-Example

In this section, we have studied the overall performance of the developed QBE by combining all techniques introduced in this paper. The performance of the

Table 4: Percentage of repeated occurrences and relevant files for specific queries for the existing keywords search engines and the developed QBE.

Multimedia type	Multimedia file title	Existing keywords search engines		Developed QBE	
		Results (millions)	Repeated occurrences	Results (millions)	Repeated Occurrences
Images	Hotels	140	35%	91	0%
	Cars	158	32%	108	0%
	Dog	130	39%	80	0%
	Baby	209	40%	126	0%
	Weather	93,6	39%	58	0%
Videos	Gangnam Style	78,6	30%	55	0%
	Baby	16,2	33%	11	0%
	On The Floor	3,7	32%	2.5	0%
	Love The Way You Lie	3,7	29%	2.7	0%
	Party Rock Anthem	0.42	25%	0.32	0%
Audios	One Day	14,8	38%	9.2	0%
	Ai Se Eu Te	6,31	41%	3.8	0%
	Merry Go Round	4,51	40%	2.8	0%
	She Wolf	5,52	39%	3.4	0%
	Diamonds lyrics	5,6	35%	3.7	0%

Table 5: Comparison between Google QBE and the developed QBE for image queries.

Number	Image	The developed QBE			Google QBE			Relevant files gain
		Results	Time (s)	Relevance	Results	Time (s)	Relevance	
1		1150	0.73	90%	1200	0.75	65%	25%
2		1030	0.77	74%	1100	0.79	45%	29%
3		200	0.73	71%	150	0.76	35%	36%
4		1660	0.81	89%	1600	0.80	80%	9%
5		710	0.65	92%	650	0.70	70%	22%
6		540	0.55	71%	410	0.50	50%	21%
7		980	0.58	76%	1080	0.63	43%	33%
8		990	0.61	95%	865	0.60	70%	25%
9		1210	0.68	87%	1290	0.62	65%	22%

Table 6: The efficiency of the developed approach for videos/audios. *(Images are shown for the purpose of explanation)

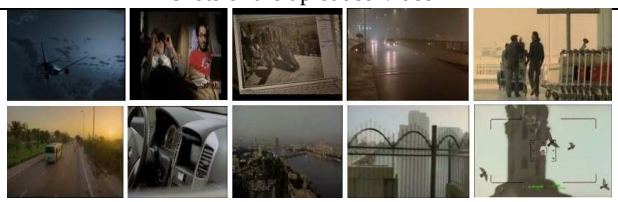
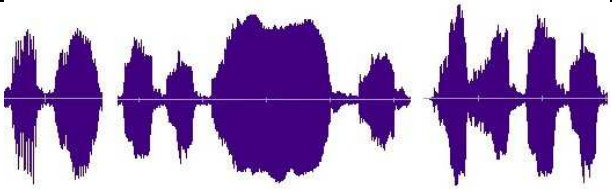
Query 1	shots of the uploaded video* 		
User's options	Number of results (thousand)	Required time (seconds)	Percentage of relevant Files
a. Find exactly similar file	0.97	3.3	100%
b. Find the full video	0.66	1.87	98%
c. Find a high quality video	0.45	1.73	98%
Query 2	shots of the uploaded video* 		
User's options	Number of results (thousand)	Required time (seconds)	Percentage of relevant Files
a. Find exactly similar file	0.75	2.1	100%
b. Find the full video	0.32	0.67	98%
c. Find a high quality video	0.2	0.71	98%
Query 3	shots of the uploaded video* 		
User's options	Number of results (thousand)	Required time (seconds)	Percentage of relevant Files
a. Find exactly similar file	2.57	2.2	100%
b. Find the full video	4.91	3.67	91%
c. Find a high quality video	5.07	3.71	95%
Query 4	Clips of the uploaded audio* 		
User's options	Number of results (thousand)	Required time (seconds)	Percentage of relevant Files
a. Find exactly similar file	0.5	1.9	100%
b. Find the full audio	0.25	1	93%
c. Find a high quality audio	0.35	1.1	91%
Query 5	Clips of the uploaded audio* 		
User's options	Number of results (thousand)	Required time (seconds)	Percentage of relevant Files
a. Find exactly similar file	1	1.75	100%
b. Find the full audio	0.52	0.89	91%
c. Find a high quality audio	0.28	0.59	90%

Table 7: Comparison between the existing keywords search engines versus the developed QBE.

Multimedia type	Multimedia file title	Average of Google, Yahoo and Bing	Developed QBE
		Percent of relevant files	Percent of relevant files
Images	Hotels	88%	91%
	Cars	92%	95%
	Dog	89%	94%
	Girls	91%	97%
	Weather	88%	92%
Videos	Gangnam Style	92%	96%
	Baby	80%	85%
	On The Floor	82%	87%
	Love The Way You Lie	80%	84%
	Party Rock Anthem	80%	88%
Audios	One Day	85%	89%
	Ai Se Eu Te	88%	92%
	Merry Go Round	80%	86%
	She Wolf (Falling To Pieces)	84%	89%
	Diamonds	82%	86%

developed QBE is compared with that of the Google QBE for image queries, as Google QBE support images only. For every query, the number of results, the percentage of relevant files and the query execution time were recorded and shown in table 5. The result shown in this table is found by three independent evaluators. From table 5, we can see the following: 1) the performance of QBE approach has been improved. 2) Although, the performance of Google QBE is good for some files, it is found that it is weak for dealing with images that have empty spaces such as the second and the seventh uploaded image. In addition, the Google QBE was found not efficient in dealing with combined images such as the uploaded image number six.

The efficiency of the developed approach for dealing with videos/audios has also been studied. In this study, we have uploaded a group of multimedia files with different lengths of duration. Table 6 shows a sample of the tested group. It is important to note that the images show in table 6 represent shots of the video and shown for the purpose of explanation.

It can be seen from Tables 6 that the approach is capable of dealing with video and audio files. Also it is clear from query number 3 that the approach can give good results even if the number of the shots which will be extracted from the uploaded video is small.

Finally, we have done an experiment to study the overall performance of the developed QBE versus the existing keywords search engines: Google, Yahoo and Bing. In this experiment, the average of the results of Google, Yahoo and Bing was compared with the result of the developed QBE. The result of this experiment is shown in table 7. It is clear from table 7 that the developed approach increases the percent of relevant files.

4 Conclusions

In this paper, QBE approach has been improved to supports all multimedia types like images, videos and audios. In addition, several techniques such as crawler, clustering and re-ranking techniques have been used to improve QBE. Furthermore, parallel implementation of the proposed approach has been done. Moreover, elimination of duplicated files in multimedia query results has also been done. Several experiments have been included, and it has found that the proposed QBE successfully supports all multimedia types, increases the percent of relevant files and decreases the process time in query process.

References

[1] Doumenis, G. Papastefanos, S. Mateevitsi, V. Andritsopoulos, F. Achilleopoulos, N. Mikhalev, A.V, Video index and search services based on content identification features ,Broadband Multimedia Systems and Broadcasting, 1-4 (2008).

[2] L. YuJie, B. Feng, L. ZongMin, L. Hua, 3D Model Retrieval Based on 3D Fractional Fourier Transform, The International Arab Journal of Information Technology, Online Publication, **10**, (2013).

[3] Y. Watai, T. Yamasaki, K. Aizawa, View-Based Web Page Retrieval using Interactive Sketch Query, Proc. International Conference on Image Processing, **6**, 357-360 (2007).

[4] I. Giangreco, M.Springmann, I. Al Kabary, H.Schuldt, A User Interface for Query-by-Sketch Based Image Retrieval with Color Sketches, Computer Science, **7224**, 571-572 (2012).

[5] Y. Wang , M. Yu, Q. Jia, H. Guo , Query by sketch: An asymmetric sketch-vs-image retrieval system, Image and Signal Processing (CISP), **3**, 1368-1372 (2011).

[6] S. M. Yoon, A. Kuijper , Query-by-sketch based image retrieval using diffusion tensor fields, Image Processing Theory Tools and Applications (IPTA), 343-348 (2010).
[7] R. Vermilyer, Intelligent User Interface Agents in Content-Based Image Retrieval, Proc.IEEE, New York, 136-142 (2006).
[8] C. Doulaverakis, E. Nidelkou, A. Gounaris, Y. Kompatsiaris, A Hybrid Ontology and Content-Based Search Engine For Multimedia Retrieval ,CiteSeerX -Scientific Literature Digital Library and Search Engine (2008).
[9] K. Shirahama, K. Uehara, Utilizing Video Ontology for Fast and Accurate Query-by-Example Retrieval, Semantic Computing (ICSC) Fifth IEEE International Conference, 395-402 (2011).
[10] X. Gao, X. Li, J. Feng, D. Tao, Shot-based video retrieval with optical flow tensor and HMMs, Pattern Recognition Letters, **30**, 140-147 (2009).
[11] H. Shen, J. Shao, Z. Huang, X. Zhou, Effective and Efficient Query Processing for Video Subsequence Identification, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, **21**, 321-334 (2009).
[12] W C. Manning, P. Raghavan, H. Schtze, Introduction to Information Retrieval, online edition, Cambridge University Press, 421-461 (2008).
[13] B. S. Manjunath, P. Salembier, T. Sikora, Introduction to MPEG-7: multimedia content description interface, 84-93 (2011).
[14] M. Kumar, P. Neelima , Design and Implementation of Scalable, Fully Distributed Web Crawler for a Web Search Engine, International Journal of Computer Applications, **7**, 8-13 (2011).
[15] L.Chen, Using a new relational concept to improve the clustering performance of search engines, Information Processing and Management, **47**, 287-299 (2011).
[16] A. Hindle , J. Shao , D. Lin , J. Lu , R. Zhang, Clustering Web video search results based on integration of multiple features, World Wide Web, **14**, 53-73 (2011).
[17] W. Chen, M. Wang, A fuzzy c-means clustering-based fragile watermarking scheme for image authentication, Expert Systems with Applications, **36**, 1300-1307 (2009).
[18] J.Alnihoud, Image Retrieval System with Self Organizing Map and Subtractive Fuzzy Clustering, International Journal on Information and Communication Technologies, **4**, 103-110 (2011).
[19] F. Tung, A. Wong, D. Clausi, Enabling scalable spectral clustering for image segmentation, Pattern Recognition, **43**, 4069-4076 (2010).
[20] V. MURTHY, E.VAMSIDHAR, J. KUMAR, P. RAO, Content Based Image Retrieval using Hierarchical and K-Means Clustering Techniques, International Journal of Engineering Science and Technology, **2**, 209-212 (2010).
[21] B. Sowmya, B. Rani, Colour image segmentation using fuzzy clustering techniques and competitive neural network, Applied Soft Computing, **11**, 3170-3178 (2011).
[22] Y. Zhu, N. Xiong, J. Park, R. He , A Web Image Retrieval Re-ranking Scheme with Cross-Modal Association Rules, International Symposium on Ubiquitous Multimedia Computing, 83-86 (2008).
[23] V. Jain, M. Varma, Learning to re-rank: query-dependent image re-ranking using click data, 20th international conference on World Wide Web, 277-286 (2011).
[24] S. Alqaraleh, O. Ramadan, Elimination of Repeated Occurrences in Multimedia Search Engines, The International Arab Journal of Information Technology, **11**, 134-139 (2014).
[25] Search keywords, http://www.wordstream.com/, last visited (28/8/2012).
[26] Multimedia Search keywords, https://www.wordtracker.com/, last visited (28/8/2012).
[27] A. Sony, K. Ajith,, K. Thomas,, T. Thomas, P. Deepa, Video summarization by clustering using euclidean distance, Signal Processing, Communication, Computing and Networking Technologies (ICSCCN), 642-646 (2011).

Saed ALQARALEH received his B.Sc degrees in the Software Engineering, Al-Hussein Bin Talal University (AHU),Jordan, in 2008, and his M.Sc.degrees in the Computer Engineering,Eastern Mediterranean University (EMU), Northern Cyprus, in 2011. He is currently a Ph.D.student in the Computer Engineering,Eastern Mediterranean University (EMU), Northern Cyprus.

Omar Ramadan received the B.Sc., M.Sc., and Ph.D. degrees in the electrical and electronic engineering from Eastern Mediterranean University (EMU), Gazimagusa, Mersin, Turkey, in 1992, 1994, and 1999, respectively. From 1992 to 1999, he was a Research and Teaching Assistant with the Electrical and Electronic Engineering Department, EMU. In 1999, he joined the Computer Engineering Department, EMU, as an Assistant Professor, and then became a Professor in 2009. He has authored or coauthored over 80 scientific journal and international conference papers. He has supervised three masters theses, and one doctoral thesis and currently supervises two masters theses and one doctoral thesis. His research interests include computational electromagnetic, digital signal processing, parallel computing, information retrieval and search engines. Dr. Ramadan is a reviewer for many international and national journals.