Multimed Tools Appl (2013) 64:77–95
DOI 10.1007/s11042-011-0848-4

An integrated semantic-based approach in concept based video retrieval

Sara Memar · Lilly Suriani Affendey · Norwati Mustapha · Shyamala C. Doraisamy · Mohammadreza Ektefa

Published online: 19 August 2011

Abstract Multimedia content has been growing quickly and video retrieval is regarded as one of the most famous issues in multimedia research. In order to retrieve a desirable video, users express their needs in terms of queries. Queries can be on object, motion, texture, color, audio, etc. Low-level representations of video are different from the higher level concepts which a user associates with video. Therefore, query based on semantics is more realistic and tangible for end user. Comprehending the semantics of query has opened a new insight in video retrieval and bridging the semantic gap. However, the problem is that the video needs to be manually annotated in order to support queries expressed in terms of semantic concepts. Annotating semantic concepts which appear in video shots is a challenging and time-consuming task. Moreover, it is not possible to provide annotation for every concept in the real world. In this study, an integrated semantic-based approach for similarity computation is proposed with respect to enhance the retrieval effectiveness in concept-based video retrieval. The proposed method is based on the integration of knowledge-based and corpus-based semantic word similarity measures in order to retrieve video shots for concepts whose annotations are not available for the system. The TRECVID 2005 dataset is used for evaluation purpose, and the results of applying proposed method are then compared against the individual knowledge-based and corpus-based semantic word similarity measures which were utilized in previous studies in the same domain. The superiority of integrated similarity method is shown and evaluated in terms of Mean Average Precision (MAP).

Keywords Video retrieval · Semantic knowledge · Content-based analysis · Similarity · Search

S. Memar (✉) · L. S. Affendey · N. Mustapha · S. C. Doraisamy · M. Ektefa
Department of Computer Science, University Putra Malaysia,
43400 Serdang, Selangor, Malaysia
e-mail: sr.memar@gmail.com

1 Introduction

In recent years, there has been a tremendous need to query and process large amount of data that cannot be easily described such as video data. Text-based and content-based methods are considered as two fundamental frameworks for video retrieval. Research on text-based methods began in 1970, and they are in relevant with information retrieval community. Moreover, content-based methods are applied in order to improve multimedia retrieval, and it is traced back to 1980s by introducing Content-based Image Retrieval (CBIR). So, the question that arises is the degree of effectiveness of content-based methods in the area of Multimedia Information Retrieval (MIR). Content-based methods can play essential roles when text annotations are not available or are not sufficient enough. In addition, in spite of having annotation, content-based methods give additional insight into media collection, and enhance retrieval accuracy.

Unlike text retrieval systems, video retrieval has encountered one of the most important challenging problems, named Semantic Gap. This is the difference between the low-level representation of videos and the higher level concepts which a user associates with video [16]. The video analysis community has taken beneficial steps towards bridging this gap by utilizing low-level feature analysis (motion, shape, texture, color histograms) and lately by using semantic content description of video, particularly when the content of video is pertaining to broadcast news. However, because the semantic meaning of the video content cannot be expressed in this way, these systems had a very limited success with the earlier mentioned approaches for semantic queries in video retrieval. Several studies have confirmed the difficulty of addressing information needs with such low-level features [24, 32]. However, low-level features have shown promising performance in video retrieval, and they can be utilized in complement to high-level semantic concepts for improving retrieval.

Lately, a number of concept detectors such as outdoors, face, building, etc., have been developed by different researchers to help with the semantic video retrieval. Among them, Large Scale Concept Ontology for Multimedia (LSCOM) with the collection of more than 400 concept annotations , Columbia374 [41] and Vireo374 [18] are considered as the largest and the most popular concept detectors. Therefore, the retrieval of desirable concept is accomplished by using the suitable concept detector and coming up with detection confidences for all video shots. After that a sorted list containing confidences of video shots is returned as result. Despite the fact that the mentioned supervised training for concept detection is desirable, providing annotations of concepts in videos manually is a very challenging and time consuming task, and it is not considered as a suitable approach for retrieving every concept in real world. Therefore, retrieving concepts provided that their annotations are not available is essential. This can be achieved through the computation of similarity measures.

Semantic similarity measures can be beneficial to bridge the gap between an arbitrary textual query and a limited vocabulary of visual concepts [13]. These similarity measures have been utilized in the area of video retrieval as well [1, 13]. Concept-retrieval, and it has been paid attention a lot recently. The quality of similarity measure employed for mapping textual query terms to visual concepts is considered as a key factor in concept-based retrieval. Various studies have made

use of different similarity measures individually. Our study on the other hand investigates the results of integration of various semantic similarity measures. Thus, the integration of semantic similarity measures is applied instead of individual semantic similarity measures in order to retrieve video shots for queries expressed in terms of semantic concepts. The proposed integrated semantic based approached combines multiple semantic similarity measurements, as previous works mostly used these measurements individually.

The remainder of this paper is organized as follows. In Section 2, an overview of related work is given. Then, in Section 3, the proposed video retrieval model is presented in details. The proposed model is evaluated experimentally in Section 4. In Section 5, results and analysis are discussed. The conclusion of the paper is presented in Section 6 with an outlook to future work.

2 Related work

2.1 Information retrieval

As the computerized documents have been increasing dramatically, the essence of retrieving documents, which are stored in databases, is inevitable. In order to make access to documents more intuitively, the field of information retrieval was revealed with the aim of retrieving documents that covers users' information needs [3]. The effectiveness of IR systems is measured by assessing the relevance.

The gap between the computational matching of documents and the way users' information need is expressed leads to semantic gap. In this manner, authors in [20] considered the role of IR system as "a retrieval system which captures the relevant relation by establishing a matching relation between the two expressions of information in the document and the request, respectively". In this definition, information shows the degree of satisfaction achieved by users. Therefore, meeting the different characteristics of information and coming up with better retrieval effectiveness are two significant factors which contribute to the construction of a retrieval system.

Various models in IR try to enhance the effectiveness of text documents retrieval. Although text retrieval is different from other multimedia content, most IR models are general, and they can be applied in other types of multimedia. Here, some classical IR models are reviewed briefly.

One of the most commonly used models for information retrieval is Boolean retrieval model which is based on such logical operators as "and", "or" and "not". This model is very prevalent in data retrieval in which the query parameters and database attributes are exactly matched. However, after coming up with results, since relevant documents must be identified by users, it will be demanding for queries that return large answers. To remove the limitation of the Boolean model, another model called vector space retrieval model was suggested for text retrieval [35]. In this model, non-binary weights are assigned to the index of documents and terms. So, the similarity between index terms and query is calculated by these binary weights. The more well-qualified the degree of similarity, the higher the relevance for documents. Nevertheless, it leads to high dimensionality of the index term space. In order to decrease the dimensionality of the index term space, the

latent semantic indexing retrieval model is proposed [7]. In this model, at first, a term correlation matrix is made by TF-IDF weights and then singular value decomposition is applied to the index term matrix. The largest singular values present index terms with lower dimensionality. For measuring similarity in this model, the query as a pseudo-document should be modeled, and the most similar documents in the projected concepts space should be found. The extended Boolean retrieval model was recommended in [33]. This model is the extension of the traditional Boolean queries with a few modifications in the conjunctive "and" and disjunctive "or" operations towards a vector space model. Authors in [31] presented the binary independence retrieval model. This naive Bayes model assesses the probability that a user will find a document in relevance to the search task. The model considered an index term as a binary value, and it is based on the assumption that the probability of appearance for each term is independent of the other index terms. Ranking of the relevant results is accomplished by minimizing the likelihood of a false judgment as it exploits the ratio of the two probabilities: a document is a relevant set and a document is an irrelevant one. The Bayesian inference network model was suggested in [39], in which random variables are assigned with the index terms, texts, documents, query concepts and queries. In this model, random variables of documents are considered as observation of the document in the search process. The network is constructed from parent nodes and extended to nodes by edges in an acyclic manner. Each node indicates the random variable, and each parent node contains prior probabilities. The graph is constructed on conditional independence, i.e., the nodes are conditioned on their parents and vice versa. In the graph, conditional relationships are presented by edges. This model can be suitable for different information retrieval ranking models, like Boolean and TF-IDF.

Some IR models are based on fuzzy sets. The Fuzzy information retrieval model [27] is on the basis of an assumption where the relevance of a document is defined as a degree of membership and the query is modeled as a fuzzy set of terms. The degree of membership is a value between 0 and 1, where 1 indicates full membership and 0 shows non-existing membership. The real advantage of the fuzzy sets is pertaining to the operators that combine separate sets using set-theoretical operations similar to Boolean model. The models explained here have been foundations for the modern IR methods, and they are considered as traditional IR strategies. More details of these methods can be found from the literature of [6, 20].

2.2 Content-based multimedia information retrieval

Multimedia retrieval has attracted much attention in the last decade. Multimedia is regarded as the combination of at least two of the following formats: text, audio, animations, images, video images. Video data is the integration of audio and motion image tracks. The term video usually refers to various storage formats for motional pictures.

As was mentioned earlier, studies in CBR were revealed in order to address problems involved with database management systems. The first years of MIR were related to computer vision algorithms which concentrated on feature-based similarity search over images, audio and video. Famous prototypes of these systems are Virage

[2] and QBIC [8]. Recently, due to the popularity of Internet, search task for images, videos, etc was directed through the Internet and web. Therefore, Internet image search engines such as Webseek [37] and Webseer [9] were proposed.

One of the prominent and novel works in automatic search is [14]. A robust local analysis approach, called PLF which is able to come up with retrieval accuracy without adopting most top-ranked relevant documents, was proposed. Moreover, automatic video retrieval has been done based on query-class model, and the effectiveness of this model has been showed by experiments of this work. In this model, firstly, a query is classified into one of the categories defined prior. Categories are named as follows: named person, named object, general object, scene and sports. While a query is classified, the ranking features of several modalities are mixed with associated weights of query-class; therefore, they can be utilized for unseen queries since they are able to be automatically introduced as part of the one of the predefined categories. Although the retrieval results based on PLF and query-class model are satisfactory, linking external semantic knowledge sources such as ontology into PLF can lead to better retrieval performance. In [40] a model, called QUCOM (query-concept-mapping) for mapping semantic concepts to queries automatically was proposed. Furthermore, they indicated that solving this problem based on both image and text are more effective in automatic search task of TECVID 2006 using a large lexicon of 311 learned semantic concept detector. The strength of this work is that, by QUCOM, all concepts are not used for search task because some of these concepts may be irrelevant and reduce the performance of retrieval. Retrieval based on QUCOM obtains the state of the art performance for query-by-concept retrieval [4, 5]. As stated by authors, by combining query classification model with QUCOM, better retrieval results will be achieved, too. The authors in [38] suggested an automatic video retrieval method in which three individual methods namely, text matching, ontology querying, and semantic visual query, are applied to select a relevant detector among a set of machine learned concept detectors. Although the retrieval results are good, it has a limitation that all these three methods choose exactly one detector. On the other hand, selecting multiple detectors can get a higher average precision score than a single one.

Managing information means many things, including analysis, indexing, summarizing, aggregating, browsing and searching [36]. As long as researchers improve new innovations and strategies in the area of content-based video retrieval, evaluation of those new strategies becomes very worthwhile. Therefore, all tasks pertaining to video information retrieval have been evaluated since 2001 by TREC Video Retrieval Evaluation (TRECVID), which is an annual benchmarking campaign under the National Institute of Standards and Technology (NIST). TRECVID promotes progress in content-based retrieval from digital video via open, metric-based evaluation [11]. A number of tasks are introduced in TRECVID as follows: shot boundary detection, story segmentation, semantic feature extraction, and search.TRECVID identifies three kinds of search task, including automatic, manual and interactive. Automatic search task is the focus of this study. The main goal of each search task is to retrieve shots which are relevant to user's information need. The TRECVID search task is defined as follows: given a multimedia statement of information need (topic) and the common shot reference, return a ranked list of up to 1000 shots from the reference which best satisfy the need [14].

2.3 Semantic similarity

Similarity is a complex concept which has been widely discussed in the linguistic, philosophical and information theory communities [12]. Semantic types were discussed in terms of two mechanisms: the detection of similarities and differences [10]. Measures of text similarity have been used for a long time in applications in natural language processing and related areas. One of the earliest applications of text similarity is perhaps the vectorial model in information retrieval, where the document most relevant to an input query is determined by ranking documents in a collection in reversed order of their similarity to the given query [34]. An effective method to compute the similarity between words, short texts or sentences has many applications in natural language processing and related areas such as information retrieval, and it is regarded as one of the best techniques for improving retrieval effectiveness [28]. In image retrieval from the Web, the use of short text surrounding the images can achieve a higher retrieval precision than the use of the whole document in which the image is embedded [6]. The use of text similarity is beneficial for relevance feedback and text categorization [21, 23], text summarization [22, 30].

Recently, text similarity or semantic similarity of words has been utilized in the area of video retrieval. There have been a large number of studies on word-to-word similarity metrics ranging from distance-oriented measures computed on semantic networks, to metrics based on models of distributional similarity derived from large text collections [25]. Word-to-word similarity metrics are mainly divided into two groups as follows: knowledge-based and corpus-based. Two prominent works in video retrieval which utilized semantic similarity of words are [1] and [13]. In [1], knowledge-based and corpus-based semantic word similarity measures, in addition to visual co-occurrence, were utilized through trained concept detector in an unsupervised manner in order to solve video retrieval problem by employing concepts from Natural Language Understanding. By using semantic word similarity, the authors tried to map query terms which stated in natural language with visual concepts. Another work is [13] in which the author only used knowledge-based semantic word similarity measures. In [13], since knowledge-based semantic word similarity measures is based on WordNet, some query terms could not be mapped due to the lack of information content (IC); therefore, the authors tried to cover this problem by presenting an approach for determining information content from two web-based sources, and demonstrated its application in concept-based video retrieval. In [19, 26], a corpus-based semantic measure was applied for computing the similarity between two terms or concepts in concept-based video retrieval. This measurement named Flickr Context Similarity (FCS) and it is based on the number of Flickr images associated with concepts.

In this study, firstly, seven of knowledge-based similarity measures which were used individually in [1], were integrated. Secondly, four of corpus-based semantic words similarity measures which were used individually in [1], and Flickr Context Similarity (FCS) [19, 26] were also integrated as a main contribution in this paper to automatically compute a score that shows the similarity of two input words at semantic level.

3 Video retrieval model

This study is mainly categorized in the search task of TRECVID. In search task, a semantic concept is given as the query, and the system should return the ranked list of documents (shots in our case) contributing to the query. Then results are stated in terms of MAP for all submitted queries. One advantage of TRECVID in regard to video search and retrieval is that it brings all groups and approaches together under the same metric-based evaluation for comparison and repeated experiments.

3.1 System overview

Different issues and parts, which are designated in the video retrieval model, are explained in this section in order to give an additional insight into the process of this study.

As Fig. 1 indicates, firstly, a query in terms of semantic concept is given to the system as input by user. Semantic concepts, which are stated as the queries, are completely new for the system. It means that no annotations are available for these semantic concepts. Moreover, it is essential to build statistical detectors for a set of semantic concepts. Since it is not possible to provide annotation for every concept in the world, just a set of semantic concepts are selected for constructing a predefined concept lexicon. These predefined semantic concepts were formerly annotated under the Large Scale Concept Ontology for Multimedia (LSCOM). Here, Columbia374 [41] is regarded as an automatic concept detector in this study. More details about the structure of Columbia374 can be found in [41]. Secondly, queries which are in terms of new semantic concepts, and there are no annotations for them, should be automatically mapped onto annotated concepts of Columbia374. Columbia374 as an automatic concept detector, acts as a trainer, and it provides scores and confidences for 374 annotated concepts on TRECVID 2005 dataset. Hence, performance of search task is enhanced significantly by mapping. For instance, a query like “rainy scene” is in close relationship with such concepts as “weather” or “cloud”. A controversial debate which arises here is how to map the annotated concepts onto user-defined queries automatically. So, the next step is mapping which will be discussed in Section 3.2.

After mapping query with trained concept detectors, confidence scores should be computed for queries. Therefore, we follow the strategy applied in [1] with some manipulations for computing the confidence score for queries in terms of new semantic concept. More details for retrieval of semantic concepts are provided in Section 3.3.

3.2 Mapping query to pre-defined concepts

Semantic similarity of words was utilized for retrieving video shots for concepts whose annotations are not available [1, 13]. Semantic similarity of words is divided into two main groups, namely knowledge-based and corpus-based [25]. In knowledge-based measures, relatedness of two words is semantically computed based on information drawn from semantic networks, i.e., WordNet hierarchy.

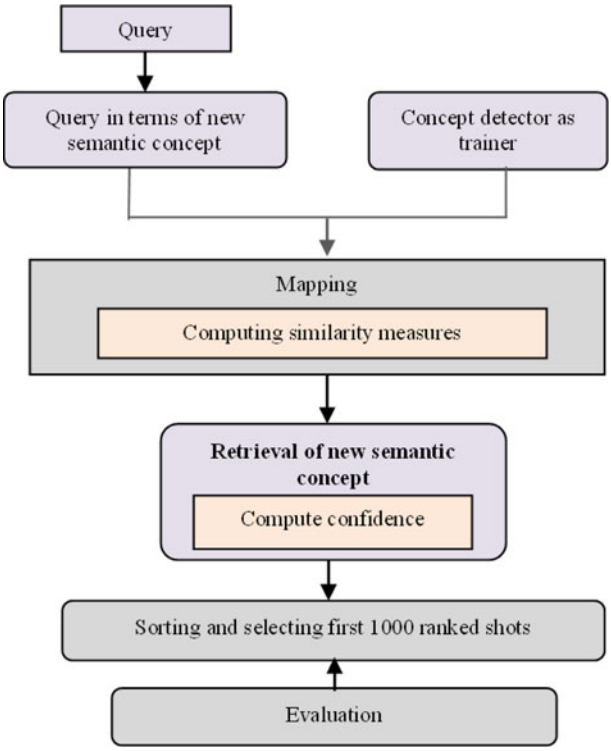

Fig. 1 Design of video retrieval model

In corpus-based similarity measures, the degree of similarity between words is computed based on information drawn from large corpora. In [1], seven measures of knowledge-based and four measures of corpus-based semantic word similarities were utilized individually for computing similarity or mapping, and no integration or combination of these measures was done. Moreover, Flickr Context Similarity (FCS) which is a corpus-based measure [19, 26] was used for such purpose in concept-based video retrieval as well. In this paper, firstly, knowledge-based similarity measures are integrated and secondly, corpus-based similarity measures are integrated, and they are applied for computing the similarity between queries and 374 concepts of Columbia374 concept detectors.

As well as [25], integration is done using a linear combination with an equal weight for each semantic measure. Linear combination is a weighted sum of some treatments. There are some reasons for applying a linear combination in integrating different semantic similarities. When relationships between the variable are linear like our case, linear combination leads to optimal results. In addition, since the result of all semantic similarity measures is numeric, linear combination is an appropriate alternative to be applied for integrating numeric outputs. In this way, queries which are in terms of semantic concept are mapped with pre-defined Columbia374 concepts. Then, the result of similarity is used in order to compute the confidence score of shots for submitted queries.

3.3 Retrieving semantic concept

The result of mapping is similarity measure between queries expressed in terms of new semantic concept and the annotated concepts of Columbia374. This similarity measure is used for computing the confidence score for all video shots. The confidence score shows the degree of relevancy between query and shot. Therefore, for each query, the confidence score of all shots should be computed. The utilized dataset in this study is TRECVID 2005. TRECVID 2005 consists of development set and test set which will be explained in details in Section 3.4.

The similarity of each new concept (c_n)—concept whose annotations are not available—is measured with annotated concepts which are provided by Columbia374 concept detector. Columbia374 acts as a learner for new concepts, and it is presented as $C = \{c_i\}_{i=1}^{R}$, where R is the number of all annotated concepts which is assumed at most 374 due to Columbia374; c_i is the *ith* concept in the corpus. Video shot database is presented as $D = \{S_k\}_{k=1}^{N}$, where N is the number of all video shots which is at most N= 64256 for test set. Then, the similarity measure is utilized in (1). Authors in [1] computed similarity measure between two concepts by utilizing seven of knowledge-based semantic word similarity, four of corpus-based semantic word similarity and visual co-occurrence. In the proposed model, for mapping, the integration of corpus-based measures and knowledge-based measures are applied and used in (1) as well.

$$Score_{cn}(S_k) = \frac{\sum_{i=1}^{R} sim(c_i, c_n) \times Scoreci(S_k)}{\sum_{i=1}^{R} Score(c_i)} \tag{1}$$

Since the number of video shots for test set of TRECVID 2005 is 64256, there will be 64256 confidence scores for each query. Then, all confidence scores should be sorted, and the first 1000 ranked shots should be selected for evaluation purpose.

3.4 TRECVID dataset

Promoting progress in content-based retrieval from digital video via open, metric-based evaluation is one of the principle aims of TREC Video Retrieval Evaluation (TRECVID). Therefore, for evaluation, automatic search development and test dataset of TRECVID 2005 are applied for search task. This dataset consists of 160 hours of multilingual television news collected in November 2004 from Chinese, Arabic and American news channels. The first 80 hours is considered as development set for such tasks as search, high/low level feature, and short boundary detection. The remaining 80 hours is related to test set. LSCOM provided annotations for the development set of TRECVID 2005 by determining the presence and absence of more than 400 concepts in each shot of development set. Columbia374 provides confidence scores for only 374 annotated concepts.

Since knowledge-based semantic word similarity measures should be experimented, and they are based on WordNet, a subset of 183 concepts which are available in WordNet is used, and this set is named fractional set. Another subset contains all 374 concepts of Columbia374, and it is called full set. Distribution and occurrences of concepts for both sets in TRECVID 2005 development set are analyzed and indicated

Table 1 Distribution and occurrences of concepts for each set

Set	Number of concepts	Least frequent concept	Most frequent concept	>5000	<5000	<20
Fractional set	183	10	41202	21 (11.47%)	88%	10
Full set	374	10	41202	36 (9.6%)	90%	20

in Table 1. As can be seen from Table 1, in both fractional and full sets, Person and Water-Tower are the most frequented (41202 0ccurrences) and the least frequented (10 occurrences) concepts, respectively. Moreover, in fractional set, only 21 concepts (11.47%) occurred in more than 5000 shots, and the occurrence of 88% of concepts is less than 5000. There are also 10 concepts which occurred in less than 20 shots.

For full set, while 90% of concepts occurred in less than 5000 shots, there are only 36 concepts (9.6%) that are in more than 5000 shots. The occurrences of 20 concepts are less than 20 as well. The main aim of data analysis is giving an additional insight into the dataset and determining the strength of contextual relation in order to come up with more precise retrieval results.

4 Experimental setup

A set of experiments are conducted on the TRECVID 2005 (TV05) search dataset [11] to evaluate the effectiveness of the proposed video retrieval model. A set of 374 concept detectors, namely Columbia374 [41], is also used. These detectors are trained based on the development set of TRECVID 2005, which is annotated by LSCOM. WordNet similarity package [29] is also applied for implementing knowledge-based semantic similarity measures. There are 10 test concepts defined by NIST as queries. In this study, only 7 out of these 10 concepts are selected due to knowledge-based semantic similarity measures. As mentioned before, knowledge-based measures are based on WordNet, and among 10 test concepts, only 7 of them are available in WordNet. These 7 test concepts are as follow: Car, Explosion-fire, Maps, Mountain, Prisoner, Sports, and Waterscape-Waterfront. The video retrieval model should return the ranked list of up to 1000 results for each query. The ground truth for all 7 queries is provided by NIST as well. A variety of measures for evaluating the performance of information retrieval systems have been presented. These measures need a collection of queries and documents. TRECVID utilizes a set of established measures to evaluate the effectiveness of retrieval results. Returning a ranked list of up to 1000 shots from the test collection, contributing to a set of target topics, is regarded as results. A common measure of retrieval effectiveness called "non-interpolated average precision" is also defined by NIST over a set of retrieved documents (shots in our case are considered as the unit of testing and performance assessment). Average precision emphasizes on ranking relevant documents higher. It is the average of precision of the relevant documents in the ranked order. Let R_j be the number of true relevant documents in a set of size S; L is the ranked list of documents returned. At any given index j, let be the number of relevant documents

in the top j documents. Let $I_j = 1$ if the jth document is relevant and 0 otherwise. Assuming the non-interpolated average precision (AP) is then defined as [15]:

$$\frac{1}{R}\sum_{j=1}^{S}\frac{R_j}{j}\times I_j \tag{2}$$

Finally, the mean of average precision is calculated for all queries which are considered as an overall performance measure.

5 Results and analysis

In this section, firstly, the integration of seven knowledge-based semantic word similarity measures is experimented on fractional set and compared directly with [1] in which these seven knowledge-based measures were utilized individually. Afterwards, the integration of corpus-based measures is implemented on both fractional and full set, and compared with individual corpus-based measures utilized in [1, 19, 26].

5.1 Integration of knowledge-based measures

Table 2 shows the performance of different knowledge-based similarity methods for fractional set with Columbia374. As can be seen from Table 2, last row presents the result for the integration of seven knowledge-based measures which is proposed in this paper, and other areas correspond to the results for seven knowledge-based measures which were applied individually in [1]. LESK and JCN, with MAP of 12.53% and 12.51% respectively, outperformed other knowledge-based similarity measures. The reason may be that such similarity measures as LCH, WUP, LIN and RESNIK cannot work well on any combination of part of speech.

Unlike [1], in this study, all seven knowledge-based similarity measures are integrated and utilized for retrieving semantic concept. Integration of knowledge-based similarity measures outperformed individual knowledge-based similarity metrics with MAP of 12.69%. Average precision-recall curves for knowledge-based methods with fractional set on TRECVID 2005 test set are shown in Fig. 2. Average-precision for each query and semantic retrieval method is also shown in Fig. 3. As mentioned before, 7 out of 10 test concepts are applied as queries in the experiment. Among all queries, the concept "car" has the best Average Precision (AP) using all methods, whereas "Explosion-fire" and "Prisoner" have the lowest retrieval results. Moreover,

Table 2 Performance of knowledge-based measures in retrieving semantic concept for fractional set with Columbia374

Method name	MAP
LESK	12.53%
JCN	12.51%
RESNIK	11.09%
HSO	10.71%
LIN	10.11%
WUP	4.34%
LCH	4.16%
Integration of knowledge-based measures	12.69%

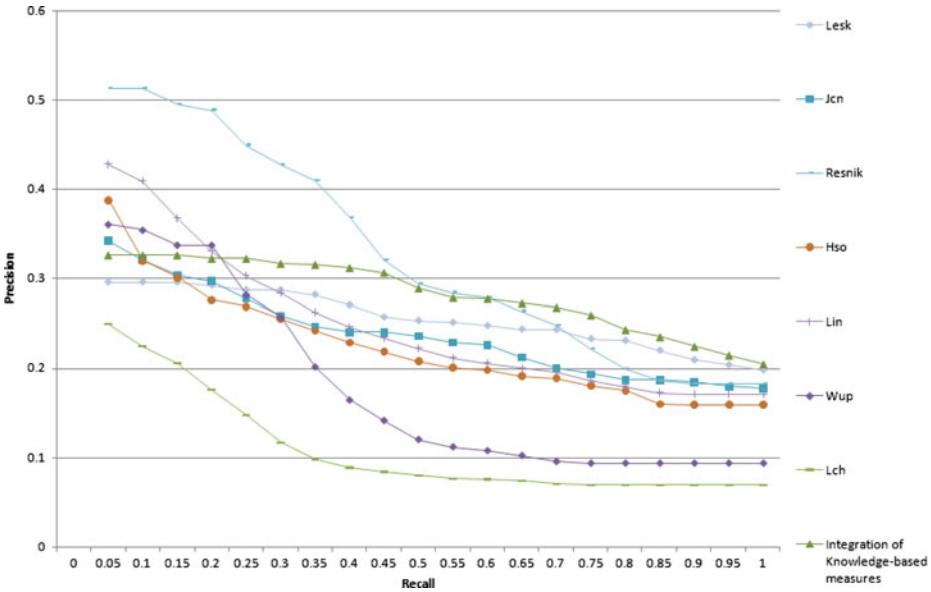

Fig. 2 Precision-recall curves for knowledge-based methods on TRECVID 2005 for fractional set with Columbia374

the mean of average precision for all queries is calculated and indicated at far right. As can be viewed from Fig. 3, among all semantic retrieval methods for fractional set, integration of knowledge-based with MAPs of 12.69% outperformed individual knowledge-based semantic retrieval measures.

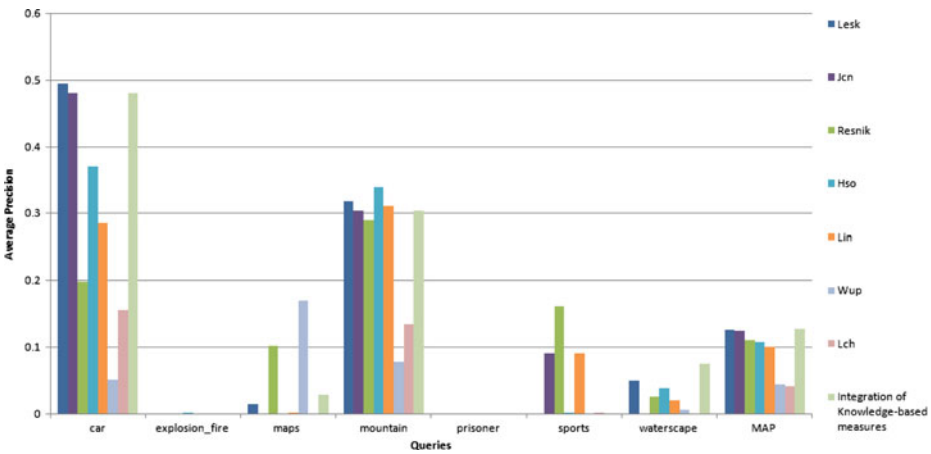

Fig. 3 Average Precision results for each query using knowledge-based methods with fractional set

Table 3 Performance of corpus-based measures for both fractional and full sets with Columbia374

Method name	MAP(fractional set)	MAP(full set)
PMI-IR-WebNEAR	11.09%	9.47%
PMI-IR-WebAND	10.72%	3.3%
PMI-IR-WebImage	12.37%	9.45%
PMI-IR-WebFLCKR	10.69%	4.85%
Flickr Context Similarity (FCS)	14.54%	11.48%
Integration of corpus-based measures	16.73%	12.53%

5.2 Integration of corpus-based measures

In corpus-based similarity measures, similarity between words is determined using information drawn from large corpora. Corpus-based similarity measures can be implemented and tested on both fractional set and full set because they are not based on WordNet.

Table 3 shows the performance of corpus-based measures for both fractional and full sets. As can be viewed from Table 3, last row indicates the result for the integration of five corpus-based measures which is proposed in this study, and other areas are related to the results for corpus-based measures which were utilized individually in [1, 19, 26]. For fractional set, FCS [19, 26] performs very well with MAP of 14.54%. Among corpus-based similarity measures utilized in [1], PMI-IR-WebImage and PMI-IR-WebNEAR are top 2 methods with MAPs of 12.37% and 11.09%, respectively. Afterwards, the integration of all corpus-based similarity

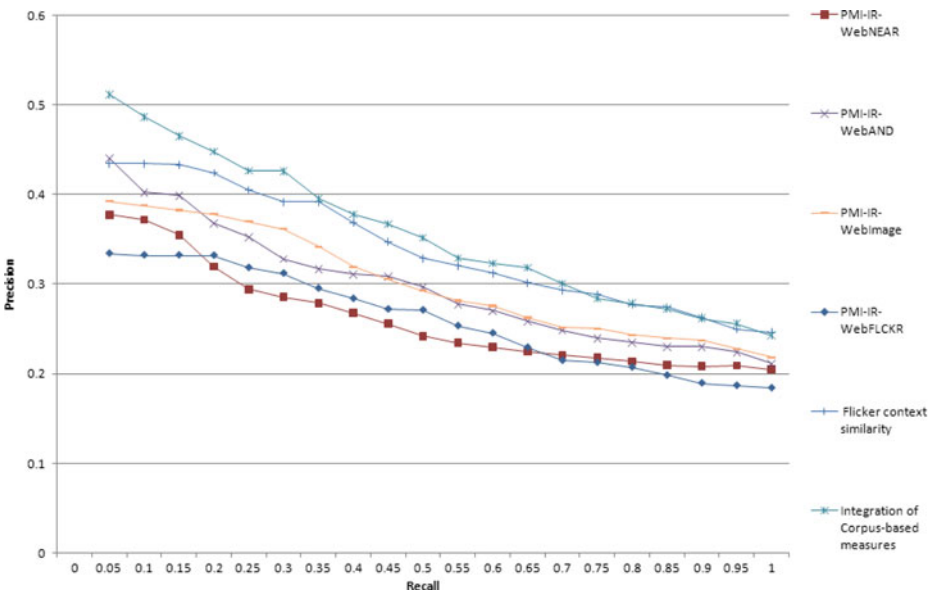

Fig. 4 Precision-recall curves for corpus-based methods on TRECVID 2005 for fractional set with Columbia374

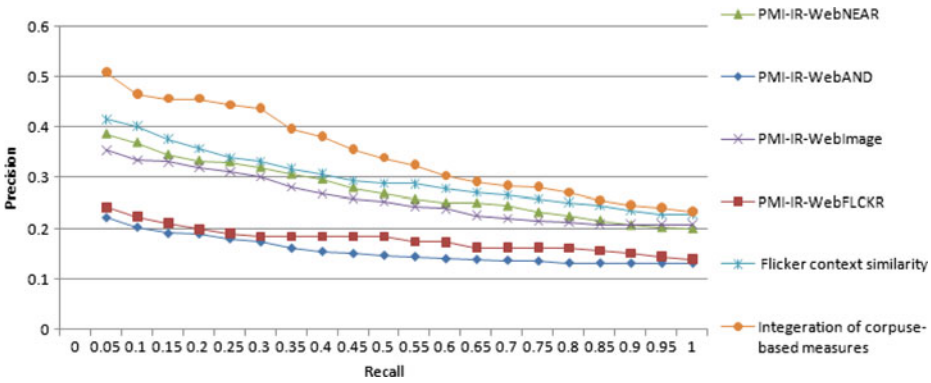

Fig. 5 Precision-recall curves for corpus-based methods on TRECVID 2005 for full set with Columbia374

measures is applied for retrieving video shots for test semantic concepts, and it outperformed all individual corpus-based methods with MAP of 16.73%.

In full set, among four corpus-based similarity measures used in [1] individually, PMI-IR-WebNEAR and PMI-IR-WebImage are top two similarity measures with MAPs 9.47% and 9.45%, respectively. Although FCS similarity measure [1, 19, 26] with MAP of 11.48% performs better than corpus-based similarity measures of [1], it is still slightly lower than the integration of all corpus-based measures with MAP of 12.53%.

Average precision-recall curves for corpus-based measures with fractional set and full set on TRECVID 2005 test set are shown in Figs. 4 and 5, respectively. Average precision of each query using corpus-based methods for both fractional and full sets is also shown in Figs. 6 and 7.

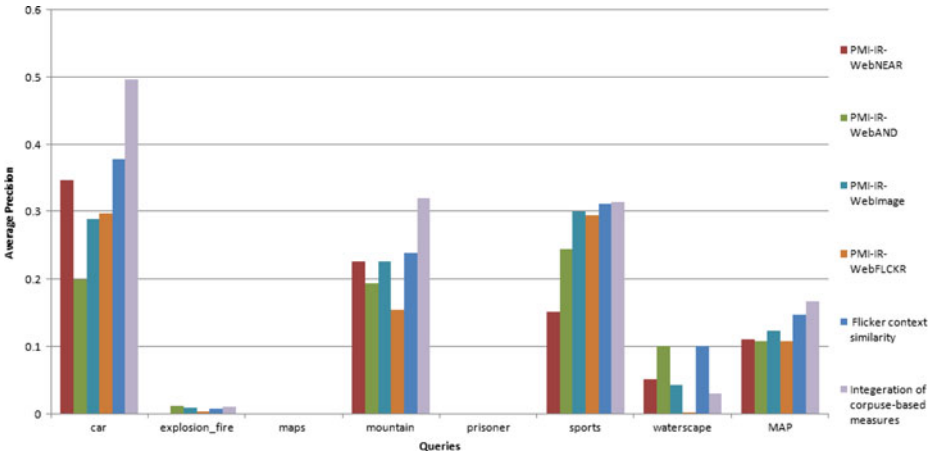

Fig. 6 Average Precision results for each query using corpus-based methods with fractional set

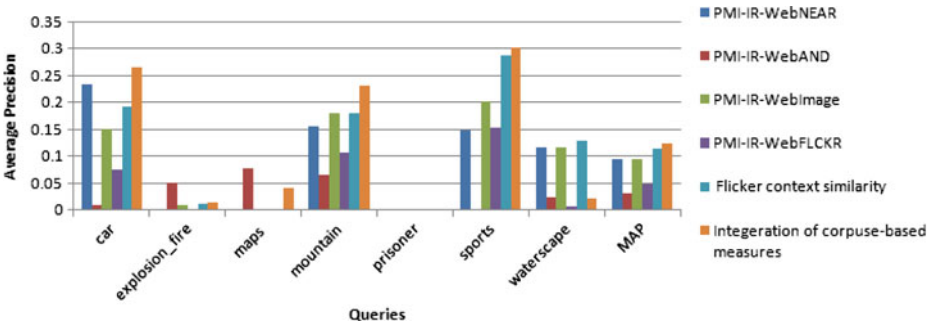

Fig. 7 Average Precision results for each query using corpus-based methods with full set

6 Conclusion

For searching desirable video, users easily express their needs by a textual description in natural language using high-level concepts. Nevertheless, there is a mismatch between the low-level interpretation of video frames and the way users express their information needs. This issue leads to the problem named semantic gap. Moreover, video needs to be manually annotated in order to support semantic query. However, annotating video is a very tedious and challenging task. In this paper, semantic video retrieval model is proposed to find new concepts without availability of annotations. One major contribution of this study is to evaluate various semantic similarity measures against the integration of them in concept based video retrieval. This study showed that the integration of knowledge-based and corpus-based measures outperformed individual ones.

Possible future work includes exploring more reliable and powerful semantic similarity measure and taking into account string similarity measure which can be beneficial in such cases [17]. Retrieving video shots for queries which are expressed in terms of a group of semantic concepts (sentence) rather than one semantic concept can be regarded as another future direction in semantic video retrieval.

References

1. Aytar Y, Shah M, Luo J (2008) Utilizing semantic word similarity measures for video retrieval. In: IEEE conference on computer vision and pattern recognition. Anchorage, Alaska
2. Bach J, Fuller C, Gupta A, Hampapur A, Horowitz B, Humphrey R, Jain R, Shu C, Sethi I, Jain R (1996) Virage image search engine: an open framework for image management. In: Storage and retrieval for still image and video databases IV, vol 2670. SPIE, San Jose, CA, USA, p 87, 76
3. Blair DC (1979) Information retrieval, 2nd edn. J Am Soc Inf Sci 30(6):374–375. doi:10.1002/asi.4630300621
4. Campbell M, Haubold E, Ebadollahi S, Joshi D, Naphade MR, Natsev PA, Seidl J, Smith JR, Scheinberg K, Xie L (2006) IBM research TRECVID-2006 video retrieval system. In: TRECVID workshop

5. Chang SF, Hsu W, Jiang W, Kennedy L, Xu D, Yanagawa A, Zavesky E (2006) Columbia university TRECVID-2006 video search and high-level feature extraction. In: NIST TRECVID workshop
6. Coelho TAS, Calado PP, Souza LV, Ribeiro-Neto B, Muntz R (2004) Image retrieval using multiple evidence ranking. IEEE Trans Knowl Data Eng 16(4):408–417
7. Deerwester S, Dumais ST, Furnas GW, Landauer TK, Harshman R (1990) Indexing by latent semantic analysis. J Am Soc Inf Sci 41(6):391–407
8. Flickner M, Sawhney H, Niblack W, Ashley J, Huang Q, Dom B, Gorkani M, Hafner J, Lee D, Petkovic D, Steele D, Yanker P (1995) Query by image and video content: the QBIC system. Computer 28(9):23–32
9. Frankel C, Swain MJ, Athitsos V (1996) WebSeer: an image search engine for the world wide web. Tech. rep., University of Chicago
10. Frawley W (1992) Linguistic semantics. Routledge
11. Guidelines for the trecvid 2005 evaluation (2005). http://www-nlpir.nist.gov/projects/tv2005/tv2005.html. Accessed 25 Aug 2009
12. Hatzivassiloglou V, Klavans JL, Eskin E (1999) Detecting text similarity over short passages: exploring linguistic feature combinations via machine learning
13. Haubold A, Natsev A (2008) Web-based information content and its application to concept-based video retrieval. In: Proceedings of the 2008 international conference on content-based image and video retrieval. ACM, Niagara Falls, Canada, pp 437–446. doi:10.1145/1386352.1386408
14. Hauptmann A, Christel M, Yan R (2008) Video retrieval based on semantic concepts. Proc. IEEE 96(4):622, 602
15. Hauptmann A, Yan R, Lin W, Christel M, Wactlar H (2007) Can high-level concepts fill the semantic gap in video retrieval? a case study with broadcast news. IEEE Trans Multimedia 9(5):958–966. doi:10.1109/TMM.2007.900150
16. Hopfgartner F (2008) Studying interaction methodologies in video retrieval. Proceedings of the VLDB Endowment 1(2):1604–1608. doi:10.1145/1454159.1454233
17. Islam A, Inkpen D (2008) Semantic text similarity using corpus-based word similarity and string similarity. ACM Trans Knowl Discov Data 2(2):1–25. doi:10.1145/1376815.1376819
18. Jiang Y, Ngo C, Yang J (2007) Towards optimal bag-of-features for object categorization and semantic video retrieval. In: Proceedings of the 6th ACM international conference on image and video retrieval. ACM, Amsterdam, The Netherlands, pp 494–501. doi:10.1145/1282280.1282352
19. Jiang YG, Ngo CW, Chang SF (2009) Semantic context transfer across heterogeneous sources for domain adaptive video search. In: Proceedings of the seventeen ACM international conference on multimedia, MM '09. ACM, New York, USA, pp 155–164. doi:10.1145/1631272.1631296
20. Jones KS, Willett P (eds)(1997) Readings in information retrieval. Morgan Kaufmann Publishers Inc
21. Ko Y, Park J, Seo J (2004) Improving text categorization using the importance of sentences. Inf Process Manag 40(1):65–79
22. Lin C, Hovy E (2003) Automatic evaluation of summaries using n-gram co-occurrence statistics. In: Proceedings of the 2003 conference of the North American chapter of the association for computational linguistics on human language technology, vol 1. Association for Computational Linguistics, Edmonton, Canada, pp 71–78
23. Liu T, Guo J (2005) Text similarity computing based on standard deviation. In: Advances in Intelligent Computing, pp 456–464
24. Markkula M, Sormunen E (1999) End-user searching challenges indexing practices in the digital newspaper photo archive. Inf Retr 1:259–285
25. Mihalcea R, Corley C (2006) Corpus-based and knowledge-based measures of text semantic similarity. In: AAAI'06, pp 775–780
26. Nowak S, Llorente A, Motta E, Rüger S (2010) The effect of semantic relatedness measures on multi-label classification evaluation. In: Proceedings of the ACM international conference on image and video retrieval, CIVR '10. ACM, New York, USA, pp 303–310
27. Ogawa Y, Morita T, Kobayashi K (1991) A fuzzy document retrieval system using the keyword connection matrix and a learning method. Fuzzy Sets Syst 39(2):163–179
28. Park E, Ra D, Jang M (2005) Techniques for improving web retrieval effectiveness. Inf Process Manag 41(5):1207–1223. doi:10.1016/j.ipm.2004.08.002
29. Pedersen T, Patwardhan S (2004) Wordnet:similarity—measuring the relatedness of concepts, pp 1024–1025
30. Radev DR (2004) Lexrank: graph-based lexical centrality as salience in text summarization. J Artif Intell Res 22:2004

31. Robertson S, Jones S (1976) Relevance weighting of search terms. J Am Soc Inf Sci 27(3):146, 129
32. Rodden K, Basalaj W, Sinclair D, Wood K (2001) Does organisation by similarity assist image browsing? In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, Seattle, Washington, United States, pp 190–197. doi:10.1145/365024.365097
33. Salton G, Fox EA, Wu H (1983) Extended boolean information retrieval. Commun ACM 26(11):1022–1036. doi:10.1145/182.358466
34. Salton G, Lesk ME (1968) Computer evaluation of indexing and text processing. J ACM 15(1):8–36. doi:10.1145/321439.321441
35. Salton G, Wong A, Yang CS (1975) A vector space model for automatic indexing. Commun ACM 18(11):613–620. doi:10.1145/361219.361220
36. Smeaton AF (2007) Techniques used and open challenges to the analysis, indexing and retrieval of digital video. Inf Syst 32(4):545–559
37. Smith JR, Chang S (1997) Visually searching the web for content. IEEE Multimed 4(3):12–20
38. Snoek CGM, Huurnink B, Hollink L, Rijke MD, Schreiber G, Worring M (2007) Adding semantics to detectors for video retrieval. IEEE Trans Multimedia 9. doi:10.1.1.62.2860
39. Turtle H, Croft WB (1991) Evaluation of an inference network-based retrieval model. ACM Trans Inf Sys 9:187–222
40. Wang D, Li X, Li J, Zhang B (2007) The importance of query-concept-mapping for automatic video retrieval. In: Proceedings of the 15th international conference on Multimedia. ACM, Augsburg, Germany, pp 285–288. doi:10.1145/1291233.1291293
41. Yanagawa A, Chang SF, Kennedy L, Hsu W (2007) Columbia university's baseline detectors for 374 lscom semantic visual concepts. Columbia university ADVENT technical report

Sara Memar is a M.Sc. student of computer science in University Putra Malaysia, since 2009. Her master research topic is concept-based video retrieval. She holds bachelor degree in Software Engineering from Karaj Azad University in 2008. Her research interests are information retrieval, concept based video retrieval, record linkage, data mining, and bioinformatics.

Lilly Suriani Affendey received her Bachelor of Computer Science from University of Agriculture, Malaysia in 1991 and M.Sc in Computing from the University of Bradford, UK in 1994. In 2007 she recived her PhD from University Putra Malaysia. Her research interest is in Multimedia Databases and Video Retrieval. She is currently a senior lecturer in University Putra Malaysia.

Norwati Mustapha is a Senior Lecturer at the Faculty of Computer Science and Information Technology, Universiti Putra Malaysia (UPM) and head of department of Computer Science since 2005. She received her Bachelor of Computer Science (1991) from UPM and Master of Science (Information System) (1995) from University of Leeds, and her PhD in Artificial Intelligence from UPM (2005). Her areas of specialization are data mining, web mining, text mining, and video mining. She is a member of the Intelligent Computing Group at the faculty.

Shyamala C. Doraisamy is a Senior Lecturer at the Faculty of Computer Science and Information Technology, University Putra Malaysia. She obtained her PhD in the area of Music Information Retrieval from Imperial College London in 2004. Her area of interest includes information retrieval, audio mining and multimedia audio technologies. She currently heads the Multimedia Information Retrieval Research Group at UPM.

Mohammadreza Ektefa received the bachelor degree in Software Engineering from Central Tehran Branch of Islamic Azad University in 2004. He is a M.Sc. student of Computer Science in University Putra Malaysia, since 2009. His master research topic is Record Linkage. His research interests include video retrieval, record linkage, data mining, bioinformatics, and text analysis.