A Survey on Optimizing Image, Video, and Audio Query Retrieval in Multimedia Databases

Ruziana Mohamad Rasli, Su-Cheng Haw, & Roznim Mohamad Rasli

Manuscript
Received:
24,Aug., 2011
Revised:
2, Feb., 2011
Accepted:
11,May, 2012
Published:
15,Jul., 2012

Keywords
Query optimization techniques, Multimedia Databases, Query Processing, Keyword based retrieval, Semantic based retrieval and Content based retrieval.

***Abstract*— With the advanced growth of the Internet, people are expecting systems to support variety of medias. This has therefore leads to new challenges such as the requirements to have new storage, retrieval and presentation of data. In addition, as the sizes continue to grow and the data become more and more complex, it is increasingly crucial and challenging to manage and analyze large amount of multimedia datasets effectively. In this paper, we suggest a query optimization classification scheme and survey the existing optimization techniques in multimedia databases. We also analyze how each approach works, the advantages and disadvantages of each approach focusing especially on video and audio data. The finding will reveal the best technique and serve as the guideline for young researchers.**

1. Introduction

Advances in computing technologies have created a demand for multimedia databases to support large collections of texts, images, videos, audios and even animations. In addition, multimedia objects become so famous because of the increasing popularity of the WWW [1]-[3]. Everyday thousands of multimedia data been uploaded in the Internet. Not only had that advance in mobile devices also made this topic a hot issue [4]. Access is no longer limited to keyword but it can be done on the content (attributes, features and concepts) and semantic of the multimedia object itself. In order to handle this large amount of multimedia data, query optimizer is usually used to speed up the retrieval process.

Query processing in multimedia database is different from query processing in traditional database system in the sense that the multimedia content is more precise and accurate. For instance, in image, there are three important representations: visual features, structural layout, and semantics of the objects. As such, image data is semantically richer than the traditional data formats since it contains both semantics and visual meanings to the users [5]. As such, current traditional database do not sufficiently support searching on content. In another words, it does handle the data appropriately. A new system known as multimedia databases is needed to integrate the management of structure and content of such data.

This paper explores the current state-of-the-art of query optimization in multimedia database. Basically there are three major classifications of retrieval techniques for multimedia databases that will be introduced in this paper.

2. Multimedia Databases

Multimedia databases are widely used in many applications, ranging from medical imaging to music recommendation systems [2]. It is a collection of varieties of multimedia data such as text, video, audio, images, graphics and animation [6]. Recently, application for multimedia data especially on audio and video has emerged greatly due to the emerging of high-speed communication system and the upgraded technologies [7]. The major goal in multimedia data is the need retrieve multimedia content to the user [6]. This imposes heavy burden of data manipulation, including integrating, searching, selecting, interpreting and skimming the information. Multimedia Database Management System (MMDBMS) has 3 data models which are object oriented data model, object relational data model and spatio-temporal data [8].

Multimedia data such as video have several information elements such as [9]:

a) Metadata or keyword – the actual video frame stream, including its encoding scheme and frame rate.
b) Content based media data – the information about the characteristics of video content such as visual feature, scene structure and spatio-temporal feature.
c) Semantic data – the text annotation relevant to the content of video, obtaining by manual or automatic understanding.

Although multimedia database has additional features in it, multimedia database also need traditional database management system (DBMS) capabilities [6].

Ruziana Mohamad Rasli is from School of Multimedia Technology and Communication, College of Arts and Sciences, Universiti Utara Malaysia (ruziana@uum.edu.my).

Su-Cheng Haw is from Faculty of Information Technology, Multimedia University, Malaysua (su-cheng@mmu.edu.my).

Roznim Mohamad Rasli is from Faculty of Art, Computing, and Creative Industry , Sultan Idris Education University (roznim@fskik.upsi.edu.my)

A. Requirement Features for Multimedia Databases

With the growing of large databases in multimedia contents, it is really important to have the ability to retrieve and browse the content from such databases. A major problem of multimedia data is the nature of the data itself. The formats of multimedia data, unlike text, have no direct access between the objects and the meaning. The major disadvantage of most existing multimedia database is they do not have a formal framework to adequately provide full-fledge multimedia operations [10]. Moreover, as mentioned earlier, although multimedia databases provide a lot of advantages, such large collection of information has brought problems in storage and management.

Query processing in multimedia databases is different from query processing in traditional database systems. This is due to the reason that multimedia data is content-dependent metadata and there exist semantic gap in the data itself. As the result, this means that, it needs a method which can structurally explain the content of video data for semantic retrieval. According to [6], there are two types of multimedia data which is discrete and continuous. An example of discrete multimedia data is image which is static whilst continuous multimedia data can be audio and video which is temporal. Many advanced applications have revealed the incapability of the current database system to represent and process the objects. This research is motivated from the absence of significance approach to optimize on the storage, retrieval and presentation in multimedia databases. Though there have been rich literatures available in recent years, however, there had been relatively little work on multimedia databases. There are several differences in the processing of multimedia data compared to traditional data which can be divided into five as follows [11]:

a) Format of multimedia data.
b) Presentation of the output results.
c) Size of multimedia data.
d) Temporal characteristics of multimedia data.
e) Automatic feature extraction and indexing using advanced tools.

Another research from [7] stated that there are three challenges arise from multimedia data which are size, time and semantic nature of multimedia data itself. These three challenges could be solved by data segmentation to achieve high resource utilization and increase concurrency, and parallelism respectively. Therefore, there is a need of query optimization to speed up the process of getting multimedia data.

3. Query Optimization Techniques

Query optimization in multimedia databases had been continuously increasing and become active issues in both academic and commercial field. It had been studied in many fields and as data set sizes and formats continue to grow, it is very important and challenging to efficiently execute data analysis applications that process these data sets [12]. Different multimedia object will use different types of techniques. It involves query segmentation which takes a query and simplified restructured query expressed in relational algorithm after modification due to views, security enforcements and semantic integrity control [13]. Not only that, multimedia database management system (DBMS) such as Oracle 10g and SQL Server had also evolved with the ability to store and retrieve large binary object such as Binary Large Object (BLOB) and Non Character Large Object (NCLOB).

Figure 1 shows the classification of several optimization techniques which are widely used in multimedia databases. In this paper, firstly, we discuss on several related optimization techniques that are used in multimedia objects. Secondly, we summarize the advantages and disadvantages of each technique.

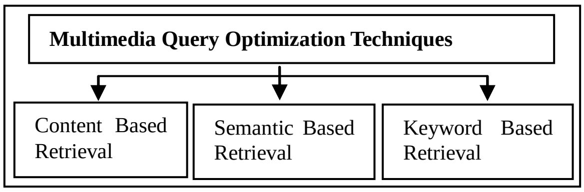

Fig. 1: Query Optimization Techniques

A. Content Based Retrieval

Content based retrieval was first introduced in the early 1980 as a new tool and is the most popular group of optimization technique [14]. CBR uses visual content of an image as features to represent and index image to be searched from large scale image databases. It is the main motivation behind recent research in multimedia databases [15]. CBR formulate queries to access semantic information contained in multimedia data such as video data and motion information [9]. The process implements the usage of color, texture, shape, and various properties gathered from the raw data to analyze the raw data in perspective of the geometric patterns, meaningful entities , and scenes with semantic. It is categorized as a low level features.

A research from [16] focusing on image data uses content based image retrieval (CBIR) to extracts parts of the image and transforms it into low level features on that particular data objects. Another research by [17] based on content based video retrieval focuses on the low level features such as color, motion, and texture index of the video. [13] had created a prototype for multimedia database with a retrieval interface focusing on the dependencies of context and content information for effective retrieval [15].

According to [17], there are two different elements in content based video similarity search which are video clip retrieval and video subsequence identification. Video clip retrieval focuses on getting similar clips from a large collection of videos which have been segmented into shots of similar length size at content boundaries. Whilst video subsequence identification focuses at searching for existent of any part of a long database video that shares similar content to a query. However, there are several disadvantages of CBR. One of the disadvantage is CBR is not suitable for image and video due to the semantic gap in its object. In order to do this, human will tends to use high level semantics to evaluate the retrieval results.

Similarity search is one of the important implementation under the CBR classification. According to [18], similarity queries can be classified into two categories which are whole matching (sequences to be compared have the same length) and sequence matching (involves the query sequence that is smaller and the comparison is done in the large sequence that best match the query sequence). Similarity search is a technique that focuses on the inner part of multimedia data. It is an operation that finds changing pattern of sequences or subsequences that are similar to the given query sequence. [19] developed a research focusing on the effectiveness of similarity search in multimedia databases using multiple representation for video and audio. The framework is able to integrate multiple representations such as audio and video featured into query processing [19]. Since this object usually consists of different features so similarity search is a suitable method to be used. The research focused on video will implement intelligent similarity search techniques whereby the objects will automatically choose the best representations. This type of representation will have different weighting function such as weighting function based on support, specific quality measures, local neighborhood and entropy.

The work at reference [20] developed a content based retrieval that focuses on image and uses the method based on contour let transform. This system uses visual contents of an image as features to represent and index image to be searched from a large scale image databases. Not only that, this research combined two types of primitives, color and texture. The advantages of using content based image retrieval compared to other methods is that CBR provides high accuracy in retrieving the result and this method normally low in computational complexity. This research implements feature extraction based on contourlet transform and RGB color space. This research also consists of two types of filter bank and three important steps.

a) Step 1 – Pre-processing

In order to extract related image features, Gaussian filter is used as the first steps. Gaussian filter is used to blur images and remove detail and noise. It is similar to the mean filter, but it uses a different kernel that represents the shape of a Gaussian or a bell-shaped hump.

b) Step 2 – Feature extraction

Color and texture features are used to extract the features in the image. Texture orientations are captured by using Contourlet decomposition with 4 levels (0,2,3 and 4) LP decomposition. At each level, the number of directional sub-bands is 3, 4, 8, and 16 accordingly. 'Maxflat' filter are used for LP decomposition and 'dmaxflat' directional subband decomposition. It results in a 32 dimensional feature vector. Standard deviation and mean vector is calculated and computed on each directional subband of the CT decomposed.

For RGB color space, first, the color is decomposed into three channels in order to calculate the mean and the standard deviation of its histogram descriptors. By combining these methods, this approach is considered as a new approach compared to Gabor based methods.

c) Step 3 – Classification

In order to calculate the similarity measures between the queried images and the images in the database, Euclidean distance is chosen to measure the degree of similarity between compared features. Mahalanobis distance also used as a similarity measures in this research

An experiment was made based on various kind of image in Wang database. In this database, there are 10 themes with each themes consists of 100 images. Query image is chosen randomly. From this experiment, contourlet transform is more robust compared to Gabor and wavelet transform. The proposed method is more efficient and gives low error rate. There are three advantages of using contourlet transform. Listed below are all the advantages:

a) It provides multiscale and multidirectional representation of an image.
b) It is usefull in detecting fine details in any orientation at various scale levels, giving it a good and effective image analysis.
c) It offers fast and structured decomposition and smoothes signal resembling to other images.

B. *Semantic Based Retrieval*

Semantic based search is defined as a type of searching technique that compares the original multimedia data to a prototypical category such as 'vehicle', clothes and others [21]. Compared to content based retrieval, semantic based retrieval is categorized as a high level features that implements user's perception According to [22] semantic query uses knowledge about the domain of relations, nature of data, and constraints related to database elements. [23] stated the efficiency of the retrieval system is based on the ability to comprehend high level features of the semantic features. The process will try to exploits and uses any available knowledge about the data. Element extracted from different modalities of a video, such as visual information, auditory information, and text in the video frames are generated to model the semantic of the video [24].

There are four main issues involved in semantic query [25]. The first issue focuses on the query and schema which should be dynamically used to select the relevant semantics for optimization without any addition search of semantic rule base. Second, a suitable mechanism should be available to combine selected semantic with the query. Third, cost analyzer is needed to evaluate the cost of equivalent queries and rank them accordingly. Lastly, a heuristic guide is needed to show the whole process in a meaningful way so that the process can be easily understood by the users.

In order to handle rich, temporal and spatial requirements of multimedia data, a visualized semantic model is used to increase the content of the information so that the cognitive load in the users is reduced [9]. Semantic query optimization is based on the semantic equivalence rather than the syntactic equivalence between different queries [25]. Query and data centric method can be incorporated to optimize the acquisition process.

However, there are several disadvantages of semantic content such as it is difficult to extract automatically [7]. For image retrieval, the process is based on an assumption that image metadata is pre-extracted and stored semantics (textual data) in the database. This approach must consider both the semantics of the query and images to be in the same category so that the query processor can defined its similarity [26]. [26] added that there are five steps involves in extracting the semantic data such as:

a) The system will identifies image regions. This is a type of low level concept.
b) The system recommends semantic candidate components by using the image component catalog.
c) User chooses the result and may select the objects to be stored in the image component catalog as “representative samples” for future use.
d) The system identifies spatial relations of the objects based on the comparison of four objects’ outer points or weight scenes. The spatial relationship are stored as two vectors, one for horizontal relationship and one for vertical relationship, similar to the 2D string representation. The ranking is calculated based on the similarity values and weight of candidate component in various query processing modules.

[23] uses Semantic Query Interpreter (SQI) as an automatic query interpretation in their research. It interprets user’s query using both lexical and semantic features using open source knowledge model, Wordnet and ConceptNet. There are four major component in SQI; core lexical analysis, common sense reasoning, candidate concept selection and retrieval and ranking of result. The steps involved in this system are as follows:

a) User submit query (single word) to the system
b) The system then analyze user query and find relevant and irrelevant data based on user’s specification.
c) Query is transferred to core lexical analyzer which converts user’s query into finite set of token by using tokenization technique. This technique transforms keyword in a set of words.
d) The word may exist in morphological forms (plural and gerund form of in past or future suffixes.
e) The word is then converted to different morphological form in a process called lemmatization. This process eliminates unusual or stops words. Nouns are represented as an entities verb which represents events and adjectives.
f) Then, selected lexeme is then transferred to WordNet knowledge base which expands the query by attaching synonyms (synsets).
g) Using the two knowledge bases, ConceptNet will reduce the semantic gap and Wordnet will reduce the vocabulary gap.
h) In this research, common sense reasoning are extracted by using Knowledge Lines (K-Lines) from ConceptNet. ConceptNet have eight different kinds of K-Line categories and 20 relationships.
i) Vector Space Model (VSM) is used to retrieve images on the basis of frequency of the selected concepts tagged with the image.

An experiment was carried out using LabelMe test collection taken from MIT Computer Science and Artificial Intelligence Lab (CSAIL) which consists of 8983 digital images, 56943 annotated images and 25040 non annotated images. Keyword based and sentence based queries are carried out and the results shows promising but fluctuates in some complex query. As a conclusion, lexical and semantic knowledge need to be combined as it is important for the system to be efficient.

C. *Keyword Based Retrieval*

Keyword based retrieval is considered as a traditional method to retrieve data using textual description (metadata) [27]. Metadata is defined as structured information describing characteristics that assist users to identify digital content itself [28]. Metadata is the data or semantic information to classify the content, quality, condition and other characteristics of the data [2]. Structured and organized metadata is needed to deal with large amount of digital content that are available on the Internet nowadays [4]. It changes the way of content search from using normal string to a conceptual level where users search for semantic contents of the data. [29]. The richer the metadata representation, the more valuable and relevant it is to both party which can be useful for several important tasks such as searching the data, summarizing it into simpler formats and relate if with other types of sources [29]. There are several metadata schemes that are famous nowadays. For this research, three famous schemes will be discussed which are MPEG-7, Dublin Core, and IEEE LOM.

MPEG-7 is different from other standard metadata in the sense that it can provide two types of schema which is divided into low level descriptions and high level descriptions [30]. Low level description is defined as colour, texture, and the shape of the multimedia data whilst the high level description is the structural and semantic descriptions. MPEG-7 major focuses are video and images and the technology enables the CBIR (Content based Information Retrieval). MPEG-7 enables the metadata to be used in different platform and applications. There are four major building blocks in MPEG-7 [31]. Table 1 list the name and its descriptions.

The second scheme is called Dublin Core schemes which conceived author generated descriptions of Web sources. The name “Dublin” came from an invitational workshop in Dublin, Ohio, and the “Core” is because of its wide and generic elements, usable for describing a wide range of recourses [32].

TABLE 1
BUILDING BLOCK FOR MPEG-7

Building Blocks	Description
1) Descriptor	Defines the syntax and semantics of the feature representations.
2) Description Scheme	The structure and semantics of the relationship between its components.
3) Description Definition Language	A language that allows the creation of new Description Schema and the Descriptors. Allow the extension and modification of existing Description Schema.
4) Systems Tools	Tools to support multiplexing of descriptions, synchronization of descriptions with content, delivery mechanism, and coded representation for efficient storage and transmission and the management and protection of intellectual property in MPEG-7 descriptions.

The main objective of Dublin Core Education Working Group which was formed in 1999 is to discuss and develop a proposal to be used in Dublin Core metadata. Its aim is to define a standard that will encourage quality resource description and encourage interoperability between tools for resource discovery [33]. Dublin Core is classified into 15 elements that will be used in resource description as shown in Table 2.

TABLE 2
ELEMENTS IN THE DUBLIN CORE METADATA.

DUBLIN CORE ELEMENTS	DESCRIPTION
1) Creator	An entity responsible for making the resource.
2) Identifier	An unambiguous context reference to the resource.
3) Title	A name given to the resource.
4) Date	A date given to the resource.
5) Type	The nature or genre of the resource.
6) Subject	The topic of the resource.
7) Description	An account of the resource.
8) Rights	Information about rights held in/over the resource.
9) Publisher	Entity responsible for making the resource available.
10) Coverage	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant
11) Language	A language of the resource.
12) Format	The physical file format.
13) Relation	A related resource.
14) Contributor	Entity responsible for making resource contributions.
15) Source	A related resource from which the described resource is derived.

The third scheme is called IEEE standards that conform, integrate and refer open standards with the existing work in related areas [29]. Learning Object Metadata (LOM) standards specify a conceptual data schema that defines the structure of a metadata instance [25]. IEEE LOM is divided into nine elements which deemed to be timely and important [10]. In addition, its specification and vocabularies were determined through discussion from the standpoints of both users and resource developers. The IEEE LOM uses hierarchical types of metadata description which is useful, and easy to be implemented in many levels of elements [34]. According to [10] LOM are classified into nine descriptors of learning resources, each of which is relatively independent and characterizes the resource from a separate aspect as shown in Table 3 below.

TABLE 3
ELEMENTS IN IEEE LOM

No.	Elements	Vocabulary
1.	General	General information about learning resources.
2.	Lifecycle	The history, current state and the contributors.
3.	Meta-metadata	Information about the metadata instance itself.
4.	Technical	The technical requirements and characteristics.
5.	Educational	The educational and pedagogic characteristics.
6.	Rights	Conditional use of intellectual property rights.
7.	Relations	The relationship between learning resources.
8.	Annotation	Comments on the resources and their creators
9.	Classification	Resource in relation to a classification system.

[8] created a generalized multimedia database data model that uses MPEG-7 data model. The system is called G3M which analyze MPEG-7 Multimedia Description Scheme (MDS) to construct user's preferred database schemes. G3M uses existing data modeling method which are multilevel abstraction, metadata interpretation and salient object based approaches. In this research, G3M can support majority of medias. Not only that, it have flexible segmentation, low level features, spatio-temporal relationships and high level semantic. G3M have added extensibility characteristics listed below:

a) New media can be easily added into the system.
b) All MPEG-7 based interfaces can be easily connected with G3M.
c) G3M supports user defined attributes, features, objects, events, concepts and relations.

In this system MPEG-7 descriptors are mapped to corresponding database data types, objects, tables and varieties of relationship which are expressed by relational keys and object references. User can choose their required interests and privileges. This system is implemented in Object Relational Database Management System (ORDBMS). As a conclusion, G3M supports user defined safety management of the database and the data.

4. Summary and Discussion

Table 4 summarizes the advantages and disadvantages of the optimization techniques mentioned above.

TABLE 4
COMPARISON ON EACH TECHNIQUE

Type of Index	Advantages	Disadvantages
1. CBR	1) Provides ease of query formulation and interpretation. 2) This technique is suitable and flexible for formulating queries.	1) Provides data dependency between multimedia data. 2) User dependency 3) Performance is lacking when queries are executed on large datasets.
2. Semantic Based Retrieval	1) Large amount of computation time can be saved compared to the index structures because of the simpler and overlap-free characteristics 2) The technique can be easily extended to all transforms available.	1) It cannot support queries on generalized concepts. 2) The retrieval is not precise enough for the process.
3. Metadata	1) Rich metadata are very effective in assisting users to navigate and find desired content items. 2) High quality metadata is important for reliable and effective Web applications.	1) The more data become available, the harder and difficult to identify and extract metadata. 2) The absence of meaningful metadata due to lack of users' attention in providing the information.

Amongst these three techniques, CBR is more suitable for temporal data for example video. Several researches had been carried out on video where the content of the video will be segmented first and compared using similarity search. This technique normally involves more than one technique to extract the content of multimedia data. However, this technique provides dependency between multimedia data in such, in order to retrieve the query output, the process will involves the extraction of multimedia segments.

Semantic based retrieval is more focus on image data. It involved algorithm to analyze the structure of the image. Using low and high feature extraction, the shape, color, and texture of the data can be analyzed. Since this technique involved the extraction of the inner part of the data, queries cannot support generalized concepts and the output retrieval is not precise for the process.

Keyword based retrieval is suitable for both of the data. However, there is a need of user interaction in defining the metadata. In order to have an effective application, high quality metadata is very important but the more data are available, the harder it is to identify and find the suitable metadata. The major problem in using keyword based retrieval is that it is hard to represent multimedia data. Each multimedia data will need to be defined by human and the extraction of the content will be time consuming. This is because, users or human need to classify and defines the data one by one. For example, if a picture of a red flower is defined by User A is a rose; User B will define it as a bougainvillea. [35] stated that most of multimedia retrieval systems are based on text motivating of keyword which gives several disadvantages. These advantages are irrelevance results and unable to meet user's requirements.

However, the main issue in the past several years regarding multimedia data is that how to represent multimedia data according to its semantics. Using the three methods on its own is not enough in representing the true meaning of multimedia data. Therefore, several researches had been done that integrates and combines methods listed above to become one.

An example of research that combines these methods is from [36]. [36] developed a content based image retrieval called I-Browse. I-Browse integrated both iconic and semantic content for histological image analysis. It combines both low level and high level features. Medical image is the major scope in this research because it plays an important part in patient diagram, therapy, references and training. In traditional medical image database, majority of indexing and retrieval process are based on the metadata of patients' information and types of examination. There exists such a different between the metadata and the higher level of abstraction in the images. Therefore, semantic is needed to give precise result based on the images.

I-Browse enables users to search over the image though the combination of icons and semantic contents. It also generated textual annotations automatically for input images. The process of I-Browse consists of six major objectives listed below:

a) To find similar image from the archive by visual similarity in image example.
b) To interpret visual properties as histological featured in similar ways to doctor.
c) To generate textual annotations for unknown images.
d) To find similar images from the archive by image example in terms of semantic similarity
e) To retrieve images using natural language queries.
f) To act as a delivery item to channel the image into the correct place.

From a series of image analysis operations, semantic units (salient histological feature) are automatically extracted. Two sets of histological features in tract images are defined as semantic label. The medical images that were used are more than 1500 digitized images with x50 magnification. By making this analysis, two histological meaningful interpretations of images are created; course feature level and fine feature level. In order to detect visual feature building blocks, three detectors are used which are course detector, semi-fine detector and fine detector. Using these detectors, semantic analyzer is used to identify errors in the labels and correct it. Semantic analyzer needs to fulfill two aims in this research:

a) Improve accuracy of recognition results using high level histological and contextual knowledge
b) Analyze semantic contents of the whole image and record semantic content that generate textual annotation for the image.

Three similarity measurements are used to compare most frequent semantic labels, local neighbor patterns of semantic labels and semantic labels frequency distribution.

a) Most frequent semantic label (MFSL) – based on 15 coarse histological regions of the images.
b) Neighborhood Similarity (NS) – uses a matrix to record the co-occurrence frequency of 63 histological labels of the 8 nearest neighbors against the sub-images.
c) Semantic Label Frequency Distribution Similarity (SLDS) – counts the frequency directly of the 63 fine histological labels if the image.

An experiment was done using these three measurements. Result shows that NS and SLDS performed better than MFSL with 78% and 80%. MFSL performed worse than the other because it uses less semantic information in the measurement. Computer that uses Intel Pentium II 450 provides computational time of 6ms, 0.23s and 4.3ms for MFSL, NS and SLDS. Here, SLDS proven to be the best and faster method to give the result compared to the other two. As a conclusion, I-Browse is a domain independent and quite useful to generate semantic features with the help of keyword (annotation) features.

The result of combining these methods brings a lot of advantages to the user. Majority of researches dated back to 2000, combines several method together to get a hybrid approaches which proven to have high precision and recall with low computational complexity.

5. Conclusion

In this paper, we have classified the query optimization techniques in multimedia database. We have also described how each technique works, the advantages and limitations of each technique. Majority of researches nowadays implemented more than one method in their system to deal with the semantic gap in multimedia data. As a conclusion, to have a perfect application, it is advisable to combine these three techniques into hybrid application so that the limitation on each technique can be fulfilled by the other technique.

References

[1] A. P. DeVries,. M. Windhouwer, P. M. G. Apers & M. Kersten, "Information Access in Multimedia Databases based on Feature Models," (2000) *In New Generation Computing*, vol. 18, no. 4, pp. 323-339.

[2] M. Wichterich, I. Assent, P. Kranen, & T. Seidl, "Efficient EMD-based similarity search in multimedia databases via flexible dimensionality reduction," (2008) *In SIGMOD Conference*, pp. 199-212.

[3] S. Kiranyaz & M. Gabbouj, "Hierarchical Cellular Tree: An Efficient Indexing Scheme for Content-Based Retrieval on Multimedia Databases," (2007) *In IEEE Transactions on Multimedia*, Vol 9, No. 1.

[4] M. Svensson, "Contextual Metadata in Practice," (2009) *In the 2009 First International Conference on Advances in Multimedia*.

[5] W. Li, K. S.Candan, K.Hirata & Y. Hara, "Supporting efficient multimedia database exploration," (2001) *In 2001 VLDB Journal*.

[6] O. Kalipsiz, "Multimedia Database," (2000) *In the IEEE Transaction*.

[7] L. Dunckley. Multimedia Databases, An Object-Relational Approach (2003), Addison-Wesley.

[8] Q. Zuo & Z. Cao. G3M, "A Gemeralized Multimedia Data Model Based on MPEG-7," (2008) *In the International Conference on Multimedia and Ubiquitous Engineering*.

[9] S. S. M. Chan, Q. Li, Y. Wu & Z. Yuetin, "Accomodating Hybrid Retrieval in a Comprehensive Video Database Management System," (2002) *In the IEEE Transaction on Multimedia*, Vol. 4. No. 2.

[10] M. Hee, Y. Y, Ik, & K. C. Kim, "Hybrid video system supporting content-based retrieval," (1999) *In the Third Conference on Computational Intelligence and Multimedia Application*, ICCIMA 99.

[11] S. Marcus & V. S Subrahmanian, "Foundations of Multimedia Database Systems," (1996) *In ACM* vol. 43, no. 3, pp. 474-523.

[12] H. Andrade, T. Kurc, A. Sussman, & J. Saltz, "Optimizing the Execution of Multiple Data Analysis Queries on Parallel and Distributed Environments," (2004) *In the IEEE Transactions On Parallel and Distributed Systems*, Vol. 15, No. 6.

[13] A.B.M. R. I. Sadat & P. Lecca, "On the performance in simulation of parallel databases : an overview on the most recent techniques for query optimization," (2009) *In the International Workshop on High Performance Computational Systems Biology*.

[14] G. Quellec, M. Lamard, G. Cazuguel, C. Roux & B. Cochener, "Case Retrieval in Medical Database by Fusing Heterogeneous Information," (2011) *In the IEEE Transaction on Medical Imaging*, Vol 30, No 1.

[15] C. Ribeiro, G. David, & C. Calistru, "A Multimedia Database Workbench for Content and Context Retrieval," (2004) *In IEEE 6th Workshop on Multimedia Signal Processing*.

[16] S. H. Kwok & J. L. Zhou, Content Based Object Organization for Efficient Image Retrieval in Image Databases (2006).

[17] W. Zhou , S.Dao, & C. C. J. Kao, "Online knowledge and Rule Based Video Classification System for Video Indexing and Dissemination," (2000) *In the 2002 Information System*, 559-586

[18] R. Aggrawal, C. Faloutsos & A. Swarmi, Efficient Similarity Search in Sequence Databases (1994).

[19] H. Kriegel, P. Kroger, P. Kunath & A. Pryakhin, Effective Similarity Search in Multimedia Databases using Multiple Representations.

[20] R. Romdhane, H. Mahersia & K. Hamrouni, A Novel Content Image Retrieval Method Based on Contourlet (2008).

[21] J. Yoon & N. Jayant, Relevance Feedback for Semantic Based Image Retrieval (2001).

[22] J. Shao, "Challenges and Techniques for Effective and Efficient Similarity Search in Large Video Databases," (2008) *In the PVLDB, August 23-28 2008, Auckland, New Zealand*.

[23] N. Aslam, Irfanullah, J. Loo, M. Looms & Roohullah, "A Semantic Query Interpreter Framework by Using Knowledge Bases for Image Search and Retrieval," (2011) *In the 2011 IEEE Transaction*.

[24] C. Choudhary, T. Liu, & C. Huang, "Semantic Retrieval of Instructional Videos," (2007) *In the Ninth IEEE International Symposium on Multimedia – Workshops*.

[25] H. T. Shen, B. C.Ooi, & X. Zhou, "Towards Effective Indexing for Very Large Video Sequence Databases," (2005). *In the SIGMOD, June 14-16 2005, Baltimore, Maryland USA*.

[26] W. Li, K. S. Candan, K.Hirata & Y. Hara, A Hybrid Approach to Multimedia Database Systems Through Integration of Semantics and Media-based Search (2009).

[27] F. M. Matsubara, T. Hanada, S.Imai, S.Miura, & S. Akatsu, Managing a Media Server Content Directory in Absence of Reliable Metadata (2009).

[28] F. Getahun, S. Atnafu, J. Tekli & R. Chbeir. The Use of Semantic-based Predicates Implications to Improve Horizontal Multimedia Databases Fragmentation (2007). Available at http://www.imdb.com.

[29] A. Steinacker, A. Ghavam, & R. Steinmetz. Metadata standards for Web-based Resouces (2001).

[30] T. Catarci, The Lift Cycle of Multimedia Data (2005).

[31] J.M Martinez., R. Koenen & F. Perceira. MPEG-7 The Generic Multimedia Content Descriptions Standard, Part 1 (2002).

[32] Dublin Core Metadata Initiative. The Dublin Core. A simple content description model for electronic resources, http://dublincore.org/documents/dces.

[33] C. Jenkins & D. Inman. Server-side Automatic Metadata Generation using Qualified Dublin Core and RDF (2001).

[34] X. Xiang, Y. Shi & L. Guo, "A Conformance Test Suite of Localized LOM Model," (2003) *In the Proceedings of the 3rd IEEE International Conference on Advanced Learning Technologies (ICAL 03)*.

[35] H. Qi, L. Zhang & Y. Gao, "Semantic Retrieval System Based on Corn Ontology," (2010) *In the Fifth International Conference on Frontier of Computer Science and Technology*.

[36] H. L. Tang, R. Hanka & H. H. S. Ip, "Histological Image Retrieval Based on Semantic Content Analysis," (2003) *In the IEEE Transaction on Information Technology in Biomedicine*, Vol 7, No 1.

Ruziana Mohamad Rasli received her B.IT (Software Engineering) from Multimedia University and M.sc (Information Technology) from Universiti Utara Malaysia in 2006 and 2008. She is currently a lecturer at School of Multimedia Technology and Communication, College of Arts and Sciences, Universiti Utara Malaysia. Her research interests includes image processing and multimedia database.

Su-Cheng Haw received her B.Sc (Physics) and M.sc (Information Technology) both from Universiti Putra Malaysia. Her doctorate was presented by Multimedia University. She is currently a senior lecturer at Faculty of Information Technology, Multimedia University. Her research interests include database query processing and optimization and data management.

Roznim Mohamad Rasli received her B.IT (Artificial Intelligence) and M.sc (Intelligence System) from Universiti Utara Malaysia in 2004 and 2005. She is currently a lecturer at Faculty of Arts, Computing and Creative Industry, Sultan Idris Education University. Her research interests include artificial intelligence, data mining and information technology.