Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 5.258

IJCSMC, Vol. 5, Issue. 5, May 2016, pg.349 – 353

A REVIEW ON MULTIMEDIA CONTENT SEARCHING AND RETRIEVAL SYSTEM

Pradnya P. Sondwale
Dept. of Computer Science & Engineering, Sipna College of Engineering and Technology, Amravati, India
pradnya.sondwale@gmail.com

Abstract: Multimedia Content Searching and Retrieval System can be defined as the process of finding interesting / useful contents and patterns from different media contents such as text, images, audio and video. This multimedia type of data are not ordinarily accessible by basic queries and associated results but uses some techniques of content searching for gaining useful information called as knowledge in terms of multi-media types of data. This paper is performing the review on Multimedia content searching and retrieval system that extract knowledge and high level multimedia information from multimedia storage system as like our personal computer system. The system performs content / pattern discovery using some techniques and rule of extraction. To extract interesting content from multimedia data different types of techniques are used. Here, we are performing survey for the different media data that having application at various areas and the techniques for searching and retrieval of multimedia data. Analysis of this huge amount of multimedia data to discover useful information and content is a challenging task which has opened the era for researchers in the context of the multimedia content searching and retrieval system. Our system is intended to maximize the usage of the discriminative power of the human researcher and apply the techniques for the multimedia content searching and retrieval.

Keywords: Multi-media Content Searching, Content Retrieval System, Features Extraction.

I. Introduction

Now a days, Multimedia has becoming an inevitable part of any presentation. It has found a variety of applications right from entertainment to education. The evolution of internet has also increased the demand for multimedia content [1]. Multimedia uses multiple forms of information content and their processing e.g. text, audio, video, graphics, animation and interactivity to inform or entertain its user. *Multimedia* also refers to the use of electronic media to store and experience multimedia content. Multimedia is similar to traditional mixed media in fine art, but with a broader scope. The term "rich media" is synonymous for interactive multimedia which is also used today for content searching and retrieval.

As in the today's world there is advancement in multimedia acquisition and storage technology. This leads to tremendous growth in very large and detailed multimedia databases. This analyzes huge amount of multimedia data to discover useful knowledge that is a challenging problem, which has opened the opportunity for research in Multimedia Content Searching and Retrieval System. Multimedia data retrieval can be defined as "the process of finding interesting patterns from the large media data such as text, image, audio and video that are not ordinarily accessible by basic queries and associated results" [2]. As our system is related to the retrieval of data, this multimedia data searching and retrieval is a new research area and it requires background knowledge from both the data searching as well as multimedia processing domain. The typical multimedia data searching and retrieval process consists of several stages and the overall process is inherently interactive and iterative. The main stages of the multimedia data retrieval process are (1) Domain understanding, (2) Data selection, (3) Data pre-processing, (cleaning and transformation) (4) Patterns Evaluation (5) Interpretation and (6) Reporting and Knowledge discovery [4].

There are many traditional, as well as future applications that can get significant benefits from the content based searching and retrieval systems. This include wide variety of application sectors as educational, business, entertainment, and sports. Generally, searching and retrieval of information is increasingly becoming a basic functionality needed for the different multimedia application or service. It can directly serves as providing desired functionality to the user, or it can indirectly allow customization of available functionalities for them.

Furthermore, this searching and retrieval is typically associated with applications, in which the user wants to locate and extract a piece of information from a storage repository. With the context of the new growing applications and also in many real-time and non-real-time application, in which the user wants portions of broadcast information to be selectively presented from many choices available. The real-time applications are those where content indexing information is generated in real-time and they should be associated with it alongside the content. In this paper, the focus is to first develop the primary concepts needed for object-based multimedia search and retrieval, and after that to develop a systems architecture that can effectively use these concepts [5]. These system is customize within context of a real-world application and demonstrate its effectiveness.

The fundamental necessities for a multimedia information retrieval system are searching for a particular media item, and browsing and summarizing a media collection. For the searching of the particular media item, the current systems have significant limitations, such as an inability to understand a wide user vocabulary and the user's satisfaction level. Current systems have not yet had significant impact on society due to an inability to bridge the semantic gap between computers and humans. Currently, the most commonly used test sets include collections involving personal photos, web images and videos, cultural heritage images, news video, and the Corel stock photography collection which is always present in our computer system. We suspect it is popular, because it is widely available and related loosely to real-world usage.

As the content are multimedia, which containing different media components such as text, image, audio and video that are correlated by the context of the application data. The problem in this context is important because, current searching and retrieval solutions are mostly proprietary and mainly associated with textual document retrieval with very little or no support for multimedia documents [3]. Even though some existing solutions claim to employ media information other than text, they do not produce satisfactory results because the multimedia information is only partially exploited. Now according to the increasing technology we have to develop the techniques to integrate features and characteristics across different media types and develop system solutions that facilitate flexible multimedia content searching and retrieval system.

II. Background

Feature extraction is very important step in content based retrieval system. Mostly for the video, the basic visual features of index frame include color and texture. From the long going, research in content based video retrieval today is a lively disciplined, expanding in breadth. Representative features extracted from index frames are stored in feature database and used for object-based video retrieval.

Texture is another important property of index frames. Various texture representations have been investigated in pattern recognition and computer vision. Texture representation methods can be classified into two categories: structural and statistical. The Structural methods including morphological operator and adjacency graph, describe texture by identifying structural primitives and their placement rules. When applied to textures, they tend to be most effective. Statistical methods, including Fourier power spectra, co-occurrence matrices, shift-invariant principal component analysis (SPCA), Tamura feature, Wold decomposition, Markov random field, fractal model, and multi-resolution filtering techniques such as Gabor and Haar wavelet transform, characterize texture by the statistical distribution of the image intensity [10].

Video Retrieval Based on Textual Queries presented an approach that enables search based on the textual information present in the video. Regions of textual information are indented within the frames of the video. Video is then annotated with the textual content present in the images. A learning framework presented an automatic Content-Based Retrieval and Classification of Video Content where construction of a high-level video index is visualized through the synthesis of its set of elemental features. This is done through the medium of support vector machines (SVM) that associate each set of data points in the multidimensional feature space to one of the classes during training. In Content-Based TV Sports Video Retrieve Based on Audio - Visual Features and Text. Here are some algorithms used for feature extraction and its retrieval application [11].

Table 1: List of Algorithm used for feature extraction

Algorithm	Approach	Features used for Retrieval
Color-Texture Classification	Segment images into regions	Color, Texture Features
Automated Scene Matching	Image matching	Wavelets, Color Descriptors
Content Based Face Retrieval	Image matching	self-organizing maps(SOMs) and user feedback Visual clustering
Multi-Modal Content Based Browsing and Searching Methods	P2P retrieval Systems	Key Frame Extraction, Shape Feature
Character Identification	Object retrieval of movie	Segmentation
Semantic Video Retrieval	Automatic Audio Categorization	Audio

Semantically Meaningful Summaries	Video sequence parsing to identify relevant camera views, and tracks ball movements	Scripted based, Co-occurrence Matrices
Recognizing Object in Video Sequences	Sequential classification methods for blob	Kalman filter applied over blob

III. Different Media Types and Their Properties

There are lots of applications and on the basis of an application context, either different elementary media types like text, image, audio, video or their combinations may be relevant. While all types of media have their associated features as timing constraints, size and different characteristics. Traditionally, different media are typically dealt with completely separate according to their coding style and content description properties. For the proper and efficient representations, we have to utilize the specific properties of different media in designing appropriate description schemes. Video is the best example of multimedia data as it contains several kinds of data such as text, image, meta-data, visual and audio. Content based video retrieval is an approach for facilitating the searching and browsing of large multimedia collections. At lower level, the proper description of data is usually derived in a bits centric mode and for the multimedia information at a conceptual level, the requirement of access becomes more content centric. And because of this, the bits-oriented description strategy may be inappropriate and inadequate for indexing and retrieval at the semantic levels. Now, we will understand the properties and description of different media types that comes in the multimedia content.

A. Text

Although more than one standard already exists for representation of text, within the context of multimedia, new standards may still be needed to facilitate search and retrieval. This is so because of the need to capture the correlation between text and other media. Text, say in ASCII, can be thought of as a collection of characters including letters and special symbols. By semantically combining continuous segments of these characters, we generate meaningful units such as words, phrases, sentences, paragraphs, etc., for the purpose of conveying information, in terms by writing, naming, speaking, etc. Depending on the application, a text transcript as an adjunct to audio and/or video may or may not be available, or required. For synchronization purpose text may also have a notion of time associated with it [6]. Thus, segments of text can be thought of as physically partitioning the text stream. Similar to audio and video, from a set of lower level physical events, many different types of logical events of text can be formed. The textual events include dialogues, stories, story categories, and news summaries.

B. Images

Image data may comprise of pictures, frames, showing / conveying some information or messages. Images conveys visual information as like video but without containing any audio / voice that includes mainly 2-D space and dimensions. The image data can be acquired as a bundle of multiple characteristics at some lines of resolution per image and an associated at one place and this image come in front of our eyes. After digitization, one can associate a spatial resolution in the horizontal and the vertical direction with each image / frame. Each image consist of distinct visual characteristics. As like partitioning in the audio, video processing a physical image is partitions as according to their characteristics such as number of objects presents, color of objects, background texture / color, and other characteristics, etc. Although characteristics to different visual events offer improvement over linear search, the search and retrieval capability is still limited to what physical data offers. To enable users the flexibility of defining these characteristics according to applications, by combining different segments of image characteristics that conveys some meaningful pictures to the viewer. Depending on the application, by combining multiple images called as stream of images it forms the video clip. For synchronization purpose, mostly for video images may also have timing constraints associated with it. Thus, segments of images can be thought of as physically partitioning different characteristics of the images.

C. Audio

The Audio media type may contain the data consisting of speech, music, or a combination of the both. Mostly the audio signal is inherently one-dimensional with the time constraints associated with it, the conventional search and retrieval on audio data is strictly linear. For enabling content-based retrieval, the audio stream has to be segmented and indexed first. Audio events can be defined as continuous audio segments that have some distinct characteristics. For example, such events can be single-speaker speech, music, multi-speaker conference speech, or speech over background music and can be analyzed and classified in a proper specification according to their feature space [8]. The classification process produce the required partitioning of the audio stream. Once an audio stream is partitioned, each event may be stored and described individually by their most unique features, indices to these events can be established, and content-based search and retrieval can then be accomplished using the indices.

Instead of the indexing and segmentation capability for the retrieval of audio content, there are other retrieval requirements, that are more logical rather than physical data chunks, called as more conceptual events. Such requirements demand the development of more advanced technologies than the ones that simply physically partition the data stream, but also the system support to manage audio data aggregation. It is obvious that logical events are related to how users view the data, as opposed to how the data exists physically, and that they are usually a composite of lower level events that can be physical or logical.

D. Video

Video data may comprise of images, audio stream or synthetic animations. Video conveys visual information that includes both 2-D space and time. The video data can be acquired as a stream of multiple frames at some lines of resolution per frame and are associated at one place and this stream of frame continuously come in front of our eyes becomes an video. Video Retrieval of Near–Duplicates using k–Nearest Neighbor Retrieval of Spatio–Temporal Descriptors [9]. It describes a novel methodology for implementing video search as retrieval of near-duplicate videos and recognition of actions in surveillance video. Videos are divided into half-second clips, the pixel regions with consistent color and motion properties are extracted from these 3D volumes by a threshold-free hierarchical space time segmentation technique. Each region is then described by a high-dimensional point whose components represent the position, orientation and when possible, color of the region.

The retrieval phase uses video segments as queries. The Fast Video Retrieval via the Statistics of Motion Within the Regions-of-Interest deals with very important issue to quickly retrieve semantic information from a vast multimedia database. In Trajectory-Based Video Retrieval Using Dirichlet Process Mixture Models [10], present a trajectory-based video retrieval framework using Dirichlet process mixture models. The main contribution of this framework is apply a Dirichlet process mixture model *(DPMM)* to unsupervised trajectory learning. Employ a time-sensitive Dirichlet process mixture model *(tDPMM)* to learn trajectories' timeseries characteristics. The framework has a good scalability and adaptability in the sense that when new cluster data are presented, the framework automatically identifies the new cluster information without having to redo the training.

Video Annotation for Content-based Retrieval using Human Behavior Analysis and Domain Knowledge [12] presents an automatic annotation method of sports video for content-based retrieval. This approach incorporates human behavior analysis and specific domain knowledge with conventional methods, to develop integrated reasoning module for richer expressiveness of events and robust recognition.

IV. Multimedia Content Searching and Retrieval Techniques

As there are different types of data involved in the multimedia storage system each of the data type has their own different feature that have to be extracted. These different characteristics or features includes Color, edges, shape, timing constraints and texture serves as the common characteristics that are used to extract features for retrieving interesting contents. Feature extraction on these attributes may be performed at the different level of access [13]. The Feature Extraction is one of the most effective method and play a vital role in the multimedia content searching and retrieval system as all the media types have their own feature associated with them that are mainly used to mine data from large storage system. Now, we will see the searching and extraction of all the useful content according to different media types.

A. Text Retrieval:

In text searching and retrieval method, its feature extraction is usually perform by means identifying the keywords that summarize the contents of the document. One way is to look for words that occur frequently in the document, most of the time these words are useful to know what the document is about. From the remaining words, by using the intelligence one has to look for words that occur frequently in documents of the same class, but rarely in documents of other classes. In order to cope with documents of different lengths, relative frequency is preferred over absolute frequency of the words occurs in the text documents.

B. Image Retrieval:

In case of Image searching and retrieval, one has to use the features as element present in that image, colors of the elements and background, etc. The special features of image content extraction can be include somewhat modality features as its brightness, blur, contrastness, color combination, etc. The cross modal retrieval can be perform mainly on the text and image annotations, this serves as very useful feature [6]. As images servers as to convey visual information as like video but without containing any audio / voice that includes mainly 2-D space and dimensions can be said as static. As we have seen above, image data can be acquired as a bundle of multiple characteristics that can be said as some lines of resolution per image and is associated at one place and this image come in front of our eyes. So, at the time of retrieving image data these characteristics / features are very much useful.

C. Audio Retrieval:

The audio searching and retrieval works mostly as like as video retrieval system. As like video both the temporal and the spectral domain features have been employed in the audio searching system. Some of the features used include short-time energy, pause rate, zero-crossing rate, fundamental frequency, spectrum, bandwidth, spectral centroid, spectral roll-off frequency and band energy ratio [15]. From audio, extracting the speech is semantically very rich. Based on these features and according to application requirement one is able to perform audio retrieval properly.

D. Video Retrieval:

Till now, the techniques that should be apply on video data for its searching and retrieval, one of the most important steps is to transform video from unstructured formatted data sets to the formatted data set. This video has one complex type as changing stream of images that serves as dynamic feature extraction and retrieval of video contents. Video as a whole is very large data to mine that creates the need of some preprocessing to get data in the suitable format for searching. Video data is consist of spatial, temporal data, images, audio clips, etc. features. Based on applications requirement these features are used for searching and retrieval. Commonly, video is the hierarchical construction, having number of frames, streams, segments, scenes, clips and full length video. Every video clips unit has its own features which are useful for getting particular useful video clips [14].

V. Conclusion

This paper provides us the useful information about different media types that comes under the word multimedia. All the media types have their own different properties associated with them. Then, this paper tries to give information about different techniques to extract these different types of media data. An important issue with features extraction from multidimensional data is how the features should be integrated for searching and retrieval. Most multimodal analysis is usually performed based on the features, and the results are brought together at a later stage to arrive at the final decision about the applications input data. The content searching and retrieval through this approach is known as cross-modal analysis because such an approach allows the discovery of semantic associations between different media types of data.

References

[1] Sarla More, and Durgesh Kumar Mishra, "Multimedia Data Mining: A Survey", *PRATIBHA: INTERNATIONAL JOURNAL OF SCIENCE, SPIRITUALITY, BUSINESS AND TECHNOLOGY (IJSSBT),* Vol. 1, No.1, March 2012.

[2] Chitra Wasnik, "Tools, Techniques And Models For Multimedia Database Mining", International Journal of Networking & Parallel Computing, Volume 1, No 2, pp. 1-5, November 2012.

[3] Avinash N Bhute, and B B Meshram,"SYSTEM ANALYSIS AND DESIGN FOR MULTIMEDIA RETRIEVAL SYSTEMS", The International Journal of Multimedia & Its Applications (IJMA) Vol.5, No.6, December 2013.

[4] Nilesh Patel and Iswar Sethi. Multimedia data mining: An overview. Multimedia Data Mining and Knowledge Discovery, Springer, 2007

[5] MICHAEL S. LEW, NICU SEBE CHABANE, DJERABA LIFL, and RAMESH JAIN, "Content-Based Multimedia Information Retrieval: State of the Art and Challenges", *ACM Transactions on Multimedia Computing, Communications and Applications,* Vol. 2, No. 1, February 2006.

[6] Shirish B. Patil, SaurabhBonde, B.B.Meshram, "A Survey on Multimedia Data Mining", *IJCEM International Journal of Computational Engineering & Management*, Vol. 14, October 2011 ISSN (Online): 2230-7893 www.IJCEM.org,IJCEM www.ijcem.org.

[7] Qian Huang, Atul Puri, *Member, IEEE*, and Zhu Liu, "Multimedia Search and Retrieval: New Concepts, System Implementation, and Application", *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY*, vol. 10, no. 5, august 2000.

[8] Z. Liu and Q. Huang, "Classification of audio events in broadcast news," in *Proc. IEEE 2nd Workshop Multimedia Signal Processing*, Dec. 1998, pp. 364–369.

[9] Daniel DeMenthon, David Doermann (2003), "Video Retrieval of Near–Duplicates using k–Nearest Neighbor Retrieval of Spatio–Temporal Descriptors", ACM Multimedia '03, November 2.8, 2003, Berkeley, California, USA.

[10] Xi Li, Weiming Hu, Zhongfei Zhang, Xiaoqin Zhang, Guan Luo (2008), "Trajectory-Based Video Retrieval Using Dirichlet Process Mixture Models", 19th British Machine Vision Conference (BMVC).

[11] B V Patel, B B Meshram, "CONTENT BASED VIDEO RETRIEVAL SYSTEMS", International Journal of UbiComp (IJU), Vol.3, No.2, April 2012.

[12] Hisashi Miyamori Shun-ichi Iisaku(2000), "Video Annotation for Content-based Retrieval using Human Behavior Analysis and Domain Knowledge" Fourth IEEE International Conference on Automatic Face and Gesture Recognition.

[13] Pukhraj P. Shrishrimal, Ratnadeep R. Deshmukh, Dr. Vishal Waghmare, "Multimedia Data Mining: A Review", CONFERENCE PAPER · AUGUST 2014 [Online] available: http://www.researchgate.net/publication/270899739.

[14] Bhatt c. A., "Probablilistic Temporal Multimedia Data mining", Ph.D. Thesis, National University of Singapore, 2012.

[15] Jadhav S. R., Kumbargoudar P., "Multimedia Data Mining in Digital Libraries: Standards and Features", *IEEE International Conference on Advances in Computer Vision and Information Technology*, Dr. Babasaheb Ambedkar MarathWada University, Aurangabad, MSIndia, 2007.

[16] Li Dongge, Dimitrova Nevenka, Li Mingkun, and Sethi Ishwar K., "Multimedia content processing through cross-modal association".

