International Journal of Research In Science & Engineering e-ISSN: 2394-8299
Volume: 3 Issue: 2 March-April 2017 p-ISSN: 2394-8280

CONTENT-BASED VIDEO RETRIEVAL USING VECTOR QUANTIZATION

Salil Manohar[1], Nalini Yadav[2], Shubham Priyadarshi[3], Vaibhav Mittal[4]

[1,2,3] *Student, Dept. of Computer Engineering, D.Y.P.I.E.M.R, Akurdi, Pune*
[1] *replysalil@email.com* [3] *priyadarshi2193@email.com* [4] *mittalvaibhav100@email.com*
[4] *Assistant Professor, Dept. of Computer Engineering, D.Y.P.I.E.M.R, Akurdi, Pune*
[2] *yadav.nalini@email.com*

ABSTRACT

In the recent years there is a significant growth in video data. Due to expressive power of video, video data has increased enormously. There is need to manage these significantly increasing data efficiently. Although there are different video database management systems available that do the retrieval of videos but the methods are not efficient enough. There exist a semantic gap in retrieval algorithms. The different methods of video retrieval are namely text based, audio or speech based, metadata based, content based and integrated based. Currently most of application that requires video retrieval , has been implemented on text based method of retrieval ,which is not efficient method of retrieval as there exist a semantic gap between retrieval algorithm and users query .The idea is to provide an efficient way of managing and retrieving video from large database efficiently .CBVR(content based video retrieval) is one of the video data management method to enhance the retrieval and management of video which reduces the semantic gap. It focuses on content of video instead of just metadata of the Video categorization and retrieval have a large spectrum of promising applications, motivating the interest of researchers worldwide. This paper offers a tutorial and an summary of the landscape of general ways in visual content-based video categorization and retrieval, that specialize in ways for video structure analysis, as well as shot boundary detection, key frame extraction and scene Segmentation, extraction of options as well as static key frame options, object options and motion options, video data processing, video annotation, video retrieval as well as question interfaces, similarity measure and significance feedback, and video browsing. Finally, we tend to analyse future analysis directions.

Keywords*: CBVR, Feature Extraction, Video Retrieval, Video Segmentation, OCR, ASR tool, Re-ranking, Vector quantization.*

---

1. INTRODUCTION

Multimedia info classification and retrieval area unit needed to explain, store, and organize multimedia system info and to help folks to find multimedia system resources handily and quickly. Dynamic video is a vital sort of multimedia system info. Videos have the subsequent characteristics:

1) Much richer content than individual images;.

2) Huge quantity of information.

3) Little previous structure.

These characteristics create the classification and retrieval of videos quite troublesome. In the past, video databases are comparatively tiny, and classification and retrieval are supported keywords annotated manually. A lot of recently, these databases became abundant larger and content-based classification and retrieval area unit needed, supported the automated analysis of videos with the minimum of human participation. Content-based video classification and retrieval have a good vary of applications like fast browsing of video folders, analysis of visual electronic commerce (such as analysis of interest trends of users' alternatives and orderings, correlations analysis between advertisements and their effects), remote instruction, digital museums, occurrence analysis, intelligent management of net videos (useful video search and harmful video tracing), and video police work. It's the broad vary of applications that motivates the interests of researchers worldwide. The subsequent 2 samples of analysis activity area unit significantly noteworthy [10].

1) Since 2001, the National Institute of Standards and Technology has been sponsoring the annual Text Retrieval Conference (TREC) Video Retrieval analysis (trecvid) to push progress in video analysis and retrieval. Since 2003, trecvid has been freelance of TREC. Trecvid provides a large-scale take a look at assortment of videos, and dozens of participants apply their content-based video retrieval algorithms to the gathering.

2) The goal of video standards is to make sure compatibility between Description interfaces for video contents so as to facilitate the event of quick and correct video retrieval algorithms. The most standards for videos area unit the show specialists cluster (MPEG) and therefore the TV-Anytime customary.

There exist several investigations that adopt the MPEG-7 to extract options to classify video contents or to explain video objects within the compressed domain. A video could have associate degree modality channel furthermore as a visible channel.

The accessible info from videos includes the following:

1) Video information, that area unit labeled texts embedded in videos, typically as well as title, summary, date, actors, producer, broadcast period, file size, video format, copyright, etc.;

2) Audio info from the modality channel;

3) Transcripts: speech transcripts is obtained by speech recognition and caption texts is scan victimization optical character recognition techniques;

4) Visual info contained within the pictures themselves from the visual channel. If the video is enclosed in a very online page, there area unit typically online page texts related to the video. During this paper, we tend to concentrate on the visual contents of videos and provides a survey on visual content-based video classification and retrieval. The importance and recognition of video classification and retrieval have LED to many survey papers, that area unit listed beside the publication years and topics. For instance, provides a smart review of video shot boundary detection throughout seven years of the trecvid activity. Snoek and Worring gift an in depth review of concept-based video retrieval. They emphasize linguistics thought detection, video search victimization linguistics ideas, and therefore the analysis of algorithms victimization the trecvid databases. Ren et al. Review the state of the art of spatiotemporal linguistics information-based video retrieval.

2. LITERATURE SURVEY

Video assortment and retrieval have a good spectrum of promising applications, motivating the interest of researchers worldwide. This paper offers a tutorial and an outline of the landscape of general methods in visual content-based video assortment and retrieval, that specialize in strategies for video structure analysis, as well as shot boundary detection, key frame extraction and scene segmentation, extraction of options as well as static key frame options, object options and motion options, video data processing, video annotation, video retrieval as well as question interfaces, similarity live and connectedness feedback, and video browsing. Finally, we tend to analyse future analysis directions [1, 6].

Retrieval in current multimedia system databases is sometimes restricted to browsing and looking out supported low-level visual options and specific matter descriptors. Linguistics aspects of visual info square measure chiefly delineate fully text attributes or mapped onto specialized, application specific description schemes. Result lists of queries square measure normally described by matter descriptions and single key frames. This approach is valid for text documents and pictures, however is usually lean to represent video content in a very substantive manner. During this paper we tend to gift a multimedia system retrieval framework specializing in video objects that totally depends on the MPEG-7 commonplace as info base. It provides a content-based retrieval interface that uses gradable content-based video summaries to permit for fast viewing and browsing through search results even on information measure restricted net applications [2].

Traditional audio-visual archives area unit these days being replaced by digital multimedia system content management systems. These systems manage the audio-visual information itself moreover as further work data (meta-data). to truly create these resources offered and make the most their content typically manual annotation by specialists is needed, that makes the task of work not solely time overwhelming however conjointly terribly dearly-won. Particularly visible of the speedy enlargement of digital media, tools to facilitate or automatize this method become indispensable. During this paper, we have a tendency to describe the combination of period speech and language technologies into a multimedia system archiving system [3].

Multimedia data retrieval (MIR) is regarding the look for data all told its forms, everywhere. Indeed, what sensible is all the data within the world if it's insufferable to search out anything? This sentiment is reflected as an ACM SIGMM grand challenge [Rowe and religion 2005]: "make capturing, storing, finding, and victimization digital media an everyday incidence in our computing surroundings." This paper is supposed for researchers within the space of content-based retrieval of multimedia system. Currently, the basic downside has been a way to modify or improve multimedia retrieval victimization content-based strategies. Content-based strategies are necessary once text annotations are non-existent or incomplete. Moreover, content-based strategies will probably improve retrieval accuracy even once text annotations are gift by giving extra insight into the media collections [4].

It is unremarkably acknowledged that ever-increasing video databases ought to be expeditiously indexed to facilitate quick video retrieval. Categorizing web-based videos is a vital however difficult task. The difficulties arise from giant information diversity inside a class, lack of tagged information etc. Similarity matching formula plays a vital role in Video Retrieval System. Most of the video retrieval systems area unit designed victimisation ancient similarity matching algorithms that area unit supported distance measures. The Accuracy of retrieval system depends on the tactic used for police work shots, reasonably video options used for retrieval. Linguistics video categorization could be a step towards automatic video categorization and retrieval. Here a Latent linguistics categorization (LSI) technique, supported Singular price Decomposition and fusion of visual options like colour and edge is planned for video categorization and retrieval. A key feature of LSI is its ability to ascertain associations between similar types of data, that the likelihood of manufacturing correct index is incredibly high [5].

3. PROPOSED SYSTEM

Proposed CBVR customized Systems enforced on six modules i.e. creation of options and stores in information and retrieval victimization feature extraction with similarity measures and personalization as shown in figure a pair of. Firstly, a user uploads or provides a text/image/video question as input to the CBVR customized system. CBVR System can divide the video into frames and will choice method of relevant frames into all frames. at the same time ASR system can method on video input and extract the keywords by ASR tool. when frame segmentation and choice, perform OCR and extract the HOG, OCR text and Gabor Filter from elite frames and additionally extract the colour, Texture and Edge detector on elite frames and ultimately additionally extract the keywords and options. Identical method of ASR, Frame segmentation, OCR and image processing is finished on videos hold on information. When pre-processing system can rummage around for similarity in keywords and options of user question information and every one video that are hold on in information. CBVR system extracts the foremost matching OCR text, ASR text and keywords and options and generates relevant final video results.

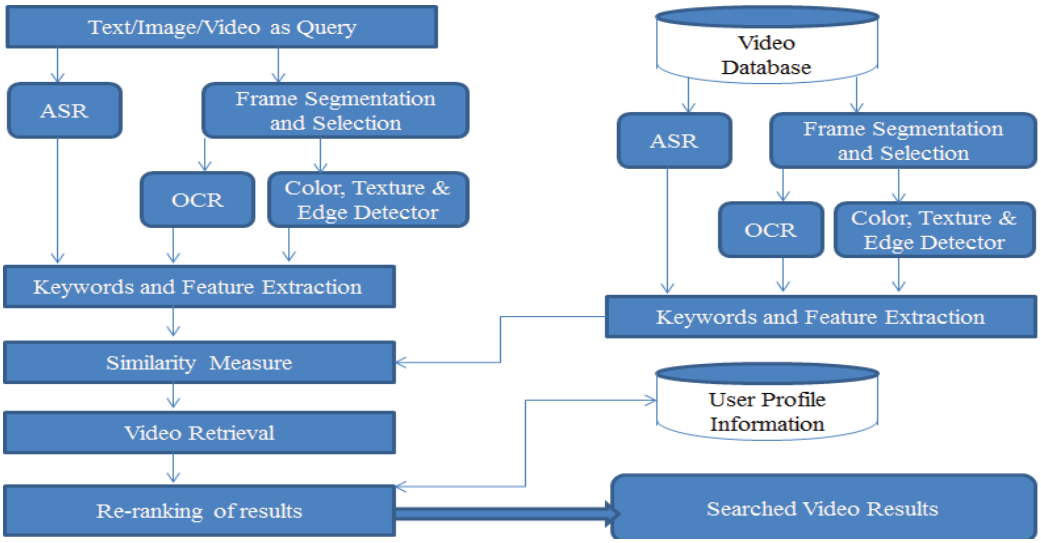

Fig.1 System Architecture

4. IMPLEMENTATION

A. Frame Segmentation and selection Technique:

If the video contains structure, i.e. many shots, then the quality techniques for video report involve:

1. Calculate the video Length

2. Divide the frame by specific interval.

• However, allow us to assume you would like to search out a motivating in close one continuous stream of frames taken from one camera supply.

• A mean colour bar graph is computed for all frames and therefore the key-frame is that with the highest bar graph i.e. system selects the most effective in close terms of its colour distribution.

• System assumes that camera stillness is associate degree indicator of frame importance. As advised by Beds, above. Then choose the still frames victimization optic-flow and use that.

• Each frame is projected into some high dimensional content house; system realizes those frames at the corners of the space and use them to represent the video. Frames are evaluated for importance victimization their length and novelty in content house.

B. Colour and Texture:

Proficient detection and segmentation of text characters from the background is necessary to fill the gap between image documents and the input of standard OCR systems [5]. The basic visual features of index frame include colour as well as texture. Features extracted from index frames are stored as feature database which is used for object-based video retrieval. Texture is yet another important property of index frames. Different texture representations have been investigated in pattern recognition and computer vision.

C. Edge detection, Colour:

Tremendous growth in digital content which includes images, audios and videos on internet and on desktop is demanding development through new technologies using different methods for representation as well as storage and retrieval of multimedia systems. Video feature database is created using entropy feature which is extracted by using key-video frames of database containing video. Same feature is extracted from video frame query [7].

D. ASR (Automatic Speech Recognition):

Apart from video ASR can be used to provide speech-to-text information from different videos, which offers the chances in improvement of the quantity of automatically generated metadata. However, as mentioned, it is difficult to achieve significant speech recognition for most of videos. "Computer speech recognition" or simply "speech to text"(STT). Some SR systems use speaker-independent speech recognition while other method utilise "training method" in which a speaker is supposed to read some portion of texts in ASR system. These systems analyse the person's specific voice and use it to fine tune the popularity of that person's speech, leading to additional correct transcription. Systems that don't use coaching are referred to as "speaker independent" systems. Systems that use coaching are referred to as "speaker-dependent" systems. Machine-controlled Speech Recognition (ASR) in engineering and applied science, Speech Recognition (SR) is that the translation of spoken words into text. it's conjointly called "automatic speech recognition" (ASR), Speech recognition applications embody voice user interfaces like voice dialing (e.g. "Call home"), decision routing (e.g. "I would really like to form a collect call"), demotic appliance management, search (e.g. realize a podcast wherever specific words were spoken), straightforward information entry (e.g., coming into a MasterCard number), preparation of structured documents (e.g. a radiology report), speech-to-text process (e.g., word processors or emails), and craft (usually termed Direct Voice Input). The term voice Recognition or recognition refers to characteristic the speaker, instead of what they're speech. Recognizing the speaker will modify the task of translating speech in systems that are trained on a selected person's voice or it may be wont to evidence or verify the identity of a speaker as a part of a security method [9, 11].

E. Optical Character Recognition (OCR):

Optical character recognition (OCR) is a very important analysis space in pattern recognition. The target of associate degree OCR system is to acknowledge alphabetic letters, numbers, or different characters, that are within the kind of digital pictures, with none human intervention.

This is often accomplished by looking out a match between the options extracted from the given characters image and therefore the library of image models. Ideally, they'd just like the options to be distinct for various character pictures in order that the pc will extract the proper model from the library with none confusion

F. Personalization of the Results:

Reranking the retrieved results according to users interest which may include different models with different user details.

G. Vector quantization

Vector quantization (VQ) contains some process of clustering. Vector quantization compression technique has two different components: VQ encoder and VQ decoder. In VQ through the given method the image is divided into non overlapping image blocks X = { x0,x1,x2,…xm} of size 2 x 2 pixels each and a clustering algorithm , is used to generate a codebook C= {Y1,Y2,Y3…Yn} for the given set of image blocks. A set of representative image blocks called code words are contained in codebook C.A closest match code word is looked in the codebook by VQ encoder for each image block and then index of the code word so generated is transmitted towards VQ decoder. In the decoding phase, the index values along with respective code words from the codebook is replaced by VQ decoder and then the quantized image, called as reconstructed image is produced[8].

Vector Quantization has been used in number of applications, speech recognition and face detection, colorization, Image segmentation, speech data compression, Contained Based Image Retrieval, face recognition, iris recognition, tumour detection in mammography etc.

Steps of Vector Quantization:

1. Temporary matrix
2. Number to double format
3. Shift mean to zero
4. 2D Cosine transform function
5. Divide by Quantization matrix

Convert Quantization matrix to 1D vector

5. CONCLUSION

We have conferred a review on recent developments in visual content-based video compartmentalization and retrieval. The state of the art of existing approaches in every major issue has been delineated with the main focus on the subsequent tasks: video structure analysis together with shot boundary detection, key frame extraction and scene segmentation, extraction of options of static key frames, objects and motions, video data processing, video classification and annotation, video search together with interface, similarity live and connectedness feedback,

and video account and browsing. At the tip of this survey, we've mentioned future directions like emotive computing-based video retrieval and distributed network video retrieval.

REFERENCES

[1] Weiming Hu, Senior Member, IEEE, Nianhua Xie, Li Li, Xianglin Zeng, and Stephen Maybank, "A Survey on Visual Content-Based Video Indexing and Retrieval," IEEE vol 41 november 2011

[2] Werner Bailer, Harald Mayer, "Content-based Video Retrieval and Summarization using MPEG-7", A-8010 Graz, Austria.

[3] Jürgen Riedler, Gerhard Backfried, Alfred Walter,Ken Hampel," Integrating Multimedia Archiving with Real-time Speech and Language Technologies", Sail Labs, Vienna, Austria.

[4] MICHAEL S. LEW, NICU SEBE, CHABANE DJERABA, RAMESH JAIN," Content-based Multimedia Information Retrieval: State of the Art and Challenges" In ACM Transactions on Multimedia Computing, Communications, and Applications, Feb. 2006.

[5] Rashmi M, Roshan Fernandes," Video Retrieval Using Fusion of Visual Features and Latent Semantic Indexing", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Issue 5, May 2014.

[6] Chivadshetti Pradeep, Mr Kishor Sadafale, and Mrs Kalpana Thakare.. "Content Based Video Retrieval Using Integrated Feature Extraction" in IPGCON 2015 at AVCOE, Sangmner on March 2015.

[7] Thakre, Kalpana S., Archana M. Rajurkar, and Ramchandra Manthalkar. "An effective CBVR system based on motion, quantized color and edge density features." In Proceedings of the First International Conference on Intelligent Interactive Technologies and Multimedia, pp. 145-149. ACM, 2010.

[8] Sudeep D. Thepade, Pritam H. Patil " Novel Keyframe Extraction for Video Content Summarization using LBG Codebook Generation Technique of Vector Quantization." International Journal of Computer Applications (0975 – 8887) Volume 111 – No 9, February 2015

[9] Haojin Yang and Christoph Meinel, "Content Based Lecture Video Retrieval Using Speech and Video Text Information", in proc. IEEE transactions On Learning Technologies, Vol. 7, No. 2, April-June 2014

[10] Arpit Jain, Xujun Peng, Xiaodan Zhuang, Pradeep Natarajan, and Huaigu Cao, "Text Detection and Recognition in Natural Scenes and Consumer Videos", in proc. IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP) INSPEC Accession Number: 14448982, 4-9 May 2014

[11] Hadi Yarmohammadi and Mohammad Rahmati, "Content Based Video Retrieval using Information Theory", in proc. IEEE Iran Conf. ,Machine vision and Image Processing, pp. 214-218, 2013