Australian Journal of Basic and Applied Sciences, 9(5) March 2015, Pages: 321-326

ISSN:1991-8178

Australian Journal of Basic and Applied Sciences

Journal home page: www.ajbasweb.com

The model approach of GLCM method of keyframe extraction and Kullback- Leibler distance similarity measure for Content based video retrieval system from video databases

[1]A.Mohamed Mustaq Ahmed, [2]M.Mohemmed Sha, [3]Dr.G.S.Anandha Mala

[1]*Dept of Computer Science & Information, Arts and Science College, Wadi Addawasir, Saudi Arabia*
[2]*Dept of Computer Science & Information, Arts and Science College, Wadi Addawasir, Saudi Arabia*
[3]*Head of the Dept, Dept of Computer Science & Engineering, St.Joseph's College of Engineering, Chennai, India*

ARTICLE INFO

Article history:
Received 12 November 2014
Received in revised form 26 December 2014
Accepted 29 January 2015
Available online 14 February 2015

Keywords:
Content Based Retrieval, Query by Content, inherent temporal dimension, shots, scenes, Shot segmentation, Keyframe extraction, 2-D Correlation Coefficient technique, HSV color model, Gray level co -occurrence matrix, Kullback- Leibler distance

ABSTRACT

In Multimedia Databases, Content Based Retrieval or Query by Content has been proposed as an alternative to text based retrieval for various media such as image, video, and audio. Text based retrieval is no longer appropriate for indexing such media for several reasons. Nowadays, the capture, storage, uploads and delivery of videos has become effortless due to the rapid advancements in digital devices, Internet infrastructures, and Web technologies. Video streams are considered the most complex form of all multimedia data because they contain almost all other forms such as images and audio in addition to their inherent temporal dimension. At the lowest level, the video consists of a set of frames. At the next, higher level frames are grouped into shots. Consecutive shots are aggregated into scenes based in story-telling coherence. Shots are basic structural building blocks in video. A shot in video information may be defined as sequence of continuous images from a single camera at a time. Content based video retrieval system from video databases uses Shot segmentation, Keyframe extraction of videos and the extraction of low level features Color, shape and texture of videos.

- In this model, each video is segmented and the key frames are extracted. 2-D Correlation Coefficient technique is applied for video shot segmentation in this video retrieval system. Here, corresponding key frames are extracted once the shots boundaries are detected from the videos. First and last frames of each shot are extracted.
- Then color feature is extracted from key frames using HSV color model. Shape feature has been extracted from key frames using canny edge detection algorithm.
- Textural features are extracted from key frames using gray level co -occurrence matrix. After extracting all of these features from key frames, feature database has been generated from these features.
- In video retrieval, the system extracts low level features for every key frame of video library and the extracted video features are stored in the feature library. Then, the same features are extracted for a query image and are compared with the features in the feature library with the aid of Kullback- Leibler distance similarity measure. Finally the relevant videos are retrieved from the collection on the basis of Kullback- Leibler distance.

To Cite This Article: A.Mohamed Mustaq Ahmed, M.Mohemmed Sha and ,Dr.G.S.Anandha Mala , The model approach of GLCM method of keyframe extraction and Kullback- Leibler distance similarity measure for Content based video retrieval system from video databases,. ***Aust. J. Basic & Appl. Sci.,*** *9(5): 321-326, 2015*

INTRODUCTION

Text based video retrieval system works based on metadata attributes like name, date of creation, tagged words, annotation, etc. ie Video data have to be manually annotated with keywords during cataloguing stage. In retrieval phase, based on the text query the desired videos retrieved. This however leads to unsatisfactory results to users because video data are manually annotated with keywords. Content Based Video Retrieval (CBVR) system works more effectively as these deals with content of video rather than video metadata. The content-based video retrieval systems that can effectively retrieve the videos to a given query frame have been attracted much attention in the recent year. Content-based retrieval system of video data consists of the following four stages; (1) Video shot boundary detection, (2) Key frames extraction, (3) Feature extraction from the selected key frames, (4) Retrieval stage where similarity matching operations are performed.

Corresponding Author: A.Mohamed Mustaq Ahmed, Dept of Computer Science & Information, Arts and Science College, Wadi Addawasir, Saudi Arabia

In CBVR, shot boundary detection is a very active research topic. Digital video information often consists of series of 25 to 30 frames per second and an associated synchronized audio track. In order to develop any content based manipulations on digital video information, the video information must be first be structured and broken down into components. At the lowest level, the video consists of a set of frames. At the next, higher level frames are grouped into shots. Consecutive shots are aggregated into scenes based in story-telling coherence. Shots are basic structural building blocks in video. A shot can be defined as a sequence of frames taken by a single camera without any significant change in the color content of consecutive images.

A shot boundary is the gap between two shots. Naturally, boundaries between shots need to be determined automatically. A number of researchers utilized robust techniques on basis of the color histogram comparison to accomplish this function.

After the boundaries are found, each shot can be represented with an appropriate key frame. In current video indexing systems, after temporal segmentation of analyzed sequence (shot detection), a set of frames that best represent the visual content of the scenes is extracted. These frames are called key-frames and are used in latter task of video search and retrieval. The popular way to choose key frame is choosing first frame of each shot. Deriving content indexes from the selected key frames is the next stage. During this step, a number of low level spatial features (Color, shape, texture,) are extracted from key frames and stored in database.

The color, edge, and texture features are extracted for a query image and evaluated against the features in the feature library. The similarity measure is carried out with the help of some similarity measure algorithm. Later, the corresponding videos are retrieved from the collection of videos on the basis of the similarity.

CBVR systems should be able to automatically extract visual features that are used to describe an image or video clip. Examples of such features are color, shape and texture. Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. Transforming the input data into the set of features is called feature extraction. Each extracted feature is encoded in n-dimensional vector called feature vector.

II. Related Works:

There are different approaches used to detect the shot in a video.

Shot boundary detection scheme based on rough fuzzy set: Han et al describe a technique for video shot boundary detection using rough fuzzy set. The selected low-level features are essential to achieve high accuracy for shot boundary detection. But there are too many features available in the frame or video, such as pixel values of different color channels, statistic features, intensity and color histogram etc. By choosing the most appropriate features to represent a shot or video, the computational burden will be reduced and the efficiency will be improved. For this purpose, the feature optimal choice method based rough sets is introduced in this section. In computation, we compute the mean of every component of each frame in the RGB or HSV model. There are two types of false detections in videos. One results from the existence of irregular camera operations during the gradual transitions. The other is due to a lot of flash effects in a shot.

The misses are mainly due to the small content changes between the frame pairs at some shot boundaries. During the experiments, we find that the false detection in the coarse detection will affect on the following feature extraction and shot boundary detection while the missed detection have less effects because the rough-fuzzy calculator weaken the mistakes in coarse detection stage, the dissimilarity function is more fit for varies video. Based on rough fuzzy set, by which the dissimilarity function for shot boundary detection is obtained. It shows that the proposed methods not only for both similarity and effective but also can decrease data dimensions and preserve the information of original video farthest.

Shot boundary detection in low-pass filtered histogram space: Han and Yoon, describes a technique for video shot detection using low pass filtered histogram space. Twin-comparison was developed to find shot boundaries among cuts and fades/dissolves using two thresholds. Gunsel and Tekalp proposed one threshold method using Otsu method to find the threshold automatically. However, this system was presented for detection of cut-type shot boundaries. In model-based method , the edit effect showing gradual changes (fades, dissolves, etc.) presents edit invariant property that is used in classifying shot boundaries. In which accentuates edit constancy effects by applying low pass filtering to histogram differences between frames, while suppressing motion effects causing false alarms.

The hidden markov model technique: Boreczky and Lynn, describe a technique for segmenting video using hidden markov model. It uses three types of features for video segmentation-the standard histogram difference, an audio distance measure and an estimate of object motion between two adjacent frames. The histogram feature measures the distance between adjacent frames based on the distribution of luminance levels. The parameters of the HMM are learned through a training phase. Once the parameters are trained, segmenting the video into its shots, camera motions and transitions is performed using the viterbi algorithm. Given the sequence of features, the viterbi algorithm produces the sequence of states most likely to have generated these features.

Shot change detection based on sliding window method: Li and Lee, describe a technique for shot change detection based on sliding window method. The Conventional Sliding Window (CSW) method has long been used in video segmentation for its adaptive thresholding technique. A hard cut is detected based on the ratio between the current feature value and its local neighborhood in the sliding window. Yet it has a relatively high rate of false alarm and missed cuts.

Histogram based detection: Colin et al. present a detailed evaluation of a histogram-based shot cut detector. The algorithm was specifically applied to large and diverse digital video collection consisting of eight h of TV broadcast video. It was found that the selection of similarity thresholds for determining shot boundaries in such broadcast video was difficult and necessitates the development of adaptive thresholding in order to address the huge variation of characteristics. The histogram creation technique used compared successive frames based upon three 64-bit histograms (one of luminance and two of chrominance).

Shot segmentation by graph partitioning: In this method an automated shot boundary detection based on the comparison of more than two consecutive frames is used. Within a temporal window we calculate the mutual information for multiple pairs of frames. This way we create a graph for the video sequence where the frames are nodes and the measures of similarity correspond to the weights of the edges. By finding and disconnecting the weak connections between nodes we separate the graph to sub-graphs ideally corresponding to the shots. The method is able to detect efficiently abrupt cuts and all types of gradual transitions, such as dissolves, fades and wipes with very high accuracy.

A key-frame is a frame that represents the content of a shot or scene. This content must be the most representative as possible. multiple key frame by unsupervised approach following a divide and conquer strategy shot boundary detection -for which reliable standard techniques exist-is used to divide key frame extraction into shot-level sub problems that are solved separately.

Feature Extraction: Video key frame edges are detected using high pass filter approaches of image processing. canny, Robert, Sobel, Kirsch, Laplacian and Prewitt are most popular edge detection techniques.

Video frame entropy is measured using Grey Level Co- Occurrence (GLC) matrix which is one of the most widely used texture measures. It also measures the frequency of adjacent pixels patterns.

III. Video Retrieval Architecture:

The proposed video retrieval architecture has shown in the Figure 1, which consists of five sub process. The proposed system consists of the following modules:

- Video shot boundary detection
- Key Frames selection
- Feature extraction (shape, color, texture)
- Feature database generation
- Video retrieval

As a first stage of this system, videos data set is collected. Then each video is segmented and the key frames are extracted. A number of researchers utilized robust techniques on basis of the color histogram comparison to accomplish this function. 2-D Correlation Coefficient technique is applied for video shot segmentation in this video retrieval system.

Later, corresponding key frames are extracted once the shots boundaries are detected from the videos. First and last frames of each shot are extracted. Then color feature is extracted from key frames. The whole HSV color space is quantified into 72 kinds of colors. Then shape feature is extracted from key frames using canny edge detection algorithm. Then texture features (entropy, energy) are extracted from key frames using gray level co-occurrence matrix. After extracting all of these features from key frames, feature database has been generated from these features. Construction of video database, a collection of video sequences and generation of feature vectors is completely processed offline in this system.

In video retrieval, the system receives query video frame input from the user. For each key frame of videos in the video database, this system extracts four different kinds of low level features, including edge, texture and color of every key frame and the extracted video features are stored in the feature library. Then, the same features are extracted for a query image and are compared with the features in the feature library.With the aid of Kullback- Leibler distance similarity measure, the comparison is carried out. Finally the relevant videos are retrieved from the collection on the basis of Kullback- Leibler distance.

This system has been validated by experiments with several kinds of video sequences. This system has been implemented in Matlab. To assess the retrieval effectiveness, this thesis has used the precision and recall metrics. Table1 shows the precision and recall performance of the content based video retrieval system for a set of given query images(frames).

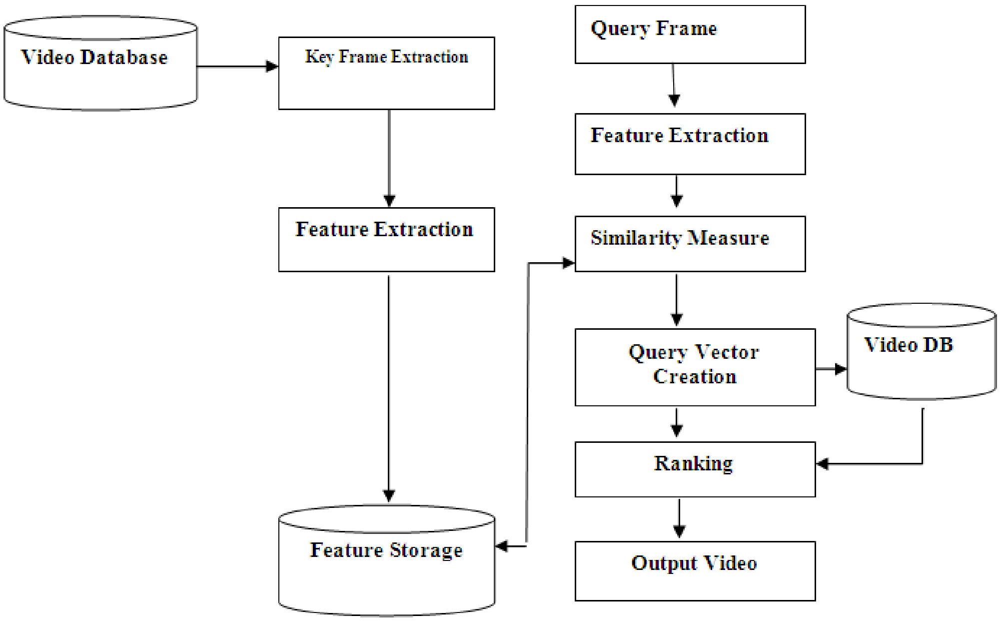

Fig. 1: Video Retrieval System architecture.

Table 1: Precision and Recall for a given set of query frames.

INPUT Frame	PRECISION (%)	RECALL (%)
F1	100	93
F2	98	90
F3	94	88
F4	94	86
F5	96	83
F6	98	81
F7	91	89
F8	90	91
F9	91	86
F10	91	84

From the table 1 it was observed that the precision is good for the video retrieval system. Recall has been reduced in certain retrieval due to the fact that the system is unable to retrieve all the related videos if the query image is not present in the database and the feature vector for the query image does not exist. Figure 2 shows the performance of the video retrieval system.

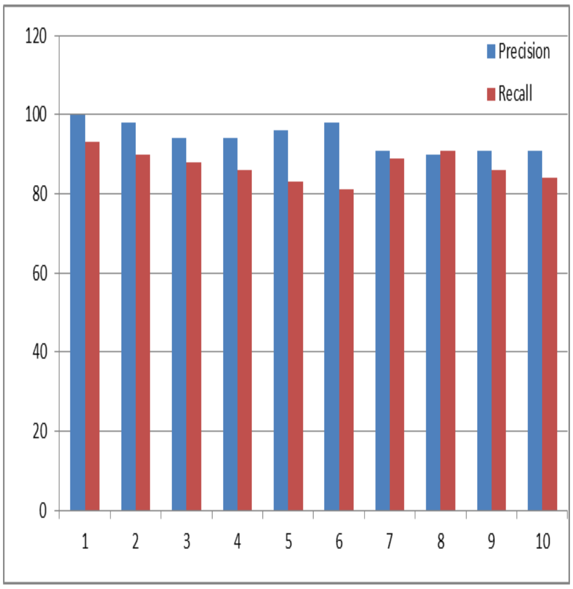

Fig. 2: Performance of Video Retrieval System.

From the graph, it was observed that the proposed system exhibits a good acceptable precision and recall which proves the system as more effective.

IV. Methodology:

Some of the methods by which information may be retrieved are:

Matlab: It has been used to analyze the color, texture and shapes and objects involved in the videos to enable intelligent content based retrieval process.

2-D Correlation Coefficient technique:It is applied for video shot segmentation in this video retrieval system. Here, corresponding key frames are extracted once the shots boundaries are detected from the videos. First and last frames of each shot are extracted.

HSV color model:Then color feature is extracted from key frames using HSV color model.

Canny edge detection algorithm:Shape feature has been extracted from key frames using canny edge detection algorithm.

Gray level co -occurrence matrix:Textural features are extracted from key frames using gray level co -occurrence matrix. After extracting all of these features from key frames, feature database has been generated from these features.

Kullback- Leibler distance similarity measure:In video retrieval, the system extracts low level features for every key frame of video library and the extracted video features are stored in the feature library. Then, the same features are extracted for a query image and are compared with the features in the feature library with the aid of Kullback- Leibler distance similarity measure. Finally the relevant videos are retrieved from the collection on the basis of Kullback- Leibler distance.

V. Conclusions And Future Work:

The proposed scheme of content based video retrieval system facilitates the segmentation of the elementary shots in the long video sequence proficiently. Subsequently, the extraction of the features for instance edge histogram, color histogram and texture features of the video sequence is performed and the feature library is employed for storage purposes. The kullback liebler distance similarity measure is employed for successful comparison between the features in the feature library and the features of the query clip extracted in a similar manner. The computed kullback liebler distance serves as the basis for the effective retrieval of the similar videos from the video database.

Content based Video retrieval system can further be enhanced by integrating content features like frequency, homogeneity, etc with mining techniques to perform context based retrieval apart from the extraction of low level features. This system works for avi and mpg format, this system can be extended in future to support all video formats. Further, this system could also be extended to retrieve content based online videos also.

REFERENCES

Bing Han, Xinbo Gao, Hongbing Ji, 2005. A shot boundary detection method for news video based on rough-fuzzy sets. Int. J. Inform. Technol., 11: 101-111. www.icis.ntu.edu.sg/scs-ijit/117/117_11.pdf

Gao, X. and X. Tang, 2002. Unsupervised videoshot segmentation and odel-free anchorperson detection for news video story parsing. IEEE Trans. Circuits Syst. Video Technol., 12: 765-776. Doi: 10.1109/TCSVT.2002.800510.

Gao, X. and X. Tang, 2000. Automatic parsing of news video based on cluster analysis. In: Proceedings of 2000 Asia Pacific Conference on Multimedia Technology and Applications, Kaohsiung, Taiwai, China, Dec. 17-19, pp: 17-19. https://dspace.lib.cuhk.edu.hk/handle/2006/5923

Han Bing, Gao Xin-bo, Ji Hong-bing, 2003. An efficient algorithm of gradual transition for shot boundary segmentation. 3rd International Symposium on Multispectral Image Processing and Pattern recognition (MIPPR'03), Beijing, pp: 956-961. http://see.xidian.edu.cn/faculty/hbji /eng.htm

Seung-Hoon Han, Kuk-Jin Yoon, and In So Kweon, 2000. A new technique for shot detection and key frames selection in histogram space. In: 12th Workshop on Image Processing and Image Understanding, pp: 475-479.http://rcv.kaist.ac.kr/publication/file/domestic_conference/18_SeungHoonHan_IPIU2000.pdf

Zhang, H.J., A. Kankanhalli and S.W. Smoliar, 1993. Automatic partitioning of full motion video. ACM Multimedia Syst., 1: 10-28. Doi:10.1007/BF01210504,http://portal.acm.org/citation.cfm?id=173856.17385

Hampapur, A., R. Jain and T.E. Weymouth, 1995. Production model based digital video segmentation. Multimedia Tools Appl., 1: 9-46. Doi:10.1007/BF01261224

Bilge, G. and A.M. Tekalp, 1998. Content based video abstraction. In: Proceedings of International Conference on Image Proceeding, 4-7: 128-132,Doi: 10.1109/ICIP.1998.727150

Song, S.M., T.H. Kwon, W.M. Kim, H. Kim and B.D. Rhee, 1998. On detection of gradual scene changes for parsing of video data. Proc. SPIE Storage Retrieval Image Video Database, 3312: 404-413. DOI:10.1117/12.298449

John, S., Boreczky and D. Lynn, 1998. A hidden markov model framework for video segmentation using audio an image features. In: Proceedings of IEEE International conference on Acoustics, Speech and Signal Processing, May 12-15, 6: 3741-3744. Doi: 10.1109/ICASSP.1998.679697

Arman, F., A. Hsu and M.Y. Chiu, 1994. Image processing on encoded video sequences. ACM

Multimedia Syst., 1: 211-219. Doi:0.1007/BF01268945

Jain, R., 1988. Dynamic vision. In: 9th International Conference on Pattern Recognition, Nov. 14-17, 1: 226-235. Doi:10.1109/ICPR.1988.28212

Zabih, R., J. Miller and K. Mai, 1995. A feature based algorithm for detecting and classifying scene breaks. In: Proceedings of the 3rd ACM International Conference on Multimedia, pp: 189- 200. http://doi.acm.org/10.1145/217279.215266.

Zhang, H.J., A. Kankanhalli and S.W. Smoliar, 2001. Automatic partitioning of full-motion video. Readings Multimedia Comput. Network.ing, 321-339. http://portal.acm.org/citation.cfm?id=383905