ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 9, September 2015

Efficient Video Feature Extraction and Retrieval on Multimodal Search

S.B. Mahalakshmi[1], Capt. Dr.Santhosh Baboo[2]

Research Scholar, Dept of Computer Science, Periyar University, Salem, India[1]

PG & Research, Dept of Computer Science, D.G.Vaishnav College, Chennai, India[2]

Abstract: The core research in Content-Based Video Retrieval (CBVR) is to automatically parse video and text to identify meaningful composition structure. To facilitate fast and accurate content access to video data, a video document should be segmented into shots and scenes. Recognizing objects sequence from videos is an important Problem of computer vision applications such as web searching, target recognition, surveillance, crime detection etc. To build an Efficient video retrieval system that focuses on features such as color, texture, shape, motion, Visual text embedded in an image .Multimodal is the capacity of system to communicate with one or more input given in search process, It can text, Image, audio embedded in it with signals based, Text embedded in an image. Thus it provides potentially accurate results. These results are used in video searching, video surveillance, text embedded in image .In CBVR, a Video is segmented for its preprocessing Key Frames are used for Feature Extraction. Then clustering and indexing is done with k-means clustering HCT(Hierarchical Clustering Tree),Then Similarity matching is done.

Keywords: Content-Based Video Retrieval, HCT (Hierarchical Clustering Tree), Keyframes, Multimodal.

1.INTRODUCTION

One of the challenges in Content Based Video Retrieval is to robustly extract its features and retrieve them in various forms .Both visual and Non-visual features are taken for retrieval whereas the traditional way of retrieving images has many drawbacks and it is also difficult to search from the long video databases as there exists a Semantic gap between the low level information extracted from the video according to users need meaningfully to higher level.so this paper focuses on the following concepts

1)Video segmenting ,retrieval and feature extraction components are discussed in a hierarchical manner
2)To bridge the gap between low level features and high level semantic content
3)To examine the each task involved in CBVR and its sub processes in detail

Content Based Video Retrieval has a wide range of applications such as Quick browsing of video folders, news event analysis[1],CCTV surveillance videos[2],Educational applications, Remote Instructions. Other applications are ,From 2001 the National Institute of Standards and Technology has been sponsoring the annual Text Retrieval Conference(TREC)Video Retrieval Evaluation (TRECVid)to promote progress in video analysis and retrieval, from 2003 onwards TRECVid has become independent of TREC. TRECVid provided a large collections of video retrieval algorithms [3,4].Many Research took place for retrieving video and image based on color, texture and shape and these were based on similarity measurement. Examples include VisualSEEK [5], Photobook [6], Videoq [7].

This paper is presents our approach to content based video retrieval. It is organized as follows., video retrieval techniques are discussed, and the third section describes about Feature Extraction, our proposed System is presented in the fourth section. The fifth section draws conclusion.

2. TECHNIQUES FOR VIDEO RETRIEVAL

The Process of splitting a long video into smaller units is called Video parsing. So video is divided into basic elements like Key frame where each frame is treated as a static image, shot is defined as a set of contiguous frames taken in continuous camera recording. A set of contagious shots make a scene.

To facilitate fast and accurate content access to video data, we should segment a video document into frames, shots and scenes

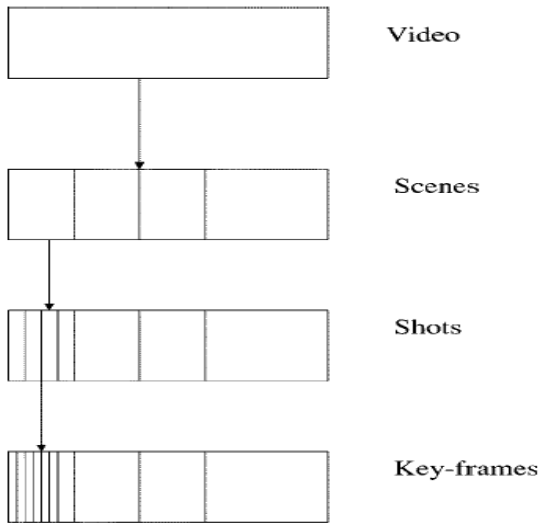

Fig1.Video Parsing

DOI 10.17148/IJARCCE.2015.4976

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 9, September 2015

After the extraction of key frames are classified into six categories: sequential comparison based, global comparison-based, reference frame based, clustering-based, curve simplification-based, and object /event-based.

2.1 Sequential comparison –based approach

The extracted key frames are sequentially compared with the key frame until a frame which is very different from the key frame is obtained. Color histogram is used to find difference between the current frame and the previous key frame.

2.2 Global comparison based approach

The algorithms based on global differences between frames in a shot distribute key frames by minimizing a predefined objective function that depends on the application. It has one of the following four forms.

(a) Even temporal variance:

These algorithms select key frames in a shot such that the shot segments, each of which is represented by a key frame, have equal temporal variance. The objective function can be chosen as the sum of differences between temporal variances of all the segments. The temporal variance in a segment can be approximated by the cumulative change of contents across consecutive frames in the segment or by the difference between the first and last frames in the segment.

(b) Maximum coverage:

These algorithms extract key frames by maximizing their representation coverage, which is the number of frames that the key frames can represent. If the number of key frames is not fixed, then these algorithms minimize the number of key frames subject to a predefined fidelity criterion; alternatively, if the number of key frames is fixed, the algorithms maximize the number of frames that the key frames can represent.

(c) Minimum correlation:

These algorithms extract key frames to minimize the sum of correlations between key frames making key Frames as uncorrelated with each other as possible.

(d) Minimum reconstruction error:

These algorithms extract key frames to minimize the sum of the differences between each frame and its Corresponding predicted frame reconstructed from the set of key frames using interpolation. These algorithms are useful for certain applications, such as animation.

2.3 Reference frame:

These algorithms generate a reference frame and then extract key frames by comparing the frame in the shot with the reference frame .The merit of the reference frame-based algorithm is that they are easy to understand and implement.The limitation of these algorithms is that they depend on the reference frame, if it does not represent the shot, some salient contents in the shot may be missing from the key frames.

2.4 Clustering:

These algorithm cluster frames and choose frames closest to the clusters as the key frames.The merits of the Clustering based algorithms are that they can use generic clustering algorithms and the global characteristics of Video can be reflected in the extracted key frames.The limitations of these algorithms are that they are dependent on the clustering results but successful acquisition of semantic meaningful clusters is very difficult for large data,the sequential nature of video cannot be naturally utilized

2.5 Curve Simplification:

These algorithms represent each frame in a shot as a point in the feature space, they are linked in the sequential order to form a trajectory curve and then searched to find a set of points which best represent the shape of the curve. The merit of the curve simplification-based algorithms is that the sequential information is kept during the key frame extraction. Their limitations is that optimization of the best representation of the curve has a high computational complexity.

2.6 Objects/Events:

These algorithms jointly consider key frame extraction object/event detection in order to ensure that the key frame contain information about it. The merit of the algorithm is that the extracted key frames are semantically important, reflecting objects are the motion patterns of object. The limitation of these algorithm is detection strongly relies on heuristic rules specified according to the application as a result these are efficient only when the experimental settings are carefully chosen. Because of the subjectivity of the key frame definition there is no uniform evaluation method .In general the error rate and the video compression ratio are used as measures to evaluate the results. Key frames giving low error rates and high compression rates are preferred.

In general a low error rate is associated with low compression rate, it depends on parameters in key frame extraction algorithms

3. EXTRACTION OF FEATURES

Feature Extraction is the required step that allows video shots to be classified, indexed and subsequently retrieved .It is done based on Scale Invariant Feature Transform (SIFT) descriptor from the region. It is based on Color, Texture and Shape.

3.1.Color Based Feature

Color is an active area of image retrieval and most widely used visual features. Color can be global and local

Global descriptor specify the overall color content of the image without spatial distribution of these colors Local descriptor relate to particular image region and in conjunction with geometric properties. The RGB is the most commonly used in display devices. HSI scheme more accurately reflect the human perception of color.

Color based feature include color histogram, a mixture of Gaussian models etc…color features can be extracted from the entire image or from image blocks into which the entire image is partitioned

3.2 Texture Based Feature

Texture measures the look for visual patterns in images

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 9, September 2015

and how they are spatially defined. They are represented by which are then placed in to number of sets, depending on how many textures are detected on the image

Texture are difficult concept to represent, It is defined as what is left after color and shape ,they also served as a Support feature for segmentation based recognition [8]

3.3 Shape Based Feature

It is one of the difficult task as it has to segment objects of interest in the images ,to apply preprocessing of an Object its region of interest in known to be darker than the background ,then the simple thresholding is applied .To isolate the object, it also includes block and edge histogram applied on it[9].

4. PROPOSED SYSTEM

In our proposed Video retrieval system the long video sequences is split it to its basic structural units and feature extraction is done based on the video shot representation and finally using Hierarchical Clustering and Indexing algorithm to find the correct match the results are determined. The best way to use the system is to determine both image and if an text in the image can be given in the search process, Thus it provides potentially accurate results. videos segmented into logical units, Features are extracted-performs clustering& indexing to find the exact image and text embedded in the video ,Retrieval of videos from image and its text content are focused.

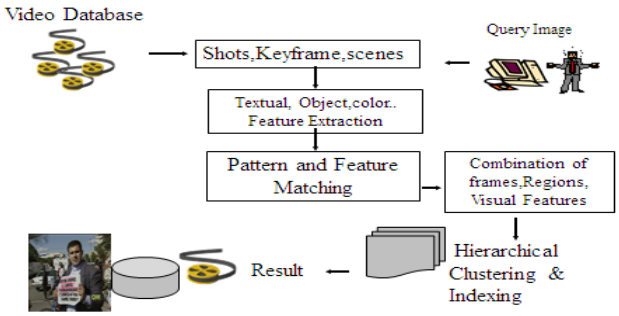

Fig2.General Framework for video search and retrieval

5.CONCLUSION

In this paper as conclusion, it provides a study of video segmentation, feature extraction and clustering for an efficient video retrieval methodology has been described.

This approach integrates both image and text embedded in an image and uses high level concepts and bridges the semantic gap. Thus the selection of these extracted features plays an important role in Content Based Video Retrieval regardless of video attributes being user consideration. Therefore thus the proposed system bridges the semantic gap between the low level visual feature and high level features for better understanding by giving images and text in the video to be retrieved as an input.

REFERENCES

[1]. Y.X.Peng and C.W.Ngo"Hot event detection and summarization by graph modeling and matching"in proc.Int.Conf image video retrieval, Singapore, july 2009

[2]. Yan Yang,Brain C Lovell, Farhad Dadgostar "Content-Based Video Retrieval (CBVR)system for CCTV Surveillance videos"Digital image computing :Techniques and application 2009

[3]. A.F.Smeaton, P.Over and A.R.Doherty "Video Shot Boundary Detection :Seven years of TRECVid activity" Computer vision image understanding vol 114 on.4.pp 411-418 2010

[4]. A.F.Smeaton ,P.Over and W.Kraaij "High level Feature Detection from video in TRECVid-A 5 years retrospective of achievement" Multimedia Content analysis signal and communication technology, Berlin, Germany, Springer 2009,pp 151-174

[5]. J.R.Smith S-F,Chang "VisualSEEK:A fully automated Content based Image Query System"ACM Multimedia Conferene, Boston, MA, November 1996

[6]. A.Pentland,R.W.Picard,S.Sclaroff,"Photobook:Content Based Manipulation of Image Databases",Int.J.Computer vision,18(3),pp.233-254

[7]. S.F.Chang,W.Chen,H.Meng,H.Sundaram,D.Zhong,"A Fully Automated Content Based Video Search Engine Supporting Spatio-Temporal Queries"IEEE Transaction on Circuits and systems for video Technology,vol.8.no.5

[8]. W.Y.Ma and B..Manjunath "Edge flow:A Framework of Boundary detection and image segmentation ".Proc.Computer vision &pattern recognition,pp 744-749

[9]. T.N.Shanmugam and Priya Rajendran "An Enhance Content Based Video Retrieval system based on Query clip" International journal of Research and reviews in applied science,ISSN:2076-743x,Volume1,issue3.