Scientific Journal of Impact Factor (SJIF) : 3.134

ISSN (Print) : 2348-6406
ISSN (Online): 2348-4470

International Journal of Advance Engineering and Research Development

Video Content Analysis

Yamini Gupta[1],Vatsal Shah[2]

[1]Student, Information Technology, BVM Engineering College, yamini_94@yahoo.co.in
[2]Assistant Professor, BVM Engineering College, vatsal.shah@bvmengineering.ac.in

Abstract *---This work aims at developing a better understanding towards the process of Video Content Analysis. Video analysis is basically how the computer interprets any video. This process of analyzing is completed in 4 steps namely, feature extraction, Structural Analysis, Clustering and Indexing, Browsing and Retrieval. In this paper, we have included various algorithms to understand how each of the above steps is applied in analyzing of any of the video. For each process, a number of algorithms are used and each algorithm has its own significance.*

Keywords*-Video, Key frames, shots, temporal boundaries, parsing, metadata, clustering*

I. INTRODUCTION

Video Content Analysis, also known as Intelligent Video Analytics (IVA) is the process of automatic analysis of CCTV images to create useful information about the content. Although humans are readily able to interpret digital video, developing algorithms for the computer to analyze videos is now in active research. The main goal of video analytics is scene understanding rather than just detecting motion. Video analytics considers motion as an object, understands the context around the object.VCA can be successfully used in a variety of applications: External and internal intruder detection, monitoring of plant or buildings for health and safety, people counting, automatic traffic event and incident detection, safety enhancements for public areas, smoke and fire detection.[1]

II. PROCESS OF ANALYSIS

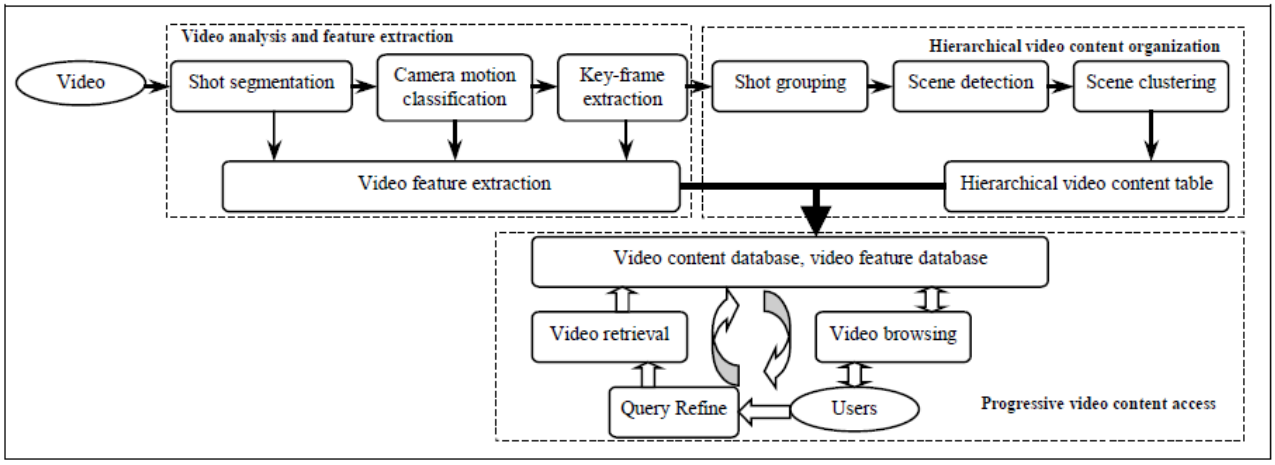

Figure 1. System flow for insight video [2]

A typical scheme of video-content analysis and indexing involves four primary processes:

1. Feature extraction,
2. Structure analysis,
3. Clustering And Indexing
4. Retrieval And Browsing

A. Feature Extraction:

Although visual content is a major source of information in a video program, an effective strategy in video-content analysis is to use attributes extractable from multimedia sources. Combined and cooperative analysis of video, audio and text components would be far more effective in characterizing video program for both consumer and professional applications.

B. Structure analysis:

Structure parsing is the next step and is the process of extracting temporal structural information of video sequences or programs. It is the process of organizing video data according to their temporal structures and thus builds an index table. This step mainly detects the main segments of the video. These temporal compositions are called Bricks. Ideally, the bricks should be categorized in a hierarchy. The top level consists of sequences or stories, which are composed of sets of scenes. Scenes are further partitioned into shots. Each shot contains a sequence of frames recorded contiguously and representing a continuous action in time or space. [3] Through structure analysis, process of video abstraction is performed. Video abstraction is a method of creating a related video having all the important segments just identified.

C. Clustering and Indexing:

The structural and content attributes extracted in above 3 processes are often referred to as metadata.
Based on these attributes, a clustering process can be performed that classifies sequences or shots into different visual categories to prepare a table of contents or video indices.
Clustering is nothing but organising the contents of video. A preview of the video content can simply be generated by showing a subset of clusters or the representative frames of each cluster. Similarly, retrieval can be performed in an efficient way since similar shots are indexed under the same cluster. [3]

D. Retrieval and Browsing:

Retrieval and browsing are performed efficiently by indexing all the similar shots into same cluster. Video retrieval techniques are almost similar to image retrieval techniques. Examples include first selecting key-frames from shots and then extracting image features such as colour and texture features from those key frames for indexing and retrieval. [3]

III. ALGORITHMS

A. Algorithms for Feature Extraction

First works proposed different method for detecting scene change detection in uncompressed data.
For example, Nagasaka et al and Zhang et al proposed three different algorithms:
Pairwise pixel comparison, a pixel intensity based algorithm; Likelihood ratio, another pixel intensity based algorithm using statistical measures and picture block division; and diagonally Histogram comparison, a colour histograms based algorithm.[4]
Lee et al. proposed an edge based method that worked with compressed data by obtaining edge orientation and strength directly from the MPEG video stream. This work presents a new video feature extraction technique based on the Generalized Hough Transform (GHT).[4] This technique provides a method to compare different frames composed by extracting different attributes from a video stream in order to obtain a basis for scene cut detection algorithm.
It also allows the rotation, scale and displacement parameters between two frames, making it possible to develop algorithms for gradual transition detection.

To illustrate the possibilities of this technique, **a scene cut detection algorithm** is proposed. This algorithm presents four key characteristics:

1. Processes the video data in the MPEG compressed domain, using estimated values of DC coefficients. The use of estimated DC values makes the algorithm much faster because there is no need to perform the complete decompression process of a frame.
2. The use of two series of similar values obtained from the video makes the cut detection algorithm more long lasting and efficient.
3. The use of an improved thresholding technique, which includes specific analysis to find cuts especially in, transitions from static to dynamic scenes or vice versa.
4. The improved thresholding technique results in an effective performance on different kinds of videos without the need of any human intervention.

B. Algorithms for Structure Analysis

- **Shot Boundary Detection**

A shot may be defined as a continuous sequence of frames generated by a single non-stop camera operation.[8]
Video Shot boundaries are defined as cut in which the transition between successive shots is abrupt and gradual transitions which include dissolve, fade in, fade out etc.
Steps for shot boundary detection includes with the first step as extracting visual features from each frame, then measuring similarities between frames using the extracted features, and, finally, detecting shot boundaries between frames that are dissimilar.
Frame transition parameters and frame estimation errors based on global and local features are used for boundary detection and classification. [5]
Shot boundary detection applications are classified into two types.

1) Threshold based approach detects shot boundaries by comparing the measured pair-wise similarities between frames with a predefined threshold.
2) Statistical learning-based approach detects shot boundary in which frames are differentiated as shot change or no shot change depending on the features that they include.

- **Key Frame Extraction**

There is a large amount of repetitions among the frames in the same shot; therefore, certain frames that best display the shot contents are selected as key frames to effectively represent the shot.
Approaches to extract 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 [5]

In Sequential comparison-based approach, previously extracted key frame are sequentially compared with each and every key frame until a frame which is very different from the key frame is obtained.
Global comparison-based approach is based on global differences between frames in a shot.
Reference frame- based approach generates a reference frame and then extracts key frames by comparing the frames in the shot with the reference frame.

- **Scene Segmentation**

Scene segmentation is also known as story unit segmentation. A scene is a group of sequential shots that are connected with a certain subject. Scenes have higher level semantics than shots. [5]Scene segmentation approaches can be classified into three categories:

- key frame based,
- visual information integration-based, and
- Background-based.

Key Frame-Based Approach: represents each video shot by a collection of key frames from which features are extracted.
Audio and Vision Integration-Based Approach: Selects a shot boundary where the visual and audio contents change simultaneously as a scene boundary. A time-constrained nearest neighbour algorithm is used to determine the correspondences between these two sets of scenes. [5] However, it is difficult to find the relation between the audio part and the video part.
Background-Based Approach: This approach segments scenes based on the assumption that shots falling under the same scene mostly have similar backgrounds.

C. Algorithm For Clustering And Indexing

Clustering algorithms can be grouped into two categories: partitional and hierarchical.
Hanjalic & Zhang have introduced a partitional clustering of video data by utilizing the colour features of selected key frames. [6] A two-level hierarchical clustering algorithm uses both colour and motion features. The top level is clustered by colour features while the bottom level is clustered by motion features. For example, in a sports video, the top level contains various clusters including wide-angle, medium-angle and close-up shots of players from different teams. Each cluster can refer to a sport event. For instance, the wide-angle shots of a basketball video usually correspond to full court advances, while the wide-angle shots of a soccer video normally correspond to bird view scenes. [6] The shots inside each cluster can be further classified on the basis their motion intensity. E.g., the sub-cluster of a close-up shot can belong either to “players running across the soccer field”, or “players standing on the field”.
The algorithm is implemented in a top-down fashion, where colour features are utilized at the top level, while motion features are used at the bottom level. At the top level, the colour feature space is partitioned to *kc* clusters. At the bottom level, each *kc* cluster is further partitioned into *km* clusters. [6]

D. Retrieval Algorithms

Given a query represented by features such as colour or motion, a retrieval system returns a set of items arranged in ascending order according to their closest relation to the query.
This is normally referred to as a *k* nearest neighbour (KNN) search problem. By coupling clustering issues with retrieval problems, the clustering structure, on one hand, inherently provides an indexing scheme for retrieval, while on the other hand, intuitively speed up the retrieval time. [6]
Suppose a query is located at the boundary of two classes, cluster-free retrieval will include items from both classes. Whereas, clustered-based retrieval compares the distance between the query and each cluster centre, and places the items whose cluster centroid is nearest to the query at the top of the index.

- **Cluster-based Retrieval**

In a hierarchical clustering structure, at the top level represents the colour characteristics of a cluster is displayed, while at the bottom level the motion features of a cluster are represented. During retrieval, clusters at the top are compared with colour related query and those at the bottoms level are compared with motion features.
A cluster with the nearest centroid is first located. Then, its sub-clusters in the bottom level are distinguished with the query. The segments in one of these sub-clusters whose centroid is the nearest to the motion attribute of the query, are arranged in ascending order of their relationship with the query, and put accordingly at the top of a ranked list.
The retrieval is processed in a depth-first search- like manner, meaning that after all sub-clusters of the most similar cluster are sorted, the next similar cluster is handled in the same way. This process is repeated until the few most similar or all clusters are visited. [6]

IV. CONCLUSION

In this paper, we have included the explanation of various steps in the procedure of video content analysis and how it is applied in various types of videos. We have defined how feature extraction is performed by cooperative methods of extracting colour, audio and text.IN the next steps i.e., structure analysing, the video is divided into smaller segments and is arranged in a hierarchy. Finally, all these segmented videos are clustered together according to be indexed in some specific order. Users browse through by this table of index and retrieve the selected video by putting forward their query video. By researching on this topic, we would like to conclude that if the researches keep going on in such a fast pace, then the time won't be far away when analysation of any video by a computer, without the aid of humans would be a child's play.

REFERENCES

[1] British Security Industry Association ,"An introduction to video content analysis industry guide", Form No.262, Issue 1, June 2009

[2] Xingquan Zhu, Ahmed K. Elmagarmid, Xiangyang Xue, Lide Wu, Ann Christine Catlin, " InsightVideo: Towards hierarchical video content organization for efficient browsing, summarization and retrieval", Multimedia, IEEE Transactions on (Volume:7 , Issue: 4), Page(s):648 - 666 ,ISSN :1520-9210 , Aug. 2005

[3] Nevenka Dimitrova, Hong-Jiang Zhang, Behzad Shahraray, Ibrahim Sezan, Thomas Huang, Avideh Zakhor, "Applications of video-content analysis and retrieval", Page(s):42 – 55, ISSN :1070-986X , Jul-Sep 2002

[4] Edmundo Saez, Jose M. Gonzalez, Jose M. Palomares, Jose I. Benavides and Nicolas Guil "New edge-based feature extraction algorithm for video Segmentation", SPIE Proceedings Vol. 5022: Image and Video Communications and Processing 2003, May 2003

[5] Ganesh. I. Rathod, Dipali. A. Nikam,"An Algorithm for Shot Boundary Detection and Key Frame Extraction Using Histogram Difference", International Journal of Emerging Technology and Advanced Engineering, ISSN 2250-2459, Volume 3, Issue 8, August 2013

6] Chong-Wah Ngo, Ting-Chuen Pong, Hong-Jiang Zhang," On Clustering and Retrieval of Video Shots", MULTIMEDIA '01 Proceedings of the ninth ACM international conference on Multimedia, Pages 51-60, ISBN:1-58113-394-4,October 2011

