International Journal of Research in advance Engineering,
Volume -1 Isuue-2, Feb-2015, Available at www.knowledgecuddle.com

A Survey on Video Content Analysis

Yamini Gupta
IT Department
Gujarat Technological University
yamini_94@yahoo.co.in
Gujarat

***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 sports video. For example, in cricket we first extract the various features by attacking on the most often occurring color, i.e. ground color. Then, we divide the video according to various high points e.g., sixes and fours. Finally we cluster these divisions and arrange them in a table of contents for users to browse and retrieve the video part they require.**

***Keywords*: Video Content analysis, Key frames, shots, temporal boundaries, parsing, metadata clustering**

I.INTRODUCTION

Video Content Analysis, also known as Intelligent Video Analytics (IVA) is the name given to the 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 perform the same task is now in active research. The main goal of video analytics is scene understanding, which differs from motion detection. In addition to detecting motion, analytics qualifies the motion as an object, understands the context around the object, and is able to track the object through the scene.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.

II. PROCESS OF ANALYSIS

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:

Video structure parsing is the next step in overall video-content analysis and is the process of extracting temporal structural information of video sequences or programs. This process lets us organize video data according to their temporal structures and relations and thus build table of contents. It involves detecting

temporal boundaries and identifying meaningful segments of video. These temporal compositions are called Bricks[i]. 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. Video abstraction is the process of creating a presentation of visual information about a landscape or the structure of video, which should be much shorter than the original video. We need to extract a subset of video data from the original video such as key frames or highlights as entries for shots, scenes, or stories. [1]

C. Clustering and Indexing:

The structural and content attributes extracted in feature extraction, video parsing, and abstraction processes, or the attributes are often referred to as metadata.
Based on these attributes, we can build video indices and the table of contents through a clustering process that classifies sequences or shots into different visual categories or an indexing structure.
Clustering is a natural solution to abbreviate and organize the content of a 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. [2]

D. Retrieval and Browsing:

Retrieval and browsing can be performed in an efficient way since similar shots are indexed under the same cluster. Video retrieval techniques, to date, are mostly extended directly or indirectly from 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. SPORTS VIDEO ANALYSIS

We, here as example, are taking sports video to explain the concepts and algorithms explained above. Identifying the different types of view is the first and the necessary step in any sports video analysis. Classification processes are divided into two parts:

- Features extraction and
- Classification of these features.

We here are applying the classification algorithm. We can identify two main kinds of features.

1. The first class relies on colour-based features. Considering sport videos, a global view is actually characterized by a large region of homogeneous colour (colour of the play field). In a baseball video, a pitching scene is detected by computing the difference between a candidate key-frame and a representative pitching image. The representative pitching image is manually extracted from the considered video.

2. The second class consists of motion-based descriptors. For example, the variation and persistence of the estimated camera motion, as well as the number of intra-coded macro blocks in a MPEG video of basketball is used to classify wide-angle and close-up shots. To efficiently capture the frame contents, some approaches mix several descriptors. In the context of soccer video, the grass pixels ratio and motion intensity in a frame are relevant features to categorize each views. The colour-based classification should be re-enforced using motion-based features. On one hand, a global view must capture at each time the main part of the court. On the other hand, in close-up views, the camera is generally tracking the player.

The second step is Structural Analysis in Video analysis. We take the example of game of Tennis to understand this step. We define different structural elements in a tennis video game: first missed serve and rally, rally, replay and break. The construction of elements takes domain-knowledge derived from tennis syntax into account as follows:

- In a broadcast video, the producers notify the viewers that a replay is being displayed by inserting special transitions.
- A first missed serve is a global view of short duration following by close-up views of short duration too (as the players do not have to change their positions) and following by another global view.
- A break is characterized by an important succession of close-up views, public views and advertisements. This set of consecutive shots has a particular long duration. It appears when players change ends, generally every two games.


The type of view, the shot duration and their temporal relations are of first importance in the discrimination of the structural elements. The activity of a shot is not a discriminatory feature: as it represents an average quantity of motion over a shot, it has quite the same value for one type of view. Consequently, it cannot help to distinguish one global view from another.

After video parsing is completed, we cluster the segments and start indexing. Video classification and segmentation are fundamental steps for efficiently searching and browsing video content. When the indexing of videos is restricted to given category, domain specific knowledge about the processed content facilitates the recovery of higher-level indexing information. One domain-specific application is the detection and recognition of highlights in sport videos. Sport video analysis is motivated by the growing amount of archived sport video material. Broadcasters need detailed annotation of video contents to select relevant excerpts to be edited for summaries or magazines. At present, this logging task is performed manually by librarians. [4]
Domain-specific video indexing can be divided into 3 main research areas:

- genre classification,
- content analysis,
- Structure analysis.

The goal of genre classification is to automatically classify TV broadcast into predetermined genres like commercials, news, sport, etc. Content analysis usually aims at detecting specific events in a video. Domain knowledge and properties of low level features are exploited for mapping low-level information extracted from video data to high-level concepts. Finally, structure analysis aims at highlighting the “table-of-contents” of videos within a given genre. The table-of-contents is obtained by finding the discontinuities of semantics in the video. It involves detecting the temporal boundaries of the coherent segments and identifying all segments of video according to predefined semantic categories. As not all of the content of a video is of interest, separating the process of structure parsing and event detection may enhance the indexing process, by first extracting the interesting segments, and then applying content analysis on them. [5]

Finally, the user browses and retrieves the video after indexing has finished. Each low-level event detected (e.g., racket hit in tennis, ball pot in snooker, goal in soccer) can yield an automated index for a sports game. Summaries can either be condensed or selective representations of a game. The condensed summary attempts to give a well-balanced summarized view of the whole video content. In contrast, the selective approach targets sequences that convey by themselves the highest interest. A selective summary of a soccer match will contain only the goals for instance, while for a cricket game it would contain only the wickets. The

creation of a condensed summary requires two stages: an off-line learning stage and a supervised recognition step. The creation of a selective summary mainly relies on an unsupervised detection of events. [6]

The identification of semantic-level events enables direct retrieval of those events. However, by constructing a kind of domain-specific language for the game description, it is possible to give the user more flexible access to the system, which would possibly allow the retrieval of events that were not specifically marked.

IV. FUTURE WORK

Many issues are still open and deserve further research, especially in the following areas:

1. Most current video indexing approaches depend heavily on prior domain knowledge. This limits their extensibility to new domains. The elimination of the dependence on domain knowledge is a future research problem.
2. Fast video search using hierarchical indices are all interesting research questions.
3. Video indexing and retrieval in the cloud computing environment, where the individual videos to be searched and the dataset of videos are both changing dynamically, will form a new and flourishing research direction in video retrieval in the very near future.
4. Video affective semantics describe human psycho-logical feelings such as romance, pleasure, violence, sadness, and anger.

V.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 the field of sports. We have defined how feature extraction is based on 2 classes, colour based and motion based. We have also defined four basic structural elements of a tennis game on which the structure analysis is based. These four elements are also interesting while they infer the following highlights in a further event detection process: first missed serve, rally, and replay. The 3rd step of indexing is defined by categorising it into 3 parts: genre classification, content analysis, structure analysis. Finally, the paper explains how each of the above steps forms the basis for constructing an index of table and presenting it to the user. 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 computers would complete the analysation of a video without the aid of any person.

REFERENCES

[1]. H.J. Zhang et al., "Video Parsing, Retrieval, and Browsing: An Integrated and Content-Based Solution," *Proc. Third Int'l Conf. Multimedia* (ACM Multimedia 95), ACM Press, New York, 1995, pp.15-24.

[2]. B. Shahraray, "Scene Change Detection and Content-Based Sampling of Video Sequences," *IS&T/SPIE Symp. Digital Video Compression: Algorithm and Technologies*, SPIE Press,

[3]. H.J. Zhang, J.Y.A. Wang, and Y. Altunbasak, "Content-Based Video Retrieval and Compression: A Unified Solution," *Proc. IEEE Int'l Conf. Image* model", *Proc. of the IEEE Int'l Conference on Multimedia and Expo*, pp. 1551-1554, 2000.

[4]. G. Iyengar and A. B. Lipman. Models for automatic classification of video sequences. In *Proc. SPIE Storage and Retrieval for Image and Video Databases VI*, pages 3312–3334, 1998.

[5]. N. Rea, R. Dahyot, and A. Kokaram, "Modeling high level structure in sports with motion driven HMMs," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, May 2004, vol. 3, pp. 621–624.