ISSN: 2321-8169
2719 – 2723

Detecting Object with Content Video Tracking

Dimple Bohra
Assistant Professor, Computer Engineering, VESIT, Chembur, Mumbai.
dimple.bohra@ves.ac.in

Gaurav Vaswani
Student, Third Year Degree, Computer Engineering, VESIT, Chembur, Mumbai.
gauravvaswani@gmail.com

Priyanka Punjabi
Student, Third Year Degree, Computer Engineering, VESIT, Chembur, Mumbai.
priyankapunjabi94@gmail.com

Ajay Chotrani
Student, Third Year Degree, Computer Engineering, VESIT, Chembur, Mumbai.
chotrani.ajay@gmail.com

***Abstract*--**The advances in the data set in the form of video are available on the internet. The paper proposes the system with the retrieval and detection of images from the video. Detecting particular objects in video is an important step for understanding the visual imaginary semantically. In content based retrieval, the ability to detect people any specified object such as animals , cycles and automobiles gives the option of advanced queries such as "Find a video clip which contains a crowded area or a fast moving car.", or we detect the various contents present in the video for which the application is trained for. The application can be designed and trained for the same from the dataset.

Keywords*--Content based video retrieval, semantic information, and colour histogram, segmentation*

*****

INTRODUCTION

Content based video retrieval may be defined as an approach in which the videos are retrieved from the large database based upon their visual contents [1][6][7]. Content based video retrieval is desirable because most image or video search engines rely purely on metadata and this produces a lot of wrong results. Also these keywords are annotated manually and are completely based upon human perception, so this procedure will not be sufficient enough to capture every keyword that describes the image or video. Thus a system that can filter images on their content would provide better indexing and return more accurate results. The content of video or image are colour, shape, texture etc. Video content can be grouped into two levels: low level visual features and high-level semantic content.

CONTENT BASED VIDEO RETRIEVAL

As video data is very complex; understanding of its unique characteristics is essential to develop techniques for managing it. There are some significant characteristics that distinguish video from other classes of data. [2]

1. Video is stored as binary; therefore, in contrast to alphanumeric data, video has higher resolution, larger data volume, larger set of data that can be originated, higher interpretation ambiguity, and needs more interpretation efforts.
2. Video has spatial and temporal dimension, whereas text is only non-spatial static and image is spatial static. Moreover, video semantic is unstructured and generally contains complex relationships.

Figure 1 shows the flow diagram of the working of CBVR.

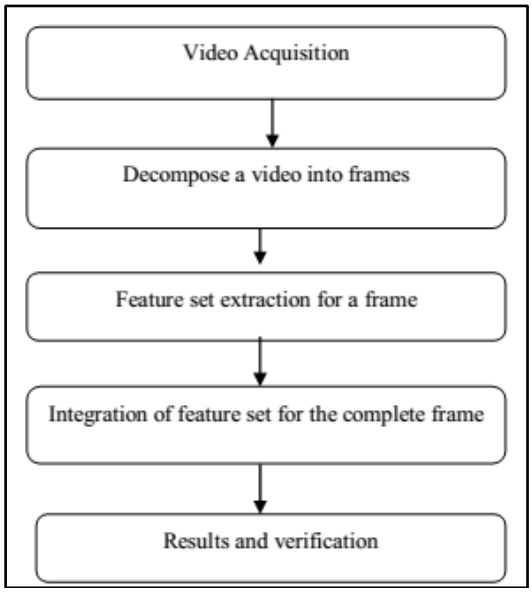

Figure 1: Flow Diagram of CBVR

The videos are segmented into frames and then how it is done is shown in the section of video segmentation.

COMPONENTS OF CBVR

The main components of a video document are semantic content and audio visual presentation. Semantic content is the idea, knowledge, story, message or entertainment conveyed by the video data.

 ISSN: 2321-8169
2719 – 2723

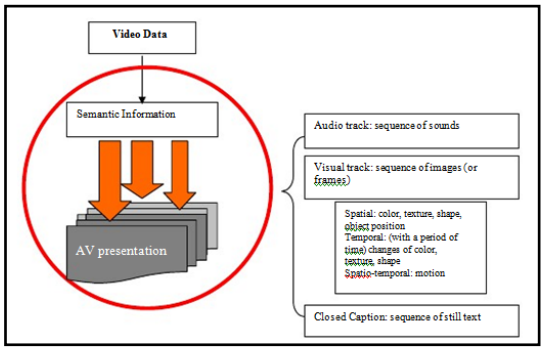

Figure 2: Video Data Components

It is the most complex part of video data as the semantic information of video can be expressed either implicitly or explicitly. Figure 2 depicts a high-level concept on the contents of a video data.

Figure 3 illustrates a very simple example on how semantic contents can be conveyed through the spatio-temporal AV presentation of video data. Since shots are merely depicted as a sequence of video frames in the second diagram, thus the sequence of three shots convey semantic information, which indicates that a car has moved from right to left (spatial-temporal information), and a tower is located in the middle of the background (spatial-static information).[8]

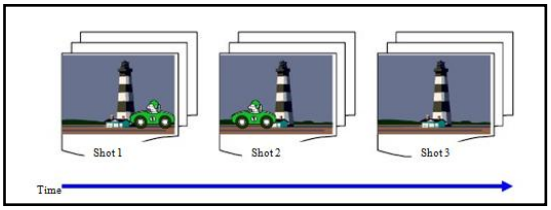

Figure 3: Semantic information through the spatio-temporal of video data

It is important to note that the AV components of video data are not always equally important in conveying the semantic content. Instead, it depends on the purpose and use of the video data. For example, in analysing the game strategy and techniques during a soccer match, the information about the motion and position of the players are most important. Along with this, depending on how the video was produced, many different AV features can represent the same semantic content and vice versa, the same AV features can represent different semantic contents due to the subjectivity of the annotators. Hence, similar to the human perception of video document, the semantic content in video will be more accurately interpreted when more channels are perceived.

A general structure of CBVR components is depicted in Figure 4 and can be described as follows.

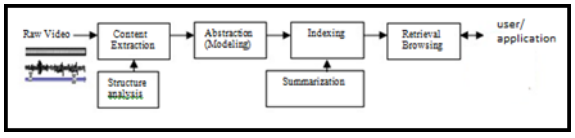

Figure 4 Components of Content based Video Retrieval Architecture

User/application requirements determine the retrieval and browsing. The success of retrieval depends on the completeness and effectiveness of the indexes. Indexing techniques are determined by the extractable information through automatic or semi-automatic content extraction. Since video contains rich and multidimensional information, it needs to be modelled and summarized to get the most compact and effective representation of video data. Prior to video content analysis, the structure of a video sequence needs to be analysed and separated into different layers such as shot and scene. Since the design of each component is affected by other components, it is generally difficult to discuss each component separately. For example, if the retrieval is based on high-level semantic features such as specific sport highlight segments like goal, the indexing can be based on the hierarchy of summarized (key) events. Each event can be abstracted using the face of the actors who participate during that event whereas the event itself can be described using some statistical measures like excitement ratio (e.g. the higher excitement, the more significant an event should be). As a result, the content extraction process aims to automatically identify and classify the event(s) that is contained within each play- break segment.

VIDEO SEGMENTATION

In general, indexing could be performed on the whole video stream but it would be too coarse [11]. On the other hand, if the indexing is based on each frame in the clip, it would be too dense as a frame often does not contain any important information. Researchers have commonly indexed on a group of sequential frames with similar characteristics. Figure 5 shows how videos are divided into frames.

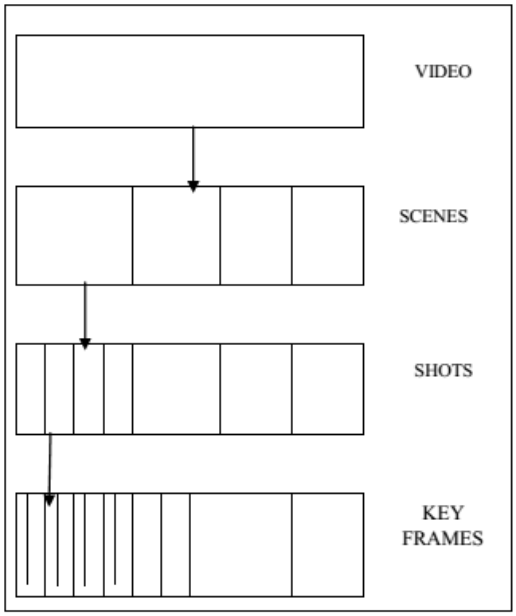

Figure 5: Segmenting videos into Frames

SHOT BOUNDARY DETECTION [8]

Shot is a sequence of video frames which have similar characteristics. Shots extraction requires the

 ISSN: 2321-8169
2719 – 2723

computation of an appropriate metric (algorithm) to characterize the change of video content between two frames and a threshold to determine whether the change is important enough to be defined as a shot boundary.

Three main methods which can be used for shot boundaries detection are described below:

- Pixel-wise frame difference.
- Histogram comparison.
- Audio assisted.

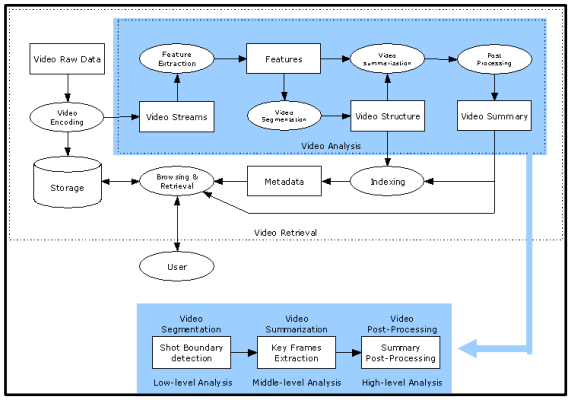

Figure 6: Shot Boundary Detection

Pixel-wise frame difference technique detects shot boundaries by measuring a qualitative change between two frames by simply comparing the spatial corresponding pixels in the two frames and determining the amount of the pixels that have changed. Thus it is also called pair-wise pixel comparison algorithm shown in Figure 7.

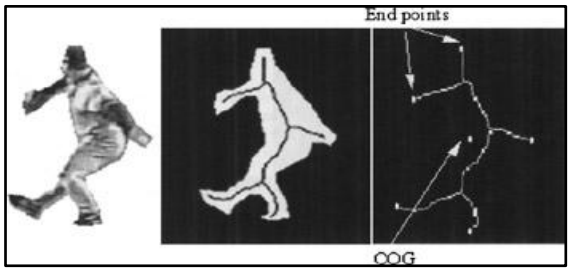

Figure 7: Pixelating the Image from the Video

While pixel-wise comparison is easy to be done, it is too sensitive against noise, illumination changes and camera motion. Alternatively, color histogram can be used due to its effectiveness in characterizing the global distribution of an image without knowing the component objects that make up the image. Color histogram shows statistically the intensities of the three-color channels in the image, such RGB (Red Green Blue) or HSV (Hue Saturation Value).

Color histogram comparison should also be less sensitive to object motion than pixel difference technique [1]. Histogram is used for shot boundary detection by assuming that two frames which have unchanging backgrounds and objects will show little difference in the histograms. This method can be summarized as follows: histogram of a frame is computed by measuring probability distributions of pixel values in the entire image on a frame; therefore each frame will produce a distinct histogram. To identify shot boundaries, the histogram difference between a frame and its consecutive frames is compared against a threshold. If the difference is bigger than the threshold, a shot boundary is found.

SCENE DETECTION

Although shot-based indexing can be fully automated, the major problem is the lack of semantic information [12]. As mentioned, a shot is only a group of sequential frames with similar characteristics; therefore, it does not actually correspond directly to semantic content. For example, the semantic content often does not change during shot boundary. To overcome this limitation, scenes need to be extracted as a sequence of shots which represent and can be described by a semantic content description. Scene detection is generally more difficult than shots detection due to a need for understanding the video contents. One approach is by measuring the semantic correlation of consecutive shots based on dominant colour grouping and tracking. For example, a shot grouping method called expanding window has been designed to cluster correlated consecutive shots into one scene. Similarly, scene can be detected via continuous coherence in which related shots are grouped into scenes which are defined as a single dramatic event taken by a small number of related cameras. Alternatively, a scene can be formed by grouping a sequence of shots which depict a particular object activity or event. Figure 8 shows frame retrieval.

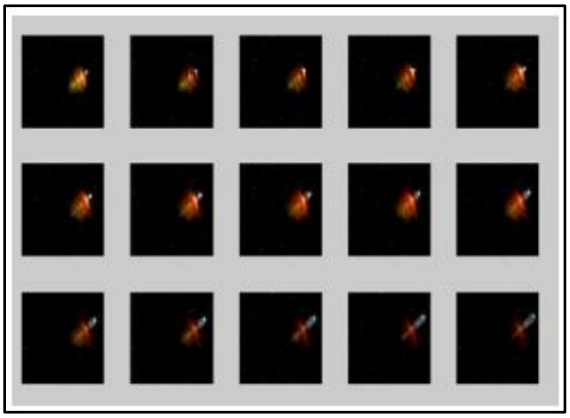

Figure 8: Frame retrieval

IMAGE ANALYSIS [2]

General image or visual features include colour, texture, and shape. The focus of this section is to review the very useful image features that can be used for video content analysis, namely,

1) Colour feature,
2) Shape feature and
3) Texture feature.

 ISSN: 2321-8169
2719 – 2723

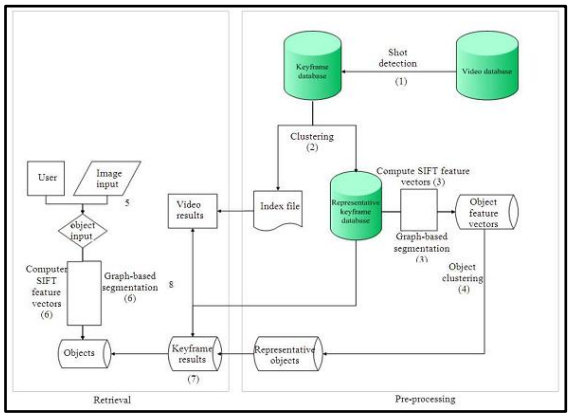

Figure 9: Image Analysis

Colour Features:

Colour is the most intensively used feature for image content management due to its robustness to complex background (occlusion), scaling (image size), orientation, and perspective. Colour histogram and colour moments are some of the fundamental features of colour. The colour feature detection is shown in Figure 10.

Figure 10: Colour Feature Detection

Colour Histogram:

Colour histogram is the most common feature representation as it can characterize the global distribution of an image effectively without knowing the components objects that made up the image [10]. It shows the intensities of the three-color channels in the image statistically. For example, Figure 11 depicts the histogram of two typical views in soccer videos; where the frame with large playing field shows dominant intensities in colour index of 0.4 to 0.6 while the frame with player close-up has a more equal distribution of colour intensities.

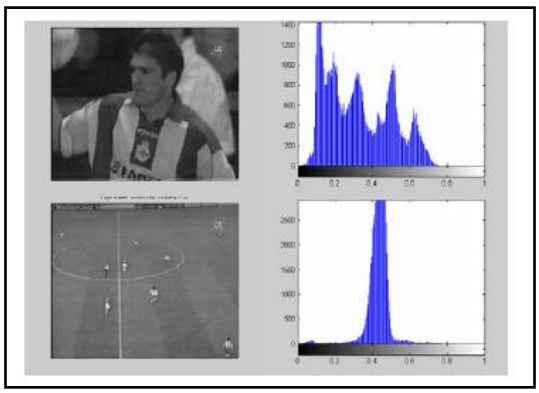

Figure 11: Colour Histogram

Colour Moments:

Colour moments of an image are chosen due to their effectiveness and simplicity. Any colour distribution can be represented by its moments. In particular, first (moment) order captures mean colour, while second and third order capture colour variance and skew-ness respectively.

Shape features:

Shape is another important image feature as any good shape representation should be unique, robust against translation, scaling and rotation. To extract shape features, image segmentation techniques are required to segment an image into its individual objects. In general, shape representations can be categorized into boundary- (outer shape) and region- (entire shape) based. The Figure 12 shows the shape feature detection.

Figure 12: Shape Feature Detection

Texture Features

Texture is an important image feature as it describes visual patterns which are homogeneous and not produced from single colour or intensity. These visual patterns distinguish structural arrangement of surfaces from the surrounding environment; therefore texture is a natural property of all object surfaces, such as clouds, woods and bricks. However, it is generally difficult to describe texture and its perception can be subjective to a certain degree. For this reason, texture specification techniques often need to mimic human perception on texture. Figure 13 shows the texture feature detection in the images from the videos.

Figure 13: Texture feature Detection

Texture can be described by six features: coarseness, contrast, directionality, regularity, likeliness, and roughness.

CONCLUSION

The content based video retrieval system can be best applied to any sports where the team wants to track the record and to improve the game of a particular player.

The CBVR can be improved and combined with histogram of oriented gradient for tacking the humans and also detecting various objects like cycles, or animals like cat or dog.

This can be implemented in a single application with a multiple detectors becomes a difficult approach when comes to the application level of the approach. Detecting humans and its application implementation for pedestrian detection and setting up an alarm as the pedestrian is detected. The various classifiers with scale invariant feature transform (SIFT) can be used to improve the detection of the object in the content based video retrieval system.

REFERENCES

[1]. Gaurav Jaswal, Amit Kaul and Rajan Parmar (2012). Content Based Image Retrieval using Color space Approaches.mInternational Journal of Engineering and Advanced Technology (IJEAT) 2(1), ISSN: 2249

– 8958.
[2]. Gulshan saluja, Ankit rokde and Richa maru (2012). Layered filtering technique for content based video retrieval. IEEE 978-1-4673-1938-6/12.
[3]. Hatice Cinar Akakin and MetinGurcan N (2012). Content-Based Microscopic Image Retrieval System for Multi-Image Queries. IEEE Transactions on Information Technology in Biomedicine 16(4).
[4]. Kui Wu & Kim-Hui Yap (2007). Content-based image retrieval using fuzzy perceptual feedback.Multimedia Tools and Applications 32 235–251. DOI 10.1007/s11042-006-0050-2
[5]. Navdeep and Mandeep Singh. Content (color) based image retrieval using RGB component Analysis. 1st National Conference on Information Technology and Cyber Security 1 171-174/ITCS13/33.
[6]. Nidhi Singhai and Shishir Shandilya (2010). A Survey On: Content Based Image Retrieval Systems. International Journal of Computer Applications (0975 – 8887) 4(2).
[7]. Padamkala S and Anandhamala GS (2011). An effective content based video retrieval utilizing color, texture and optical key frame features. International Conference on Image Information Processing. 978-1-61284-861-7/11/$26.00 ©2011 IEEE.
[8]. Ravi Mishra, Singhai SK and Monisha Sharma (2013). A comparative based study of different video shot boundary detection algorithms. International Journal of Advanced Research in Computer Engineering & Technology 2(1).
[9]. Anh, N.D., P.T. Bao, B.N. Nam and N.H. Hoang, 2010.A new CBIR system using SIFT combined with neural network and graph-based segmentation. Lecture Notes Comput. Sci., 5990: 294-301. DOI: 10.1007/978-3-642-12145-6_30
[10]. Anh, T.Q., P. Bao, T.T. Khanh and B.N.D Thao, 2011.Shot Detection Using Histogram Comparison and mage Subtraction. GESTS Int. Trans. Comput.Sci. Eng.

[11]. Boreczky, J.S. and L.D. Lynn, 1998. A hidden Markov model framework for video segmentation using audio and image features. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, May 12-15, IEEE Xplore Press, Seattle, pp: 3741-3744. DOI: 10.1109/ICASSP.1998.679697
[12]. Cao, Y., W. Tavanapong, K. Kim and J.H. Oh, 2003. Audio-assisted scene segmentation for story browsing. Lecture Notes Comput. Sci., 2728: 446-455. DOI: 10.1007/3-540-45113-7_44
[13]. Felzenszwalb, P.F. and D.P. Huttenlocher, 2004. Efficient graph-based image segmentation. Int. J. Comput. Vision, 59: 167-181. DOI: 10.1023/B:VISI.0000022288.19776.77
[14]. Flickner, M., H. Sawhney, W. Niblack, J. Ashley andQ. Huang, 1995. Query by image and video content: The QBIC system. IEEE Comput., 28: 23-32. DOI: 10.1109/2.410146
[15]. Geetha, P. and V. Narayanan, 2008. A survey of content-based video retrieval. J. Comput. Sci., 4:474-486. DOI: 10.3844/jcssp.2008.474.486
[16]. Han, B., G. Xinbo and J. Hongbing, 2005. A shot boundary detection method for news video based on rough-fuzzy sets. Int. J. Inform. Technol., 11:101-111.
[17]. Jain, A.K. and R.C. Dubes, 1988. Algorithms for Clustering Data. 1st Edn., Prentice Hall, Englewood Cliffs, New Jersey, ISBN-10: 013022278X, pp: 320.
[18]. Li, S. and Lee, 2005. An improved sliding window method for shot change detection. Proceeding of the 7th IASTED International Conference Signal and Image Processing, Aug. 15-17, USA., pp: 464-468.
[19]. Lowe, D.G., 1999. Object recognition from local scaleinvariant features. Proceedings of the 7th IEEE International Conference on Computer Vision, Sep. 20-27, IEEE Xplore Press, Kerkyra, Greece, pp: 1150-1157. DOI: 10.1109/ICCV.1999.790410
[20]. O'Toole, C., A.F. Smeaton, N. Murphy and S. Marlow,1999. Evaluation of automatic shot boundary detection on a large video test suite. Proceeding of the 2nd U.K. Conference Image Retrieval: The Challenge of Image Retrieval, Feb. 25-26, UK., pp: 1-12.
[21]. TRECVID, 2006. An overview of up-to-date methods in content-based video retrieval --- by examining top performances in TREC video retrieval evaluation. TRECVID.
[22]. Zhu, X., X. Wu, A.K. Elmagarmid, Z. Feng and L. Wu,2005. Video data mining: Semantic indexing and event detection from the association perspective. IEEE Trans. Knowl. Data Eng., 17: 665-677. DOI: 10.1109/TKDE.2005.83