International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Performance Analysis of Content Based Video Retrieval System Using Clustering

Poonam O. Raut[1], Nita S. Patil[2]

[1]Research scholar, Department of Computer Engineering, Datta Meghe College of Engineering, Airoli Navi Mumbai
[2]Assitant Professor, Department of Computer Engineering, Datta Meghe College of Engineering, Airoli Navi Mumbai

Abstract: ***the multimedia storage grows and the cost for storing multimedia data is cheaper. So there is huge number of videos available in the video repositories. It is difficult to retrieve the relevant videos from large video repository as per user interest as users shift from text based retrieval systems to content based retrieval systems. Video retrieval is very important in multimedia database management. This paper offers an overview of the landscape of general strategies in visual content-based video retrieval, focusing on methods for video structure analysis, including shot boundary detection, key frame extraction, extraction of features including static key frame features, object feature, video retrieval including similarity measures and the proposed procedure consists of the unique aspect of clustering techniques.***

Keywords: Content-Based Video Retrieval; Video Structure Analysis; Shot Boundary Detection; Key Frame Extraction; Similarity Measure; Video Retrieval

1. Introduction

Due to the increase of available network bandwidth, many users access the videos from large video repositories like YouTube. For example, in YouTube, over 48 h of new videos are uploaded to the site every minute, and more than 14 billion videos were viewed in May 2010. It is difficult to manually index and retrieve from large video repositories. It is also difficult to search with in long video clips in order to find portions of segments that the user might interested and all these data are almost not usable in the absence of a proper search method.

The demand for intelligent processing and analysis of multimedia information has been rapidly growing in recent years. Researchers have actively developed different approaches for intelligent video management, including shot transition detection, key frame extraction, video retrieval. Among these approaches, shot transition detection is the first step of content-based video analysis and key frame is a simple yet efficient form of video abstract. It can help users to understand the content at a glance and is of practical value.

Content-based Video Retrieval (CBVR) systems appear like a natural extension (or merge) of Content-based Image Retrieval, "Content-based" means that the search will analyze the actual content of the video. The term 'Content' in this context might refer colours, shapes, textures. However, there are a number of factors that are ignored when dealing with images which should be dealt with when using videos. In visual content-based video retrieval, focusing on methods for video structure analysis, including shot boundary detection, key frame extraction and extraction of features and similarity measure [1].

2. Review of Literature

The first step for video-content analysis, content based video browsing and retrieval is the partitioning of a video sequence into shots. A shot is defined as an image sequence that presents continuous action which is captured from a single operation of single camera. Shots are joined together in the editing stage of video production to form the complete sequence. Shots can be effectively considered as the smallest indexing unit where no changes in scene content can be perceived and higher level concepts are often constructed by combining and analyzing the inter and intra shot relationships.[2] Almost all shot change detection algorithms reduce the large dimensionality of the video domain by extracting a small number of features from each video frame. These are extracted either from the whole frame or from a subset of it, which is called a region of interest (ROI). Such features include Luminance/color, Luminance/color histogram, Image edges, Transform coefficients (DFT, DCT, wavelet). Luminance/color where the simplest feature that can be used to characterize a ROI is its average grayscale luminance. This, however, is susceptible to illumination changes. A better choice is to use some statistics of the values in a color space. In Luminance/color histogram, A richer feature for a ROI is the grayscale or color histogram. It is quite discriminant, easy to compute and mostly insensitive to translational, rotational and zooming camera motion, for the above reasons it is widely used. In Image edges, an obvious choice of feature is edge information in a ROI .Edges can be used as is, be combined into objects or used to extract ROI statistics. They are invariant to illumination changes and most motion, and they correspond somewhat to the human visual perception. Their main disadvantage is computational cost, noise sensitivity and high dimensionality. In Transform coefficients (DFT, DCT, wavelet), These are a classic way to describe the texture of a ROI. The DCT coefficients are also present in MPEG encoded video streams or files. Their greatest problem is that they are generally not invariant to camera zoom and other features such as the color anglogram [3].

Key-frames are still images extracted from original video data that best represent the content of shots in an abstract manner. Key-frames have been frequently used to supplement the text of a video log, though they were selected manually in the past. Key-frames, if extracted properly, are a very effective visual abstract of video contents and are very useful for fast video browsing. A video summary, such as a movie preview, is a set of selected segments from a long

video program that highlight the video content, and it is best suited for sequential browsing of long video programs. Apart from browsing, key-frames can also be used in representing video in retrieval video index may be constructed based on visual features of key-frames, and queries may be directed at key-frames using query by retrieval algorithms. key frames extraction can be done using Sequential comparison based Approach, Global comparison based Approach, Reference frame-based Approach, Clustering based Approach, Curve simplification-based Approach, Object/event-based Approach[2].

Once key frames are extracted next step is to extract features. The features are typically extracted off-line so that efficient computation is not a significant issue, but large collections still need a longer time to compute the features. Features of video content can be classified into low-level and high-level features. Low-level features such as object motion, color, shape, texture, loudness, power spectrum, bandwidth, and pitch are extracted directly from video in the database. High-level features are also called semantic features. Features such as timbre, rhythm, instruments, and events involve different degrees of semantics contained in the media. High-level features are supposed to deal with semantic queries [4].

The key frames of a video reflect the characteristics of the video to some extent. Traditional image retrieval techniques can be applied to key frames to achieve video retrieval. The static key frame features useful for video in retrieval are mainly classified as color-based, texture-based, and shape-based.

Color-based features include color histograms, color moments, color correlograms , a mixture of Gaussian models, etc. The exaction of color-based features depends on color spaces such as RGB, HSV, YCbCr and normalized r-g, YUV, and HVC [5].

In Shape-Based Features, Shape-based features that describe object shapes in the image can be extracted from object contours or regions. A common approach is to detect edges in images and then describe the distribution of the edges using a histogram.

In Texture features in common use include Tamura features, simultaneous autoregressive models, orientation features, wavelet transformation-based texture features, co-occurrence matrices, etc [2].

3. Proposed System

3.1 Shot Boundary Detection

Partitioning a frame into blocks with m rows and n columns, and computing histogram matching difference between the corresponding blocks between consecutive frames in video sequence and then computing histogram difference between two consecutive frames then threshold calculated by computing the mean and standard variance of histogram difference over the whole video sequence. Shot candidate detection: if $D(i,i+1) \geq T$, the ith frame is the end frame of previous shot, and the (i +1)th frame is the end frame of next shot[6].

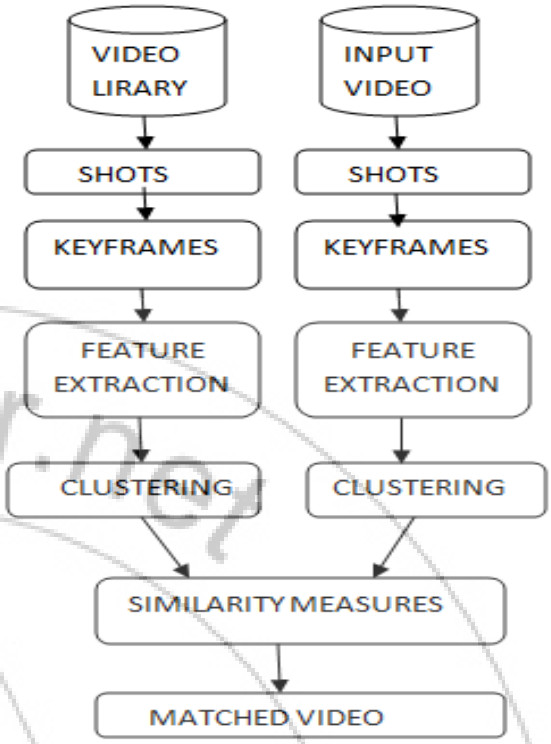

Figure 1: Proposed System

3.2. Key Frame Extraction

Computing the difference between all the general frames and reference frame with the above method, Searching for the maximum difference within a shot, Determining "ShotType" according to the relationship between max(i) and MD: Static Shot (0) or Dynamic Shot, Determining the position of key frame: if ShotTypeC=0, with respect to the odd number of a shot's frames, the frame in the middle of shot is choose as key frame; in the case of the even number, any one frame between the two frames in the middle of shot can be chose as key frame. If ShotTypeC=1, the frame with the maximum difference is declared as key frame [6].

3.3. Feature Extraction

3.3.1. Color Extraction

Color descriptors of images and video can be global and local. Global descriptors specify the overall color content of the image but with no information about the spatial distribution of these colors. Local descriptors relate to particular image regions and, in conjunction with geometric properties of these latter, describe also the spatial arrangement of the colors. In particular, the MPEG-7 color descriptors consist of a number of histogram descriptors, a dominant color descriptor, and a color layout descriptor (CLD).

• RGB Moment

The frame is divided into 5x5 blocks and the first, second, and third central moments of RGB components are calculated for each block and then all the moments are concatenated in order of blocks to form a feature with 255 dimensions.

3.3.2. Shape Feature

A common approach is to detect edges in images and then describe the distribution of the edges using a histogram. Use

the edge histogram descriptor (EHD) to capture the spatial distribution of edges for the video search. The EHD is computed by counting the number of pixels that contribute to the edge according to their quantized directions. To capture local shape features, first divide the image into 4×4 blocks and then extract an edge histogram for each block. Shape-based features are effective for applications in which shape information is salient in videos [2].

3.3.3. Texture Feature

In LBP operator method, the image is divided into a set of blocks and for each block, the LBP operator labels the pixels by applying threshold operation over the 3x3-neighbourhood of each pixel with the center value. For every pixel in a block, the pixel to each of its 8 neighbors (on its left-top, left-middle, left-bottom, right-top, etc.) is compared. Then, the pixels along a circle, i.e. clockwise or counter-clockwise is followed and we put "1" if the center pixel's value is greater than the neighbor; otherwise, we put "0". This gives an 8-digit binary number which is converted into decimal value and the procedure is repeated for every pixel in the block. Subsequently, the histogram is computed over the block according to its frequency of each "number" occurring and finally, normalized histograms of all blocks are concatenated. This provides the texture feature vector for the input frame [2].

3.4. Region Thesaurus

Generally, a thesaurus combines a list of every term in a given domain of knowledge and a set of related terms for each term in the list. Here the constructed Region Thesaurus contains all the Region Types that are encountered in the training set. These region types are the centroids of the clusters and all the other feature vectors of a cluster are their synonyms. It is important to mention that when two region types are considered to be synonyms, they belong to same cluster, thus share similar visual features, but do not necessarily share the same semantics. By using a significantly large training set of keyframes, our thesaurus is constructed and enriched [8].

3.5. Clustering

Suppose we don't have a clear idea how many clusters there should be for a given set of data. Subtractive clustering is a fast, one-pass algorithm for estimating the number of clusters and the cluster centers in a set of data. The cluster estimates obtained from the subclust function can be used to initialize iterative optimization- based clustering methods and model identification methods. The subclust function finds the clusters by using the subtractive clustering method. This algorithm can have a large reduction on the number of training samples, based on the density of surrounding data points. Namely, all data points in a small dense zone of one point center will be replaced by this typical one. On the other hand, the sparse points in the input space will remain as cluster centers themselves. So this algorithm is noise robust, outliers have little influence on the choice of cluster centers [9].

3.6. Similarity Measures

Similarity measurement plays an important role in retrieval. A query frame is given to a system which retrieves similar videos from the database. The distance metric can be termed as similarity measure, which is the key-component in Content Based Video Retrieval. In conventional retrieval, the Euclidean distances between the database and the query are calculated and used for ranking. The query frame is more similar to the database frame if the distance is smaller.

3.8. Quantitative Analysis

The performance of the proposed CBVR system is evaluated on the input dataset using the precision, recall and F-measure. For quantitative analysis, videos from each category are given to the proposed system and results are evaluated with the defined measures as follows: Precision (P) and Recall (R).

$$P = \frac{(\text{Similar video}) \cap (\text{Retrieved video})}{(\text{Retrieved video})}$$

$$R = \frac{(\text{Similar video}) \cap (\text{Retrieved video})}{(\text{Similar video})}$$

$$\text{F-measure} = \{2*PR\}/(P+R)$$

Recall reflects the system's ability of retrieval related videos, while the precision reflects the ability of rejecting the unrelated videos. The evaluation results obtained by employing the input parameters shows the performance of the proposed system in retrieving the relevant videos and it clearly differentiate the results obtained for two different input parameters [8].

Support vector machines (SVMs) are a kind of machine learning method for both classification and regression problems. They base on the structural risk minimization principle. SVMs have been applied to deal with a wide range of problems due to their high generalization ability and good classification precision. At present, SVMs have been applied to many classification and recognition fields, such as handwriting recognition, text classification, face recognition, and speech recognition, etc. Support vector machines outperform conventional classifiers especially when the number of training data is small. However, for the large and high dimensional data sets, the kernel computation and optimization time for training a SVM is time consuming [10].

4. Conclusion

Video retrieval is very important in multimedia database management. Video stream will become the main stream in the years to come. Better off if we had an efficient CBVR search engine ready. Still many area needs to be improved. To overcome the problem video is divided into shots n shot boundary detection technique use statistical difference method, key frame extraction method use reference frame-based method and feature extraction based on color, texture and shape and we use clustering methods to cluster the frame based on their low level visual features. Here the number of comparisons is reduced. Then the similarity measure between the clustered frames and the query frames is found and get the retrieved results. It works faster than the previous approaches.

5. Future Scope

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) Fusions of multiple model information in multiple levels are all difficult issues in the fusion analysis of integrated models.

References

[1] B. V. Patel, B. B. Meshram (2007), "Retrieving and Summarizing Images from PDF Documents", International Conference on Soft computing and Intelligent Systems (ICSCSI-07), Jabalpur, India.
[2] Weiming Hu, Senior Member, IEEE, Nianhua Xie, Li Li, Xianglin Zeng, and Stephen Maybank "A Survey on Visual Content-Based Video Indexing and Retrieval" IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, VOL. 41, NO. 6, NOVEMBER 2011.
[3] Purnima.S.Mittalkod, Dr.G.N.Srinivasan "Shot Boundary Detection Algorithms and Techniques" Research Journal of Computer Systems Engineering- An International Journal http://technicaljournals.org ISSN: 2230-8563; e-ISSN-2230-8571
[4] B V Patel and B B Meshram "content based video retrieval systems" International Journal of UbiComp (IJU), Vol.3, No.2, April 2012.
[5] S.Padmakala, Dr.G.S.AnandhaMala, M.Shalini "An Effective Content Based Video Retrieval Utilizing Texture, Color and Optimal Key Frame Feature" 2011 International Conference on Image Information Processing (ICIIP 2011)
[6] ZHAO Guang-sheng "A Novel Approach for Shot Boundary Detection and Key Frames Extraction" 2008 International Conference on MultiMedia and Information Technology
[7] David G. Lowe, Proc. of the International Conference on Computer Vision, Corfu, September, 1999. " Object Recognition from Local ScaleInvariant Features
[8] Evaggelos Spyrou and Yannis Avrithis " High-Level Concept Detection in Video Using a Region Thesaurus"
[9] Xianglin Zeng, WeimingHu, Wanqing Liy, Xiaoqin Zhang, Bo Xu "Key-Frame Extraction Using Dominant-Set Clustering" NSFC (Grant No. 60672040
[10] Yu Xiaohong,Xu Jinhua "The Related Techniques of Content-based Image Retrieval" 2008 International Symposium on Computer Science and Computational Technology"

Authors Profile

Poonam O. Raut, Research Scholar, Department of Computer Engineering, Datta Meghe College of Engineering, Airoli Navi Mumbai, India

Nita S. Patil, Assistant Professor, Department of Computer Engineering, Datta Meghe College of Engineering, Airoli, Navi Mumbai, India