Communications on Applied Electronics (CAE) – ISSN : 2394-4714
Foundation of Computer Science FCS, New York, USA
Volume 7 – No. 5, August 2017 – www.caeaccess.org

Video Retrieval System using Shot detection and Analysis of Frame Dissimilarities using Different Parameters

Snehal Harishbhai Patel
PG Student, Dept. of EXTC,
DJSCE, Vile Parle
Mumbai, India

Vivek Deodeshmukh
Asst. Prof, Dept. of Biomedical,
DJSCE, Vile Parle
Mumbai, India

ABSTRACT

Today, nearly above 400million of population uses internet. Internet is mostly used for search engines like google, where an individual search for information. Roughly 100million hours of videos are uploaded over internet daily (Viz. YouTube, Netflix, Dailymotion, Vimeo, Veoh, Metacafe, etc.) due to this tremendous amount of data is generated. Semantic/context based search is use which matches fast but only with correct tags. Each and every video is assigned with tags, the desired video is retrieved if and only if correct tags are used. Due to involvement of large number of frames in videos, it is difficult to extract desired video using context based matching. The proposed system is developed to extract desired video from huge data base. Algorithm consists of content based shot detection method and features are extracted for each data set of videos. Further, the frames dissimilarities are analyzed by different parameters like entropy, probability, color, etc. User can search desired video by using image as an input to the system. Proposed system successfully achieved 100% accuracy in content based search. This algorithm also reduces the search time than the existing one that is roughly 0.3ms/video, which is much faster and reliable. Expectation oof system is algorithm fits for content based video search and also gives alternative to context based search (e.g. Netflix, YouTube).

General Terms

Video Retrieval, Shot detection, Different parameters for matching, SVM, Image as query, Content based video retrieval system

Keywords

Video retrieval, content based matching, frames, feature vector, entropy, probability

1. INTRODUCTION

Video Retrieval [1] system proposed in this paper is content based retrieval system. Now-a-days semantic/context based retrieval system is used where context/words are used as query. To search information that means referring to search data from biggest and best search engine called ‘google’. Data collection of google is huge, out of an individual’s imagination. To find one desired video from such huge data quickly and efficiently is next to impossible. When an individual search for a video on internet using context based search he/she uses common and easy words to find single video from huge database. As videos are made up of over thousands of frames depending upon their duration and each video is assigned with correct tags, hence very difficult for an individual to keep a track of each and every tag. Example when an individual search “Eiffel tower” on google videos get over 17 lakh videos in result. Now a common man does not have that much time to look each and every video of result to get desired video. If searching factor relies totally on text or metadata, then it is very difficult to retrieve desired video in fewer attempts and thus consumes more time.

Interest in content based video retrieval [2] system has grown due to limitation in text/metadata [3] based search retrieval system. CBVR system can also be called as query based retrieval system. However, query here will be an image itself. In textual based retrieval the person has to justify the correct assigned tag for each and every image to retrieve video from huge database like YouTube, Netflix etc. So, it may happen that person may miss the desired video due to use of synonyms of words for text based query. While on the other hand content-based retrieval analyses all the contents like color, texture, shape or other parameters which describes the images/frames of the video. Thus, makes it easier, reliable and efficient retrieval system.

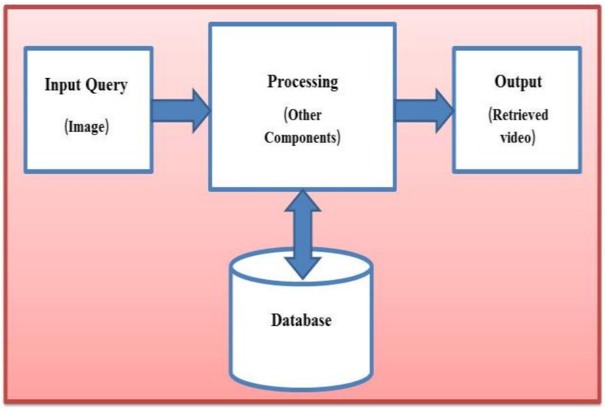


Fig. 1: Block diagram of video retrieval system

Fig 1 demonstrates CBVR system model where the system input is image of the particular desired video person needs to retrieve. Database is collection of all the key-frames and videos. After processing all the features of key-frames, videos and query image with other components, the desired and relevant output video is retrieved for the query image.

2. METHODOLOGY

First and essential step for this system is to generate database. For database all the different types of videos are collected. These videos can be of any format (viz. mpg, mp4, avi, ogg, etc.). Next is to generate frames from these videos. For frame generation proposed system have used shot detection method. Shot is a part of video taken by single camera recorder without any disturbance. It is unbroken sequence frames of the video. Aim is to detect shots in pre-processing step and provide user friendly interface for further process. Scene changes or cuts are first indication of shot boundary detection. First the extraction of visual features is prepared from each frames of the specific video, then the similarities between those extracted frames are measured and finally shot boundaries are detected between dissimilar frames of that video.

Frame Generation: The dissimilarities are measured by following parameters like canny edge detection, correlation, probability, entropy, color features and SURF.

2.1 Canny edge detection method:

Canny edge detector [4] is used for object in motion like video. For example, if an individual take photo of running fan or waving hand the picture taken comes out blur image as capturing picture of moving object. Sometimes the videographer hands shake while recording video which leads to shaky footage. Here system use canny edge detector for extracting highly defined key-frames from object in motion. Steps for canny edge detection method are:

1. Canny uses Gaussian filter to remove image noise to prevent false detection and smooth the image.
2. Intensity gradient of each pixel is calculated, algorithm uses four filters to detect edges horizontally, vertically and diagonally in the blurred image.
3. Apply edge thinning technique, as the edge extracted from gradient value is still moderately blurred.
4. After edge thinning, remaining pixels provides more precise representation of edges in an image. However, for more accuracy double thresholding is done due to noise and color variation. Edge linking is use to accomplish the accurate result.

2.2 Correlation

In digital image, correlation is used to measure the changes in two or more image. It is used to calculate the similarities between those compared images say if the image is correlated (same) then it shows value 1 or else 0. Here, system will use to measure pixel intensity between two or more images to verify correlation.

2.3 Probability of frames

Probability difference is another algorithm use for shot detection. Probability of each pixel is calculated of an image. Each pixel is divided by 255, as it is gray scale image. Sum of probability of each pixel of an image is calculated. Mapping of each pixel is done and then summation of total probability of image is divided by size of an image [m n].

2.4 Entropy

The uncertainty of random variables and the average information content within an image is measured by entropy [6]. Gray level intensities of an image are used to track changes in image time to time. As video contains abrupt changes in frames, entropy is very effective for such abrupt boundary detection and calculates the information.

2.5 Color

Color [5] is very effective and important feature of an image for retrieval system as it gives maximum information of intensity and brightness of an image. System generates color histogram for each RGB distribution (red, blue & green) of an image which is further use for HSV plane. HSV stands for hue, saturation and value. HSV is cylindrical model with angle around vertical axis corresponds to ‘hue’, distance from axis is ‘saturation’ and distance along the axis is ‘brightness’.

2.6 SURF

SURF[6] stands for ‘Speeded-Up Robust Feature’ which is also used for abrupt boundary detection, as the frequency of abrupt shot detection occurs more often than the gradual shot boundaries. This algorithm matches the key points between the images. SURF is use for fast and robust point matching between two images. SURF employs under scale, rotation, noise, changes in illumination and also clustered background.

These were the feature vectors system used for extracting the shot boundaries by correlating and identifying dissimilarities between the frames of respective video. The differences of each algorithm are also calculated meanwhile to relate the key-frames and analyses the correlation, probability difference, entropy difference, color feature vector and also SURF (to know the matching point) between two images.

Fig. 2: Query image

As the frames are generated now next step is to retrieve the desired video using image as query. Any image can be used as a query image as shown in fig.2 Feature vector of size 23x74 is generated which relates to following feature datasets like color moment, LL, LH, HH and HL features, dwt (discrete wavelength transform)[7] as shown in figure 3, and the mean values from each of these features.

For the matching vector, average wavelet feature values are used. Color autocorrelogram is used to correlate the color of the frames. To determine the exact color sequence and its nearby values, autocorrelation function is used . Edge detection algorithm allows us to match the shapes and textures. The main edge detection algorithm system followed in this project is canny edge detection method and the shapes in the image are determined by the Hough transform with rho resolution of value 0.5 and theta resolution of value 0.5. Once the shape and texture is determined for each and every frame, now the HSV model is used to compare the three planes i.e. Hough saturation value and the histogram comparison of each plane of HSV in 10 different banalizations i.e. 10x10x10 classes where defined and matched.

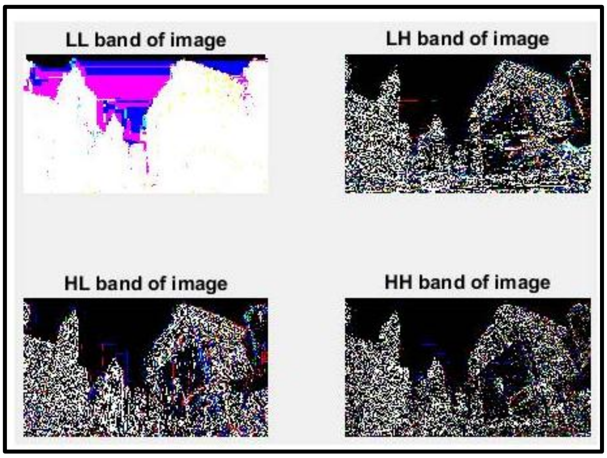


Fig. 3: DWT of query image

Finally, the feature vector matrix consists of color moment, average of LL, average of LH, average of HH, average of HL and C (color autocorrelogram) which is then matched with the training dataset fraction. In the concluding part of the algorithm a group train sequence is used i.e. a training vector and corresponding feature vectors both are included. Here the known values of videos and corresponding feature vector are feed in for any unknown sample input given randomly. This portion of classification is developed using SVM classifier which is put up in a self-training mode. SVM (Support Vector Machine) which is type of machine learning with highest accuracy used for classification of data. Finally the output of SVM classifier is multiclass SVM with different weights and highest weighted among all value is displayed as output retrieved video and its corresponding class.

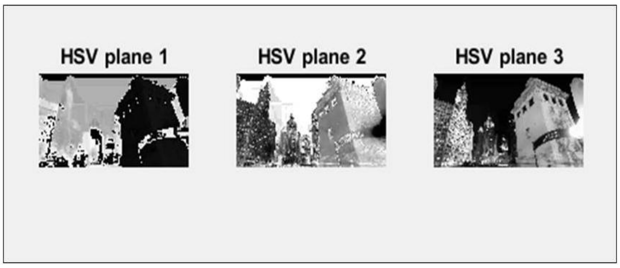
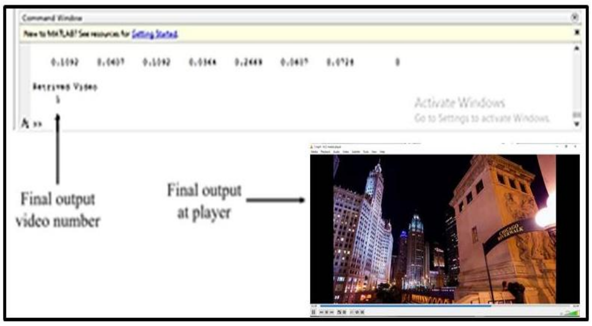


Fig. 4: HSV plane

3. RESULT

This retrieval algorithm is tested over 229 image inputs and 12 videos with 458 shots. The accuracy with database images is 100%. For blind testing with unknown images the accuracy goes down nearly 97.2% which can be mostly because of size of the image at search input and illumination conditions of image. Time taken by proposed algorithm is 0.3msec per video for retrieval which is approximately 100 times faster reported in literature. This system also analyses the different parameters like correlation, probability, entropy, color and SURF to differentiate the frames generated by shot boundary detection of particular video [8].


Command Window
New to MATLAB? See resources for Getting Started.
Activate Windows
Go to Settings to activate Windows.
Final output video number
Final output at player

```
    0.1192    0.0407    0.1192    0.0344    0.2449    0.0417    0.0724         1

Retrived Video
     1

>>
```

Fig. 5: Output of System

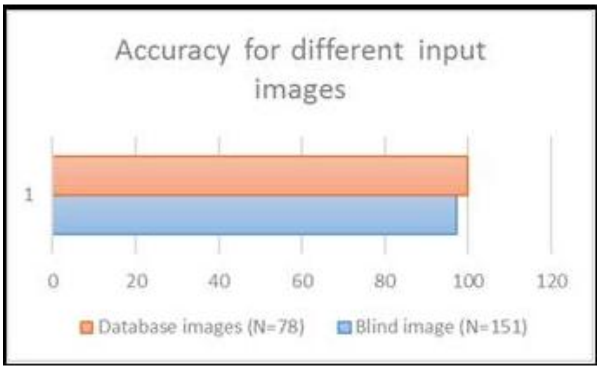


Fig. 6: Accuracy for Different Test Conditions

4. CONCLUSION

Proposed algorithm for effective video retrieval system is totally based on content based search where query will be image itself. This method gives 100% accuracy when the system input image is within the database and it can reach nearly up to 97.2% accuracy when the system input image search is blind i.e. out of database search. Time taken for the retrieval is roughly 0.3msec/video which is approximately 100 times faster than the existing algorithms. Different parameters are analyzed to measure the dissimilarities between two or frames. This video retrieval system is dependent on spatial domain as well as frequency domain information and varies mainly depending upon the luminance and aspect ratio of the signal. More robust video retrieval systems can be modeled using frequency and time domain analysis methods and computational speed can also further be improved using high end computers with high capabilities in terms of processor, clock speed and RAM.

5. ACKNOWLEDGMENTS

Authors are grateful and would like to thank SVKM for the infrastructure and facilities provided.

6. REFERENCES

[1] Jiang, Yu-Gang, Chong-Wah Ngo, and Jun Yang. "Towards optimal bag-of-features for object categorization and semantic video retrieval." Proceedings of the 6th ACM international conference on Image and video retrieval. ACM, 2007.

[2] Zhang, Hong Jiang, et al. "An integrated system for content-based video retrieval and browsing." Pattern recognition 30.4 (1997): 643-658.

[3] ZHAN Chaohui DUAN Xiaohui, et al. “An Improved Moving Object Detection Algorithm Based on Frame Difference and Edge Detection.” Fourth International Conference on Image and Graphics IEEE 2007

[4] Hua Zhang, et al. "A Shot Boundary Detection Method Based on Color Feature." 2011 International Conference on Computer Science and Network Technology

[5] Junaid Baber, et al. "Shot boundary detection from videos using entropy and local descriptor", IEEE 2011

[6] Serdean, C. V., et al. "DWT-based high-capacity blind video watermarking, invariant to geometrical attacks." IEE Proceedings-Vision, Image and Signal Processing 150.1 (2003): 51-58.

[7] Sivic, Josef, and Andrew Zisserman. "Video google: A text retrieval approach to object matching in videos." iccv. Vol. 2. No. 1470. 2003.