INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)
Dundigal, Hyderabad-500 043

Project Based Learning
(Prototype / Design Building)

External Evaluation Report

Title of your Idea : **Video Frame Extraction**

Thrust Area / Sector : **Image Processing**

Branch : **ECE-C**

Year / Semester :**3rd Year/5th Semester**

S. No	Name of the Student	Roll Number	Mobile Number	Signature
1	SHAIK ZEESHAN	19951A04F4	9515770299	

1. Background of the Idea (Min. 500 words):

Rapid increase in use of multimedia content is very commonly observed in current generations. Representation of information content retrieved from internet in the form of multimedia content comprehends the users understanding leading to increase in retrieval and processing of multimedia content. Video is one of the most widely retrieved multimedia data which provides instant gratification and quick fixes for all the types of user information desires. Design of video processing algorithms to simplify the computational loads is one of the challenging research objectives. Video classification and key frame detection from videos are the vital protocols in the sequel of video retrieval tasks. In this paper, the performance study of various key frame detection methods from videos is carried out towards the sports videos. The data sets of sports videos are downloaded from YouTube and Google search engine. Key frame detection is carried out using Histogram differencing; Region of Interest and Shot Boundary detection and further the performance of three methods are analysed by comparing the results with ground truth data. It is observed that Shot boundary detection method is found to be more performance efficient compared to other two methods.

Video classification is the process of grouping videos based on similarities of contents. Video classification has become significant in the current technology trends to contend the variety of user needs.

The challenging factors of video classification include size of the videos, number of videos, different categories of videos, and formats. The video classification directs the user towards more desired choices by associating the same kind of videos together. One of the vital discriminative factor for video classification is its content.

Videos are basically composition of frames and it is obvious that frames directly extracted from videos are comprised of lots of duplicate frames as same frame is casted from 1to 2 or more seconds of time duration. It is very essential to identify the salient frames that convey the key features of the videos which is accomplished through key frame detection.

The removal of duplicate frames from videos is often termed as key frame detection. The detection of key frames from videos is more challenging due to the length of videos and this desires the efficient algorithms that can identify the key frames. Key frames are candidate frames defining the salient contents of videos uniquely without any overlap or repetition.

Detection of key frames from videos reduces the computational burdens involved in processing of videos and also enables optimal classification.

Key frame detection includes variety of benefits in real time including video abstraction, compression, browsing and searching, text summarization from videos, story generation, automatic index structure generation and various other performance concerns of retrieval algorithms. Varieties of techniques are reported in the past for key frame detection the insights into some of the key works in the literature are as discussed subsequently.

2. Problem Statement (Min 100 words):

This section formally defines the problem of reliably extracting text from video as addressed by this thesis. We have an input color video stream that may be in MPEG-1 compressed, Motion-JPEG compressed, or raw uncompressed formats, which may be noisy, with blocking artifacts, and unknown bit rate. The scene content is unconstrained and may be both indoor or outdoor scenes under any lighting or contrast conditions. The nature of illumination may be such that the imaged text suffers from specularity and variation of color, intensity and contrast across the character strokes. The input video stream may have artificial text composited on top of it and may also contain scene text. Horizontal orientation is usually assumed for artificial text but in some cases this condition may not be satisfied. The orientation of the scene text is unconstrained. Artificial text may be stationary or scrolling vertically or horizontally. More unusual movements are rare but cannot be ruled out. The scene text may be painted or printed, usually on a planar surface. The case of curved surfaces or 3-D text objects is assumed to be less frequent. Scene text may move in the image plane due to camera or object motion. Its image on the video frame may also be distorted by perspective projection.

3. Proposed Solution (Min 100 words):

Based on the above considerations, the frame difference method is used to extract key frames by analysing the presence of spatial redundancy and temporal redundancy. In order to improve operational efficiency, it is worth mentioning that this method is different from the traditional shot segmentation method for that the traditional approach is to conduct a video shot segmentation, then to extract key frames from each shot, and finally to compose key frame sequence of the video. In this method, the segmentation is not considered and then to extract key frames directly from the video.

In existing method the algorithm was Implemented in c , c++ and open cv2. It is used to frame the image in particular intervals and the time to frame ratio is more high . In the existing method they used open cv2 to frame the image of the video

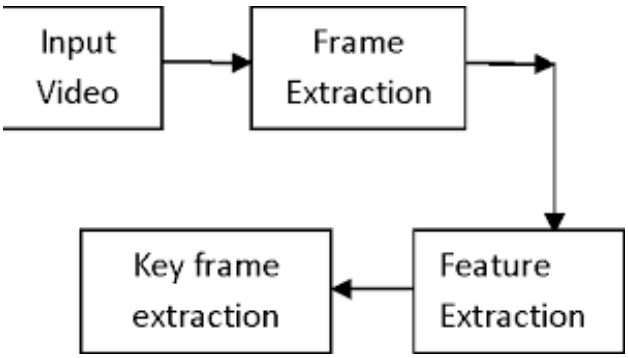
4. Technology concept formulation:

There are some distinct features about the key frame for video copyright protection. So, the key frame for video copyright protection is defined firstly before video pre-processing and key frame extracting.

The key frames should meet the following three conditions. The gray value of a key frame is within a certain range to allow viewers to have subjective perception about the video content.

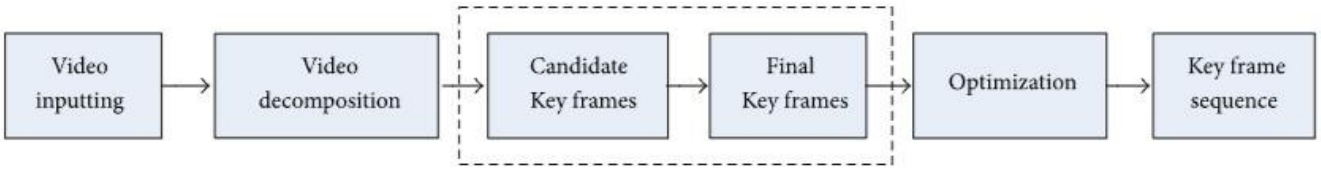
Four images with low gray value in Figure are extracted from a single video, which is difficult for almost viewers to recognise the content.

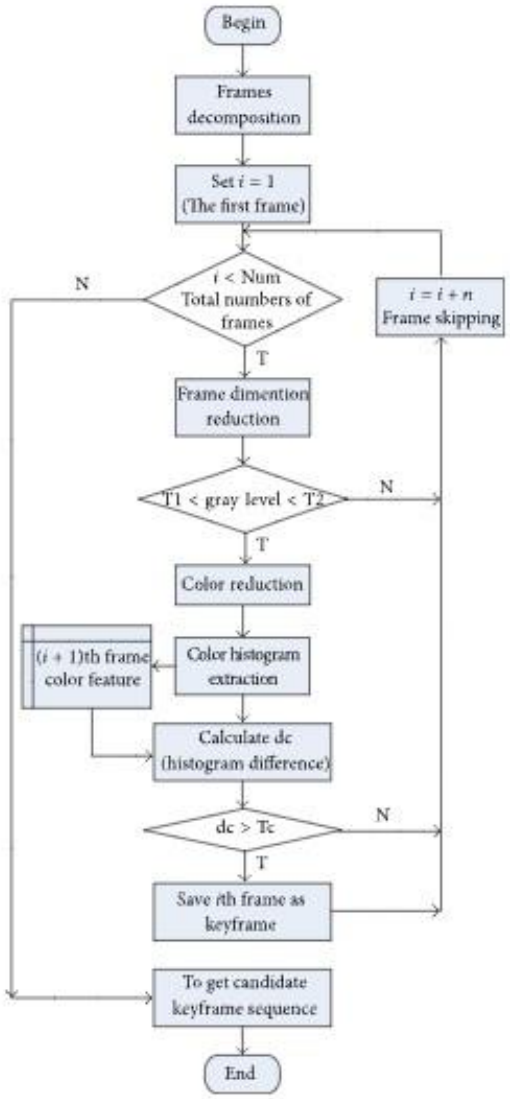
The final key frame sequence must be arranged in chronological order consistent with original video sequence, in order to satisfy temporal features and to be different from the short promotion trailer. Appropriate redundancy of some key frames is allowed to ensure the periods or intervals along the processing of video content. Figure indicates the condition by selecting four images from a tested video, which are with similar content, that is to say, one judge in the show every once in a while.




```python
import cv2


# Function to extract frames
def FrameCapture(path):
    # Path to video file
    vidObj = cv2.VideoCapture(path)

    # Used as counter variable
    count = 0

    # checks whether frames were extracted
    success = 1

    while success:
        # vidObj object calls read
        # function extract frames
        success, image = vidObj.read()

        # Saves the frames with frame-count
        cv2.imwrite("frame%d.jpg" % count, image)

        count += 1


# Driver Code
if __name__ == '__main__':
    # Calling the function
    FrameCapture("C:\\Users\\zeeshan\\OneDrive\\Documents\\pbl\\WhatsApp Video
2022-03-02 at 11.13.03 PM.mp4")
```

5. Prototype of proposed system (UI screens / block diagrams / circuits / designs):





6. Detailed description of prototype / product / project (Min 1000 words):

Key frame extraction from videos is the process of eliminating the redundant frames from videos, so that a comprehensive interpretation about a particular video can be achieved within a short span of time. In this paper, the attempts are addressing the key frame extraction from sports videos, in this direction, the key frame extraction is carried out using three efficient methods and a comparison is performed to understand the performance of each method. In the proposed research the efficiency of key frame extraction analysed using histogram differencing, region of interest and shot boundary detection techniques. The details of the techniques are discussed subsequently.

Image Differencing

Image differencing is one of the simple yet significant techniques for estimation of measures that represents the variability from one image to other. The most commonly employed image differencing techniques include absolute difference and histogram differences.

The overall efficiency of any technique is determined by its computational complexity, robustness and reliability. In the perspective of computational complexity, absolute differencing is an ideal method, however the method susceptible to noise and diacritics in the images. Hence, histogram differencing is an effective method for image differencing.

Histogram difference

Images with slighter distortions due to noise and motions can also be recognized effectively using this technique [17,18]. If difference from one image frame to other image frame is represented by, N indicates the dynamic

range of the display system or the number of intensities. Let and designates the current and previous frames, then histogram differencing from one frame to other is given by (1). $_{c}i_{p}i$

Where and represents the histogram of current and previous frames with respect to an arbitrary intensity ‘’. Similarly, histogram for an image frame defined in the RGB color space is given by and refers to the histogram differencing of red, green and blue spaces respectively. The occurrence of scene changes in particular frames to can be estimated as a result of change in motion or changes in spatial and temporal relationships in frames. In this technique a threshold set based on the mean and standard deviation for identification of key frames. The frame which exceeds the threshold limit is considered as key frames.

Region of Interest

Region of interest is an effective technique for detection of key frames in the video particularly when user interest is focused only on a desired object [19].

This technique greatly reduces the computational loads involved in detection of frame, since key frames are detected based on the change in the motion of objects considered as region of interest. Prior to processing of actual image frames, the region of interest objects is detected using Canny edge detection technique [20] and the object of interest bound by particular width, size and height within the frames are only considered as the key frames.

Shot boundary detection

A shot is generally considered a basic unit of video. Detecting shot boundary is retention of elementary video units, which provides video abstraction or a high-level video [21]. In the video production process, in order to support the content and context of the video sequence, we carefully selected each conversion type. Automatic retention of positions and types of objects facilitates in deducing high-level semantics of video sequences. For example, in a feature film, dissolution is often used to convey a period of time. News, sports, comedies and shows featured in feature films, documentaries, biographies and landscape videos often appear. Wet wipes the opposite situation. Therefore, video types can be automatically identified using automatic detection of conversions and their types. Shot detection is also useful for color black and white movies. For each shot, choose a different gray to color lookup table.Video structure analysis is a prerequisite for automated video content analysis. At different levels of structure (i.e., frames, shots, scenes, etc.), the lens-level organization is considered suitable for browsing and content-based retrieval. The shot includes a continuous sequence of frames captured by a single camera action. Depending on whether the conversion between shots is sudden or gradual, the shot boundaries can be divided into two types: Shot transition detection also called as CUT is a field of research of video processing and Generalized random transform (GRT). According to the characteristics of different editing effects, GT can be further broken down into digestion, erase, fade (FOI) and so on where a GT is a binary image with pixel inside the lesion curve labelled 1 and 0. Shot Border Detection (SBD) identifies the transition between adjacent shots. In the perspective of visual aspect, video is a kind of three-dimensional signal, in which two of them reveal the visual content in the horizontal and vertical frame direction, and the third one reveals the variations of the visual content over the time axes.

7. Final version of prototype / product (only images):


frame10


frame11


frame12


frame13


frame14


frame15


frame22


frame23


frame24


frame25


frame26


frame27


frame34


frame35


frame36


frame37


frame38


frame39


frame16


frame17


frame18


frame19


frame20


frame21


frame28

frame29


frame30


frame31


frame32


frame33


frame40


frame41


frame42


frame43


frame44


frame45

8. Any other information:

Here in this paper, we have compared three key frame extraction techniques. I.e. Histogram Differencing method, Region of Interest method, and Shot Boundary detection method for key frame extraction. The testing result of the experiment shows that Shot Boundary detection method has good accuracy for the detection of key frames. This method takes very less computational time to extract the key frames from the video. And also, we can see in the results that the duplicate frames are very minimal compared to the other two methods. This allows us to work on minimal frames and speeds up the process. It basically resolves the difficulties of detection caused by sub-window. The framework also greatly improves accuracy of gradual transitions of shot as compared to the other two methods.

References

[1] Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R. and Fei-Fei, L. Large-scale video classification with convolutional neural networks. *In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2014, 1725-1732.

[2] Rani, N. S., Neethu, O. P. and Ponnath, N. Automatic Vehicle Tracking System Based on Fixed Thresholding and Histogram Based Edge Processing. *International Journal of Electrical and Computer Engineering* **5** (4) (2015).

[3] Pawar, M. S., Perianayagam, L. and Rani, N. S. Region based image classification using watershed transform techniques. *In International Conference on Intelligent Computing and Control (I2C2)*, 2017, 1-5.

[4] Rani, N. S. and Vasudev, T. Automatic detection of telugu single and multi-character text blocks in handwritten words. *International Conference on Computing and Network Communications (CoCoNet)*, 2015, 234-240.

[5] Lei, S., Xie, G. and Yan, G. A novel key-frame extraction approach for both video summary and video index. *The Scientific World Journal* (2014).

[6] Zhuang, Y., Rui, Y., Huang, T. S., and Mehrotra, S. Adaptive key frame extraction using unsupervised clustering. *International Conference on Image Processing*, 1998, 866-870.

[7] Liu, T., Zhang, H. J. and Qi, F. A novel video key-frame-extraction algorithm based on perceived motion energy model. *IEEE transactions on circuits and systems for video technology* **13** (10) (2003) 1006-1013.

[8] Zhao, L., Qi, W., Li, S. Z., Yang, S. Q. and Zhang, H. Key-frame extraction and shot retrieval using nearest feature line (NFL). *In Proceedings of the ACM workshops on Multimedia*, 2000, 217-220.

[9] Gianluigi, C. and Raimondo, S. An innovative algorithm for key frame extraction in video summarization. *Journal of Real-Time Image Processing* **1** (1) (2006) 69-88.

[10] Lee, Y. J., Kim, J., and Grauman, K. Key-segments for video object segmentation. *In IEEE International Conference on Computer Vision (ICCV)*, 2011, 1995-2002.

[11] Sivic, J. and Zisserman, A. Video Google: Efficient visual search of videos. *In Toward category-level object recognition*, 2006, 127-144.

[12] Shah, K. M. and Makwana, R. M. Shot Boundary Detection using Logarithmic Intensity Histogram: An Application for Video Retrieval. *International Journal of Applied Engineering Research* **12** (22) (2017) 12616-12624.

[13] Prabavathy, A. K. and Shree, J. D. Histogram difference with Fuzzy rule base modeling for gradual shot boundary detection in video cloud applications. *Cluster Computing*, 2017, 1-8.

[14] Shi, Y., Yang, H., Gong, M., Liu, X. and Xia, Y. A fast and robust key frame extraction method for video copyright protection. *Journal of Electrical and Computer Engineering*, 2017.

[15] Zong, Z., and Gong, Q. Key frame extraction based on dynamic color histogram and fast wavelet histogram. *IEEE International Conference on Information and Automation (ICIA)*, 2017, 183-188.

[16] Sheena, C. V. and Narayanan, N. K. Key-frame extraction by analysis of histograms of video frames using statistical methods. *Procedia Computer Science* **70** (2015) 36-40.

[17] Bovik, A. C. *The essential guide to video processing*. Academic Press, 2009.

[18] Huang, T. Automatic region of interest extraction and tracking in coded video sequence. *IEEE International Conference on Control Science and Systems Engineering (CCSSE)*, 2014, 175-179.

[19] Canny, J. A computational approach to edge detection. *In Readings in Computer Vision*, 1987, 184-203.

[20] Ferreira, L., Da Silva Cruz, L. A. and Assuncao, P. Towards key-frame extraction methods for 3D video: a review. *EURASIP Journal on Image and Video Processing* (2016).

Dean-CLET **Dean -TIIC**