INTERNATIONAL JOURNAL OF RESEARCH – GRANTHAALAYAH
A knowledge Repository

RACSIT - 17

HYBRID APPROACH FOR KEY FRAME EXTRACTION FROM VIDEO SEQUENCE

N. Satish Kumar [1], Shobha G [2]
[1] Research Scholar, R V College of Engineering, Bangalore, India
[2] Prof & Head CSE Department, R V College of Engineering, Bangalore, India

DOI: https://doi.org/10.5281/zenodo.583896

Abstract

This paper proposed and developed hybrid approach for extraction of key-frames from video sequences from stationary camera. This method first uses histogram difference to extract the candidate key frames from the video sequences, later using Background subtraction algorithm (Mixture of Gaussian) was used to fine tune the final key frames from the video sequences. This developed approach show considerable improvement over the state-of-the art techniques and same is reported in this paper.

Keywords: Histogram; Background Subtraction; Key-Frame; Mixture of Gaussian (MoG).

Cite This Article: N. Satish Kumar, and Shobha G. (2017). "HYBRID APPROACH FOR KEY FRAME EXTRACTION FROM VIDEO SEQUENCE." *International Journal of Research - Granthaalayah*, 5(4) RACSIT, 97-104. https://doi.org/10.5281/zenodo.583896.

1. Introduction

Key frame is a frame in a video that provide the best summary of the video content. Many videos are emerging on the internet, due to complexity of these videos it becomes more critical to search and catch required video quickly and effectively. In the series of video frames key frame can be used to describe the key image of a video shot. Key frame extraction is essential for many video processing applications like video summary, video analysis, video organization and video compression. Key frames are used to evaluate the value of information of a video within a short time for decision making.

Key frame extraction algorithms should attain some requirements. First, the key frames should represent the whole video content without missing the important information and second, these key frames should be non-repetitive, in terms of video content information. Finally, it is highly desirable that the number of key frames should be specified automatically without any prior knowledge about the video content. In the process of video retrieval video frame is referred to as a static image which is the basic unit of video data. Video frame sequence is defined as a set of

frame with some time interval. The individual frames are separated by frame lines. A video stream will contain frames, shots, scenes and sequences. Video shots are generated by physically related frame sequences. Then related scenes are combined to form sequences. Based on the complexity of the content in video shot one or more number of key frames can be extracted from a unique shot. The extracted key frames must summarize the characteristics of the video, all the key frames in a time sequence gives visual summary of the video. There are some great repetitions among the frames in same shot, so only those frames that best reflect the contents of shot are selected as key frames to represent the shot. The key frames which are extracted should contain as much important content of the shot as possible and avoid as much redundancy as possible. In the process of video indexing and video retrieval it includes the analysis of structure for detecting shot boundaries, extracting key frames and segment scenes; feature extraction for object features and motion features; video annotation for building a semantic video index; query is returned for searching the desired video in video database using the index and similarity features; and video browsing and feedback for response to a query returned to browse in the form of video summary and subsequent search results are optimized [1-2].

2. Literature Review

Literature classifies key frame extraction based on the following approaches.

2.1. Improved Heirarchical Clustering Algorithm

Huayong Liu, Huifen Hao proposed a method for key frame extraction based on improved hierarchical clustering algorithm (IHCA). In this method with they use the characteristics of image information entropy, which tells how much information there is in an event to measure the degree of similarity between two frames, if the similarity attains certain value the frames are merged into the same cluster. Then they extracted clustering center as key frames.

In this approach to extract key frames along with the improved hierarchical clustering algorithm another algorithm which is k-means algorithms is proposed. IHCA is used to obtain an initial clustering result and k-means is conducted to optimize the initial clustering result and obtain the final clustering result. Finally, the center frame of each clustering is extracted as key frame.

In their approach, they Assumed that video shot boundary detection is already done and some video shots have been identified. Starting from video data, extracting the feature of each frame and calculating the interframe similarity by using Euclidean distance formula. For initial clustering result, they conducted the hierarchical clustering algorithm and to optimize initial clustering result they conducted k-means algorithm. Finally, they output the key frames which are eligible. If the information entropy of certain image is larger, the image contains more information [3-4]. Here each image is divided into B blocks for 256 grey-scale image and information entropy of each block is calculated by equation (1)

$$H = -\sum_{i=0}^{255} p(x_i) \log_2 p(x_i) \qquad (1)$$

Where, i is grey scale of each image
$p(x_i)$ is Probability of x_i

Feature vector of each image can be done by using equation (2)

$$F = \{H_j | j = 1,2,\ldots,B\} \quad (2)$$

Where, H_j = information entropy of block j-th
F = Feature vector

Steps of proposed algorithm:

Step 1: Calculate interframe similarity. Set each frame as a cluster and calculate the distance between ever two clusters using Euclidean distance formula i.e., equation (3)

$$d(F_a, F_b) = \left(\sum_{r=0}^{N} (F_a[r] - F_b[r])^2 \right)^{1/2} \quad (3)$$

Where, $N = \frac{n(n-1)}{2}$, n denotes frame number of the video
F_a= Feature vector of image 'a'
d (F_a,F_b) = distance between image a and b.

Step 2: Assuming the number of key frames are K, we can calculate the average M and variance V of distance vector by using equation (4) and (5)

$$M = \frac{1}{N} \sum_{i=0}^{N-1} D_i ; \quad (4)$$

$$V = \frac{1}{N} \sum_{i=0}^{N-1} (M - D_i)^2 ; \quad (5)$$

Where, D_i is distance vector
In the proposed work K is equal to number of frames whose distance d(F_a,F_b)> M+ 2* √v

Step 3: Merge the nearest clusters into a new one and then recalculate the distance between the new cluster and the old ones.

Step 4: Repeat step 3, until the number of cluster is K

Step 5: Calculate the mean value of each cluster. Assign each frame to nearest cluster per minimum distance. Recalculate the clustering center.

Step 6: Repeat step 5, until the cluster objects do not change. Select the clustering center as key frames.

Experimental results of IHCA:

They made lot of experiments on videos which have different characteristics and selected five representative videos for verifying the effectiveness of proposed algorithm. For testing the

algorithm Precision Ratio(PR) and Recall Ration(RR) are used and they can be found by using the formulas.

$$PR = \frac{N_c}{N_c + N_f} * 100\% \quad (6)$$

$$RR = \frac{N_c}{N_c + N_m} * 100\% \quad (7)$$

Where, N_c denotes number of correct key frames
N_f denotes number of false key frames
N_m denotes number of missing key frames

The following tables shows the comparison between a widely-applied algorithm for key frame extraction namely Sequence Difference Histogram Algorithm (SDIF) and the proposed algorithm i.e., Improved Hierarchical Clustering Algorithm.

Table 1: Results for using SDIF

Video name	Total frames	Key frames	PR (%)	RR (%)
Ad	1235	21	62.1	65.5
Cartoon	720	27	68.6	70.4
Movie	1087	31	63.4	64.1
News	1802	21	77.4	74.8

Table 2: Results for using improved hierarchical algorithm

Video name	Total frames	Key frames	PR (%)	RR (%)
Ad	1235	21	90.5	89.2
Cartoon	720	27	89.7	88.9
Movie	1087	31	86.5	87.8
News	1802	21	86.9	90.4

As shown in above table, IHCA improves the PR to 86% and RR to 87% showing the best results than SDIF.

The main advantage of proposed method is that the redundancy of the algorithm is relatively low and is effective. The drawback lies in evaluation criteria of key frame extraction are not perfect and calculation of image formation entropy is expensive.

2.2. Key frame extraction using Discrete Cosine Transform

In conventional text based video retrieval the query is difficult to form and process with the textual properties of the video in database. A technique which is extraction of only essential frames using Discrete Cosine Transform (DCT) is proposed to overcome this problem. This type

of method include the extraction of only selected important frames which contain maximum video information.

The whole proposed method is divided into two categories.

a) Computation of frame difference
b) Selection of key frames

2.2.1. Computation of Frame Difference

Here three red, green and blue planes are considered to extract the features. DCT is applied on each plane to get the DCT equivalent of each consecutive frame. DCT block computes the related DCT of each channel in m x n input matrix. In m x n matrix the 2D DCT compresses all the energy information of image and concentrates it in a few coefficients located in the upper left corner of the resulting real valued m x n matrix [5]. The features from video frames are extracted to three separate red, green and blue planes. Individual feature vectors are then transferred using DCT [6-7]. Then their absolute difference is further considered to identify key frames. These differences are used with threshold value to get key frames.

$$d(f_i, f_{i+1}) = abs(d_r(f_i, f_{i+1}) + d_g(f_i, f_{i+1}) + d_b(f_i, f_{i+1})) \quad (1)$$

Where d_r,d_g,d_b are consecutive differences in three different planes.

2.2.2. Selection of Key Frames

Here cumulative difference value is used, it shows the exact change in transition with their differences [8-9-10-11]. Using DCT and dividing total size by the factor of 2,4 and 8 and here the approximate value of total difference is reduced so that total computations are minimized. Energy is concentrated at certain points. So, there is no need to read complete video information. Threshold value is calculated by constant *p* and standard deviation *std.*

***Threshold* = p*std**

If the difference is greater than the threshold value, then respective frame is considered as key frame.

Implementation steps:

Here input video V is read frame by frame from 1^{st} to n^{th} frame.
Step 1: Read an input video with N frames with n x n size.
Step 2: Read consecutive i^{th} and $(i+1)^{th}$ frames.
Step 3: Resize each frame as power of 2.
Step 4: Apply DCT on each frame with 25% , 6.25% and 1.56 % coefficient.
Step 5: Find the consecutive frames.
Step 6: Compute *std* and threshold value.
Step 7: If difference >threshold
Write $(i+1)^{th}$ frame as key frame
Else
Discard it.
Step 8: Repeat this for all frames till the end of the video. At the end key frames are collected.

The advantage of this method is that total number of frames can be reduced by inclusion of only key frames and removal of non-significant key frames and overcomes the problem in searching and retrieving the videos due to more keywords stored in database. The drawback of this method is that for 25% DCT coefficient average completeness shows better performance, whereas for 6.25% and 1.56% the average completeness is comparatively lesser.

Key Frame Extraction from Surveillance Video

It is based on intensity of motion energy. The main aim of this approach is to extract meaningful key frames efficiently. Motion energy is computed between the frames and used to extract frames with maximum motion energy. The proposed framework consists of two modules, the first module is used to reduce the size of input video using global similarity feature and the second module extracts the desired key frame.

Reduction of video:

In a video, there are significant number of near similar frames. These are detected and eliminated on the bases of peak signal to noise ratio(PSNR). This PSNR value is calculated by

$$PSNR = 10 \log_{10} I^2_{MAX}$$

I_{max}-maximum possible intensity value of a pixel in the image

MSE- Mean Square error

If PSNR>30 then no difference is noticeable [12].

If PSNR<30 the difference between is noticeable.

So, by eliminating the frames whose PSNR>30, the size of the video significantly reduces.

Algorithm:

Input: video

Step 1: r=1

Step 2: Calculate PSNR between F_r and all successive frames till F_i is encountered for which PSNR<25

Step 3: Eliminate all the frames between F_r and F_i

Step 4: r=i

Step 5: If F_r is not the last frame, go to step 2

Step 6: Stop.

Output: visually dissimilar frames in the same sequence as in the input video shot.

Extraction of Desired Key Frames

Motion is an intrinsic attribute of video. Hence, the frames which have significant change based on motion are the frames which convey some semantic information. If an object moves in a frame, then the coordinates of the points lying on that object change in the next frame of same video. The change in coordinates of a point is called motion vector of that point and this is used to measure the motion energy of that point. The optical flow can be used to detect motion vectors [13]. The motion vectors of the points lying on the same edge are same. The corner points are detected by Harris Corner Detector [14].

Algorithm:

Input: Reduced video shot obtained from previous modules.
Step 1: Let r=1
Step 2: Add F_r to the set of extracted key frame.
Step 3: Apply Harris Corner detector algorithm to frame F_r to get the corner points. Let the number of corner points obtained be k.
Step 4: Using the algorithm for optical flow, find the value of j, such that
Intensity (F_{rj})>Intensity (F_{ri}) for all r<i<j where

$$Intensity(F_{rq}) = \Sigma_{m=1}^{k} \sqrt{x_m^2 + y_m^2}$$

X_m and Y_m denotes two components of motion
Vector m along x and y axis
Step 5: If F_r is not the last frame, go to step 2
Step 6: Stop
Output: Desired key frame.

The main advantages of proposed method are time complexity is reduced by detecting motion vectors and it overcomes shaky and redundant key frames extracted by IBM Multimedia Analysis and Retrieval System [15-16]. The drawback of this method is that it gives quality results only for uncompressed video shots i.e., without camera motion.

References

[1] Sudeep D.T Hepade Ajay A. Narvekar,Ameya V. Nawale, "Color Content Based Video Ret rieval Using Discrete Cosine Transform Applied on Rows and Columns of Video Frames with RGB Color Space", ISO 9001:2008 Certified, International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 11, May 2013
[2] Costas Panagiotakis, Anastasios Doulamis and Georgios Tziritas, "Equivalent Key Frames Select Ion Based on Iso-Content Principles", article submit ted to IEEE Trans. On Circuits Systems for Video Technology, 2008.
[3] Sun, Lina, and Yihua Zhou, "A key frame extraction method based on mutual information and image entropy," IEEE conference, 2011 International Conference on Multimedia Technology (ICMT'11), 2011:35-38.
[4] Angadi, Shanmukhappa, and Vilas Naik, "Entropy Based Fuzzy C Means Clustering and Key Frame Extraction for Sports Video Summarization," IEEE conference, 2014 Fifth International Conference on Signal and Image Processing (ICSIP'14), 2014: 271-279.
[5] H.B.Kekre, Tanuja K. Sarode , Sudeep D. Thepade, " Image Ret rieval using Color-Texture Features from DCT on VQ Code vectors obt ained by Kekre"s Fast Codebook Generation", ICGST -GVIP Journal, Volume 9, Issue 5, September 2009, ISSN: 1687-398X.
[6] H. B. Kekre, Dr. T. K. Sarode, Prachi J. Natu, Prachi J. Natu, "Performance Comparison of Face Recognition using DCT and Walsh Transform with Full and Partial Feature Vector against KFCG VQ Algorihm", Proceedings published by International Journal of Computer Applicat ions® (IJCA).2011.
[7] Sudeep D. T hepade, Ashvini A.T onge,"An Opt imized Keyframe Exat raction For Detection of Near Duplicates In Content Based Video Retrieval", presented in IEEE conference ICCSP "14, Tamilnadu,3rd to 5th April, 2014.

[8] Sudeep D. T Hepade, Ashvini A. Tonge," An improved approach of key frame extraction for Content Based Video Retrieval", presented in CPGCON, National Symposium Post Graduate Conference in Computer Engineering, March 28th and 29th, 2014.

[9] Sanjoy Ghat ak, Debotosh Bhattacharjee, "Extraction of Key Frames from News Video Using EDF, MDF AND HI Method for News Video Summarization", ISO 9001:2008 Certified, International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 12,June 2013.

[10] Jasmeet Kaur, Rohini Sharma, "A Combined DWT -DCT approach to perform Video compression base of Frame Redundancy", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 9, September 2012 ISSN: 2277 128X.

[11] Guozhu Liu, and Junming Zhao, "Key Frame Extraction from MPEG Video Stream", Proceedings of the Second Symposium International Computer Science and Computational Technology (ISCSCT'09), Huangshan, P. R. China, 26-28, Dec. 2009, pp. 007-011, ISBN 978-952-5726-07-7 (Print), 978-952-5726-08-4.

[12] C. C. Lee, H. C. Wu, C.S. Tsai and Y.P Chu, "Adaptive lossless stenographic scheme with centralized difference expansion," Pattern Recognition, vol. 41, no. 6, pp. 2097–2106, 2008.

[13] Jie-Ling Lai and Yang Yi, "Keyframe extraction based on visual attention model," Journal of Visual Communication and Image Retrieval, vol. 23, no. 1, pp. 114–125, 2012.

[14] Gary Bradski and Adrian Kaebler, "Learning Open CV", First ed., vol.1. OReilly Media, Inc. Sebastopol, 2008, pp. 316–337.

[15] Pascal Kelm, Sebastian Schmiedeke, and Thomas Sikora, "Featurebased video key frame extraction for low quality video sequences," in Proceeding of IEEE conference on Image Analysis for Multimedia Interactive Services, pp. 25-28, 2009.

[16] Iron-horse. (2011, June 21). IBM Multimedia Analysis and Retrieval System (IMARS) [online]. Available: https://www.ibm.com/developerworks/community/groups/service/html/communityview?communityUuid=7dc62548-8bc8-42c4-b2e9-150dde7c649a

*Corresponding author.
E-mail address: satish.rmgm@gmail.com