INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

CONTENT BASED VIDEO RETRIEVAL USING KEY FRAME EXTRACTION ALGORITHM

ANURADHA B. RATHOD[1], P. S. GAWANDE[2]

1. M.E. Student, Department of Computer Science and Engineering, RSCE, Maharashtra, India.
2. Department of Electronics and Telecommunication, RSCE, Maharashtra, India.

Accepted Date: 05/03/2015; Published Date: 01/05/2015

Abstract: Key frame extraction has been recognized as one of the important research issues in video information retrieval. Although progress has been made in key frame extraction, the existing approaches are either computationally expensive or ineffective in capturing most important visual content. Video summarization aimed at reducing the amount of data that must be examined in order to retrieve the information desired from information in a video, is an essential task in video analysis and indexing applications. We propose an innovative approach to the selection of representative (key) frames of a video sequence for video summarization In this paper, we discuss the importance of key frame selection; and then briefly review and evaluate the existing approaches. To overcome the shortcomings of the existing approaches, we introduce a new algorithm for key frame extraction.

Keywords: Video summarization, video skimming, lecture videos, key frames, Video Retrieval

Corresponding Author: MS. ANURADHA B. RATHOD

Access Online On:

www.ijpret.com

How to Cite This Article:

Anuradha B. Rathod, IJPRET, 2015; Volume 3 (9): 1608-1616

PAPER-QR CODE

INTRODUCTION

Digital video has become a popular storage and exchange medium due to the quick development in recording technology, recovered video compression techniques and high-speed networks in the last few years. Therefore audiovisual recordings are used more and more frequently in e-lecturing systems. A number of universities and research institutions are taking the opportunity to record their lectures and publish them online for students to access independent of time and location. As a result, there has been a huge increase in the amount of multimedia data on the Web. Therefore, for a user it is nearly impossible to find desired videos without a search function within a video files. Even when the user has found related video data, it is still difficult most of the time for him to judge whether a video is useful by only glancing at the title and other global metadata which are often brief and high level. The relevant metadata can be automatically gathered from videos by using appropriate analysis techniques. They can help a user to find and to understand lecture contents more efficiently, and the learning effectiveness can thus be recovered. Traditional video retrieval based on visual feature extraction cannot be simply applied to lecture recordings because of the homogeneous scene composition of lecture videos. Now people tend to produce lecture videos by using multi-scenes format by which the speaker and his presentation are displayed synchronously. Video summarization aimed at reducing the amount of data that must be examined in order to retrieve a particular piece of information in a video is consequently an essential task in applications of video analysis and indexing.

Summarization utilizing moving images is called video skimming; the product is similar to a video trailer or clip. Both approaches must present a summary of the important events recorded in the video. We focus our attention here on the creation of a visual summary using still images, called key frames, extracted from the video. Although video skimming conveys graphic, motion and, where used, audio information, still images can summarize the video content in more rapid and compact way users can grasp the overall content more quickly from key frames than by watching a set of video sequences. Key-frame based video representation views video abstraction as a problem of mapping an entire segment to some small number of representative images. The extraction of key frames must be regular and content based so that they maintain the salient content of the video[1]. Besides providing video browsing capability and content description, key frames act as video “bookmarks” that designate interesting events captured supplying direct access to video sub-sequences. Key frames, which visually represent the video content, can also be used in the indexing process, where we can apply the same indexing and retrieval strategies developed for image retrieval to retrieve video sequences [2].

Low level visual features can be used in indexing the key frames and thus the video sequences to which they belong. The use of low level features in the indexing process should not be considered as a less powerful strategy; high level features with different levels of semantic can be derived from them.

2. KEY FRAME EXTRACTION SYSTEM

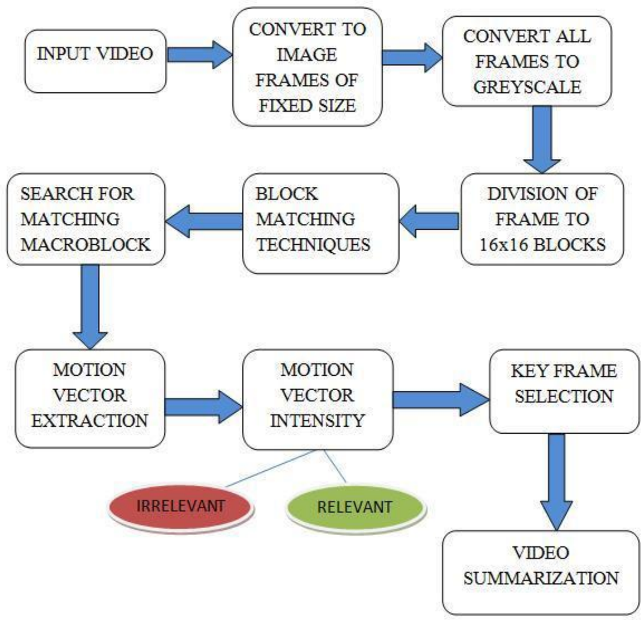

Fig 1: System Architecture diagram

The aim of the algorithm is to provide a summarized video which produces a gist of the original video without losing semantics of the video. Fig1 provides the blueprint for our process. The initial process involves converting the input video into frames [1][2][3]. After which the frames are grey scaled. Later, each frame is further divided into a fixed number of macro blocks (16x16 in this case) which facilitates the use of an individual macro block as comparison units. The first macro block of the first frame is then compared with the macro blocks in the second frame to search for the closest match to the original macro block. Comparing all macro blocks in the second frame is a tedious process and hence an astute method of selection of macro blocks is required which gives the correct match yet saves processing time [4].

This is implemented with the use of block matching algorithms which form the crux of this system. Each block matching algorithm specifies which blocks are to be compared and in what order. Once a block of the first frame is matched with the block of the second frame, the motion activity descriptor of the block can be established. This process is then repeated for each block of the first frame, and sum of all such motion descriptors is considered to produce the cumulative motion descriptor between the two frames. Such a cumulative motion descriptor is obtained between each pair of consecutive frames. These motion descriptors are then compared to categorize them into irrelevant and relevant. The motion descriptors signify the amount of motion present between two consecutive frames. Absence of motion signifies no or minimum difference between two frames, whereas a high motion descriptor signifies a vast difference between two frames and thus leads to the conclusion of them being key frames. Summation of all such key frames lead to the formation of the summarized video[5].

2.1 New key frame extraction approach

The proposed key frame extraction algorithm functions as shown in Fig2. The algorithm employs the information obtained by a video segmentation algorithm to process each shot. We have developed a prototypical segmentation algorithm that for the moment detects only abrupt changes and fades, since these are the most common editing effects. For abrupt changes we have implemented a threshold based algorithm coupled with a frame difference measure computed from histograms and texture descriptors. The same frame difference measure is used for the key frame extraction algorithm [5]. To detect fades we have implemented a modified version of the algorithm proposed. The results obtained by these algorithms are submitted for evaluation to a decision module which gives the final response. This allows us to cope with conflicting results, or with groups of frames that are not meaningful, such as those between the end of a fade out and the start of a fade-in, increasing the robustness of the detection phase. A gradual transition detection algorithm is currently being developed, and integrated in a similar manner [6][7][8].

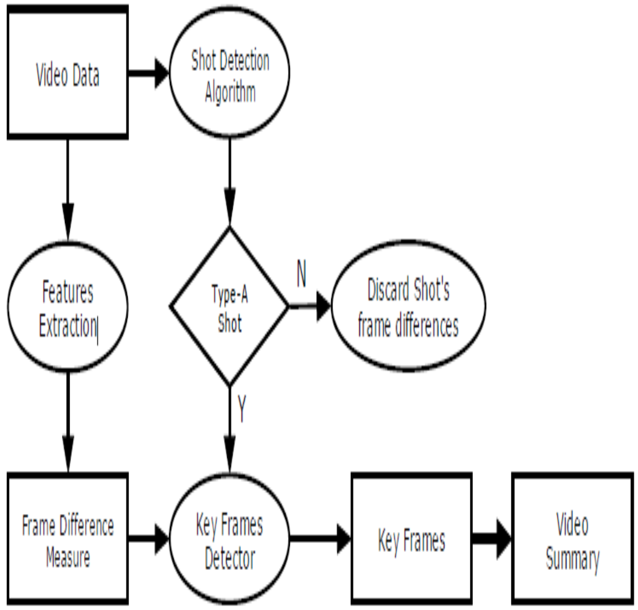

Fig 2: The Key Frame Extraction Algorithm

We distinguish two types of shots: "informative" shots (type A) and "uninformative" shots (type B). Type B shots are those limited by: a fade-out followed by a fade-in effect, a fade-out followed by a cut or a cut followed by a fade-in. If we were to extract key frames from these shots the resulting set of frames would contain uniformly colored images that are meaningless in terms of the information supplied. Key frames are extracted from type A shots of only.

2.2 Algorithms of key frame extraction

The video sequence constructed using video handover has to be sampled to extract key frames. Our intention is to create a summary that is both complete and compact. To achieve this, we have to minimize the number of redundant key frames while ensuring that important key frames are not missed. Table1. Summarizes the algorithms we implemented for key frame extraction. In all entries, *T* is a constant time interval. Adaptive spatio-temporal sampling attempts to sample more key frames when the person makes a lot of movement without changing to a different camera view. We designed and conducted an experiment to evaluate the performance of these algorithms. This has been described in detail in [9]. Adaptive spatio-temporal sampling performed much better than the other algorithms and 15 sec was found to be the optimal value for the parameter T. The key frames extracted using this algorithm were compared with the average key frames that the subjects who took part in the evaluation believed to be the best set of key frames for the corresponding sequence.

Sampling algorithms	Condition for sampling key frame
Spatial	At every camera change
Temporal	Once every T second
Spatio-temporal	Sampling a key frame • At every camera change • If T second elapsed with no camera Change after the previous key frame
Adaptive spatio-temporal	Sample a key frame • At every camera change • If T second elapsed without camera Change where: t = T(1-n/20) if 1<= n<=10 t = T/2 if n>10 (n= No. of footsteps since last key frame)

Table1: Algorithm steps for key frame extraction

2.3 Key frame selection

The key frame selection algorithm that we propose dynamically selects the representative frames by analyzing the complexity of the events depicted in the shot in terms of graphic changes. The frame difference values initially obtained are used to construct a curve of the cumulative frame differences which describes how the visual content of the frames changes over the entire shot, an indication of the shot's complexity: sharp slopes indicate significant changes in the visual content due to a moving object, camera motion, or the registration of a highly dynamic event. These cases must be taken into account in selecting the key frames to include in the shot summary [5]. They are identified in the curve of the cumulative frame differences as those points at the sharpest angles of the curve (curvature or corner points). The key frames are those corresponding to the mid points between each pair of consecutive curvature points. To detect the high curvature points we use the algorithm proposed. The algorithm was originally developed for shape analysis in order to identify salient points in a 2D shape outline. The high curvature points are detected in a two-pass processing. In the first pass the algorithm detects candidate curvature points. The algorithm defines as a "corner" a location where a triangle of specified size and opening angle can be inscribed in a curve[10][11].

3. MATHEMATICAL MODEL

A video frame is a solo picture or motionless shot that is shown as fraction of a larger video. Key-frame is the frame that can represent the content of a section of a video. The key frames are the frames which has drastic change than the previous frame. In proposed system, the key-frame extraction is done with the help of color histogram.

The algorithm of color histogram is given below:

1. Set„ d as threshold and initialize the algorithm.

2. Select the first frame and set it as a key frame.

3. If no frame is captured go to 6.

4. Start the comparison of current captured frame with the selected key frame.

5. Calculate the difference between them

a. If the *d* > difference then select the current frame as next key frame.

b. If the *d* < difference then go to c.

c. Go to next frame.

6. End.

4. RESULTS

The different sequences of videos are scene from table-2. It summarizes some of the characteristic of the algorithms tested. Based on that for an amount of interval obtain key frame from above mention steps of algorithm.

Sr no	Parameters	Specifications
1	Video length (assumption)	60 sec
2	Frame rate	25 Fps
3	Resolution	352 * 25=101376
4	Number of frames	60*25=1500
5	Bit rate (compressed)	60*512kbps=30720kbps=30.720Mbps

Table 2: Experimental setup

Sr No	Threshold (k)	Actual key frames	Detected key frames	Transmission bandwidth (Mbps)
1	K=1	10	19	19*512kbps=9.7
2	K=1	10	15	15*512kbps=7.6
3	K=1	10	10	10*512kbps=5.1
4	K=1	10	7	7*512kbps=3.58

Table3: Result Analysis

Here the key frames are changing according to threshold value, for higher threshold we get less key frames and vice-versa, so based on application we can make a relevant threshold so that we can save the transmission bandwidth accordingly.

5. CONCLUSION

This paper discusses the proposed key frame extraction algorithm. However, the key frames of a particular event are often taken into account. The experimental result shows that the key information of live video are extracted efficiently resulting in saving transmission bandwidth The differences between pairs of consecutive frames of a video sequence, the algorithm determines the complexity of the sequence in terms of changes in visual content as expressed by different frame descriptors. Similarly measures are computed for each descriptor and combined to form a frame difference measure

REFERENCES

1. Haojin Yang and Christoph Meinel, Member, IEEE "Content Based Lecture Video Retrieval Using Speech and Video Text Information," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process., 2014, val.7 no.2

2. G. Ciocca1, R. Schettini "Key frame extraction for video summarization using motion activity descriptors".

3. Gamhewage C. de Silva, Byoungjun Oh, Toshihiko Yamasaki, Kiyoharu Aizawa "Experience retrieval in a ubiquitous home" CARPE'05, November 11, 2005, Singapore.

4. Kavitha Mamindla, Dr. V. Padmaja, and R. Neeharika "Key Frame Extraction Of Live Video Based On Optimized Frame Difference Using Cortex-A8" International Journal of Electronics and Communication Engineering & Technology (IJECET), ISSN 0976 – 6464(Print), ISSN 0976 – 6472(Online) Volume 4, Issue 5, September – October (2013).

5. Yueting Zhuangy , Yong Rui Thomas S_ Huang and Sharad Mehrotra “An adaptive key frame extraction using unsupervied clustering,” in ARL Cooperative Agreement No. DAAL01-96-0003and in part by NSF/DARPA/NASA Digital Library Initiave Program under Cooperative agreement No94-11318

6. D. Lee and G. G. Lee, “A korean spoken document retrieval system for lecture search,” in Proc. ACM Special Interest Group Inf. Retrieval Searching Spontaneous Conversational Speech Workshop, 2008.

7. J. Glass, T. J. Hazen, L. Hetherington, and C. Wang, “Analysis and processing of lecture audio data: Preliminary investigations,” in Proc. HLT-NAACL Workshop Interdisciplinary Approaches Speech Indexing Retrieval, 2004, pp. 9–12.

8. Zhang H J,Wu J H, Zhong D et al, “An integrated system for content-based video retrieval and browsing”. Pattern Recognition, 1997, 30: 643-658.

9. Yeung M M, Liu B, “Efficient matching and clustering of video shots”. Proceedings of IEEE ICIP, 1995: 338-341.

10. Zhang Y J, Lu H B, “Hierarchical video organization based on compact representation of video units”. Proc. Workshop on Very Low Bitrates Video’99, 1999: 67-70.

11. Calic J, Izquierdo E, “Efficient key-frame extraction and video analysis”. Proceeding of the International Conference on Information Technology: Coding and Computing (ITCC’02),2002: 28-33.

12. He Xiang, Lu Gung-ho, “Algorithm of key frame extraction based on image similarity”. Fujian Computer, 2009, 5: 73-74.