Video Summarization Using Key Frame Extraction

Tayyab Tariq
tayyabt@gmail.com

Naveed Ejaz
naveed.ejaz@nu.edu.pk

Department of Computer Sciences
FAST NUCES, Islamabad, Pakistan

Abstract

The nature of video data does not make it a good candidate for conventional retrieval, indexing and storage techniques due to redundancy. Video summarization is a method to reduce this redundancy. In this paper we present a technique for video summary generation using key frame extraction called Key Frame Extractor (KFE). KFE uses 2D auto-correlation, color histogram comparison and moment invariants for key frame extraction. An adaptive formula is used to make KFE partially tolerant to lighting condition changes. KFE allows a tradeoff in computation, memory complexity and accuracy of key frame extraction. The video summarization results compare very well to the TRECVID 2007 benchmark and CMU's submission to TRECVID 2007.

1. Introduction

Recently there has been an explosion of video data on the internet. However, techniques for its indexing, retrieval, and storage have not advanced at the same pace. This is due to the substantially different nature of video data which is not suited to conventional retrieval, indexing and storage techniques [1]. The nature of video itself provides a solution to this problem. Videos contain a lot of redundant information. This redundant information can be removed to make video data more suitable for retrieval, indexing and storage.

The first step in video summarization is to cut down the video into smaller units called shots. These shots are semantically and visually homogeneous. The second step is to use these shots to extract key frames. The technique presented in this paper is developed to addresses the second problem. However, it has been tested on videos with multiple shots with promising results.

For KFE the problem is defined as follows. Find the minimal set of key frames that cover all significant events or maximize the number of key frames while minimizing redundancy of information in these key frames. These problem definitions are information theoretic in natures and thus a quantitative evaluation of results is not easy. The technique described in [2] is used for evaluation (more details in the results section).

The rest of this paper discusses the related work, their short comings, the proposed design for KFE and its results.

2. Related Work

One of the simplest possible approaches to key frame extraction is to choose fixed frame index or indexes as the key frames. Such an approach is used by Otsuji & Tonomura (1993) in which they choose the first frame as the key frame given that the video is divided into shots [3]. Rui, T.S., & Mehrotra (1998) used a similar technique where they used the first and the last frame as the key frame [4].

Another approach is to comapre consecutive frames in video. Hanjalic & R.L. (1996) compare the difference in color histograms of consequtive frames with a threshold to obtain keyframes [5]. A similar technique is used in KFE.

Towards the more complicated end, techniques based on clustering have also been developed. Girgensohn & J., (2001) clustered visually similar frames and used constraints on key frame positions in the clustering process [6]. Gong & Liu (2000) use single value decomposition (SVD) in their clustering. The video frames are time sampled and features are extracted. The refined feature space obtained from SVD is clustered and a key frame is extracted from each cluster [7].

Chitra A. & Sanjeev (2008) developed a technique in which they used multiple visual features for key frame extraction [8]. They used a weighted descriptor of edge direction and wavelet coefficients to obtain results comparable to Niblack, Yue, Kraft, Amir, & Sundaresan, (2000) [9].

The problem with the above mentioned techniques is that they need the number of key frames, the interval between them or some similar feature to be specified by the user. The specification of number of key frames or the interval does not guarantee a loss less or redundancy free video summary. Moreover, the techniques are either too complex [6, 7] or too naïve [3, 4]. The simpler techniques heavily compromise the quality of key frame extraction and the more sophisticated techniques are computationally very expensive. KFE provides a good tradeoff between complexity and quality of results.

3. Design

A technique based on comparison of frames is developed in KFE. KFE compares the current frame with the last key frame instead of comparing consecutive frames. This allows a summary with lesser redundancy as consecutive frames are unlikely to have large differences. KFE also divides the frame into sections and compares the corresponding sections. To compare two frames KFE utilizes three comparison measures; 2D auto-correlation, color histogram difference and moment invariants. We pass the result of a comparison measure to the adaptive formula that combines them with the results of past comparisons. The result of the adaptive formula is compared with a threshold to obtain the confidence measure expressed by the said measure. A voting mechanism in used to combine the results of each measure. Any frame that gets more than τ votes is declared to be a key frame. The value 2 was found to be a good choice for τ.

The three measures, the adaptive formula and the voting mechanism are explained below.

3.1. Correlation Comparison Measure

The 2D auto-correlation between corresponding sections of frames being compared is calculated. The values from all sections and color channels are combined using mean function to obtain the result of this comparison measure.

3.2. Histogram Comparison Measure

The color histograms of the corresponding sections of the frames being compared are subtracted and the absolute difference is normalized. This mean of these values is used as the result of the histogram comparison measure.

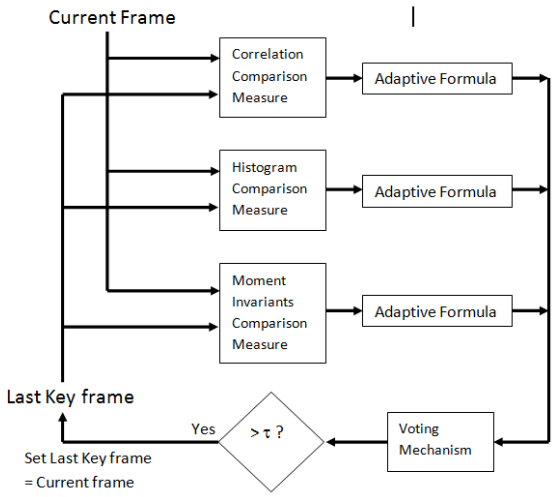

Figure 1: KFE Framework

3.3. Moment Invariants Comparison Measure

The moment invariants method is used to compute the 9 moments from each section of a frame (3 for each color channel). The city block distance between the moments is computed. The mean of these measures is used as the result of this comparison measure.

3.4. The adaptive formula

The adaptive formula combines the previous comparison results with the current comparison result to give a smoother output function. This function is insensitive to small changes in lighting conditions and helps reduce redundant key frames. A weight α is given to the previous results whereas 1-α to the current result, where α is an input parameter. This parameter is discussed in detail in the results section.

3.5. Voting Mechanism

KFE uses a voting mechanism in which the vote cast by each comparison measure is a real number. One measure can cast a vote with a value greater than 1. This vote value is obtained by comparing the result of the comparison measure with a threshold and computing the percent change from this threshold. The votes from different comparison measures are combined by addition.

4. Results

To evaluate the performance of KFE two sets of experiments were conducted. The first set of experiments was carried out to find out the optimal values for different thresholds. The second set of experiments were conducted to score the summaries generated by KFE and compare them to the TRECVID benchmark and CMU's submission to TRECVID 2007 using the method described in [2].

To determine the values of thresholds KFE was tested on 7 different videos containing a total to 8,846 frames. Each video was tested for a combination of values for; number of sections, threshold values and alpha value. The following table summarizes these inputs parameters. We use a combination of these parameters and the total number of combinations for one video is 3x3x4=36. Thus the total number of frames processed was 8,846*36=318,456.

Parameter Name	Values used
Number of sections	1 x 1 4 x 4 8 x 8
Comparison Measure Thresholds	0.75 0.50 0.25
Alpha (α)	0.75 0.50 0.25 0.00

Table 1: Input Parameter Values

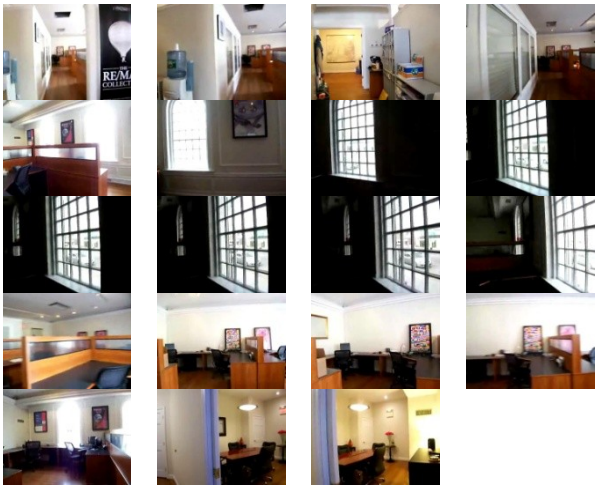

Figure 2: Key Frames extracted from Office Tour Video. Original Video Available at: http://www.youtube.com/watch?v=7qwp06xNuCs

Figure 2 shows key frames extracted from an office tour video. There are two main reasons for testing on this video (1) It is easy to assess the quality of key frame extraction results (2) The lighting conditions vary greatly in the video allowing testing the system to the limits. The system is able to handle mild changes in lighting conditions (such as the bottom left frame). However, extreme changes in lighting conditions define (such as those in the third row) define an upper limit for insensitivity to lighting changes.

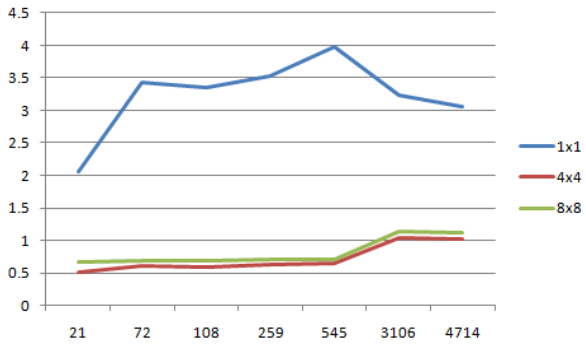

Figure 3: Number of frames (x-axis) against time taken for different number of frame sections.

The graphs in Figure 3 shows time taken (y-axis) against number of frames in videos for different frame divisions. The time taken is greatest for 1x1 (no frame division) because the time taken by moment invariants and correlation methods are not linear in number of pixels to process and the time taken increases rapidly with increase in the size of individual sections.

Figure 4 shows the change in fraction key frames (Number of key frames/Total frames) against different values of alpha (α). It is observed that giving equal weight to history and current result gives the best compression ratio. Increasing the value of alpha gives too much weight to result history causing significant events to be missed.

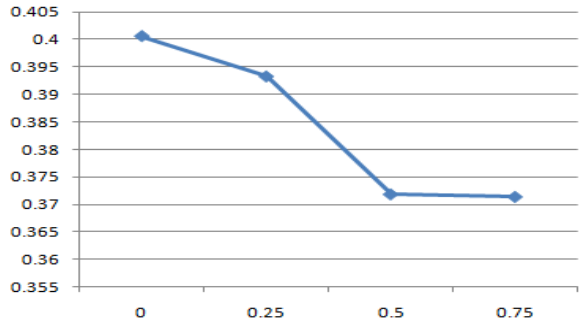

Figure 4: Alpha against fraction key frames

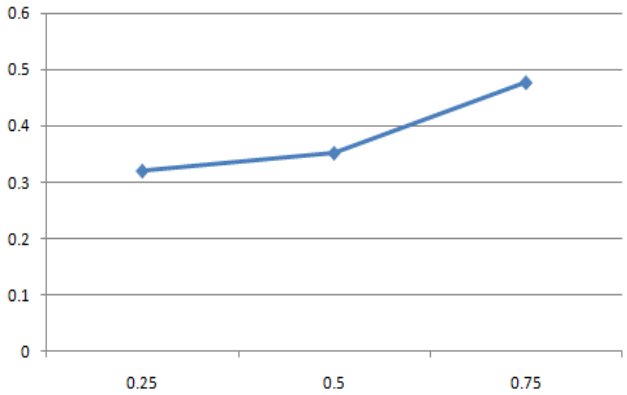

Figure 5: Threshold values against fraction key frames

The graph in Figure 5 shows the relation between fraction key frames and the threshold values. As expected a more succinct summary is obtained as the number threshold is made smaller.

Once the optimal values for the thresholds have been obtained we use the method described by P. Over, A. Smeaton & P. Kelly [2] to score the summaries generated by KFE. The first step is to create an output summary video. For this purpose a MPEG video is created using the key frame is with a frame rate five times as slow as the original video. 10 videos were used for this evaluation; the ground truth of significant events was created by a panel of ten student volunteers from FAST NUCES. Each video was then rated for;

- INclusion (fraction of significant events covered by key frames)
- Lack of REdundancy (non-redundant key frames/total key frames scaled to 5)
- Target time - Summary time (XD), reported in percent with a target time of 4% of original video (the 4% values was chosen arbitrarily in (2)).
- Time taken to spot significant events (TT) as a fraction of summary time.
- Play duration of summary as a percent of total video time (DU).

Table 2 compares the results of KFE with TRECVID 2007 benchmarks (CMU Base 1 & CMU Base 2) and CMU's submission to TRECVID 2007.

	CMU Base 1	CMU Base 2	CMU's Submission	KFE
IN	0.59	0.58	0.6	0.85
RE	3.52	3.50	3.62	3.83
XD	-0.15%	-0.06%	0.12%	0.37%
TT	1.7	1.65	1.75	2.71
DU	4.15%	4.06%	3.88%	3.63%

Table 2: KFE, CMU Base 1, CMU Base 2 and CMU Submission results (Mean) [2]

The values for XD and DU have been converted from seconds to percents.

References

1. *Applications of Video Content Analysis and Retrieval.* **Dimitrova, N., et al.** s.l. : IEEE Multimedia, 2002.

2. *The TRECVID 2007 BBC Rushes Summarization Evaluation Pilot.* **Over, Paul, Smeaton, Alan F. and Kelly, Philip.** Augsburg, Bavaria, Germany : International Workshop on TRECVID Video Summarization, 2007. 978-1-59593-780-3.

3. *Projection detecting filter for video cut detection.* **Otsuji, O. and Tonomura, Y.** s.l. : First ACM Int. Conf. Multimedia, 1993.

4. *Exploring video structure beyond the Shots.* **Rui, Y., T.S., Huang and Mehrotra, T.S.** s.l. : IEEE International Conference Multimedia Computing and Systems (ICMCS), 1998.

5. *A new key-frame allocation method for representing stored video stream.* **Hanjalic, A. and R.L., Langendijk.** 1996. 1st International Workshop on Image Databases and Multimedia Search.

6. *Time-constrained key-frame selection technique.* **Girgensohn, A. and J., Boreczky.** 2001. Multimedia Tools and Application.

7. *Generating optimal video summaries.* **Gong, Y. and Liu, X.** 2000. IEEE International Conference Multimedia and Expo.

8. *Summarization, A Novel Approach Towards Keyframe Selection for Video.* **Chitra A., Dhawale and Sanjeev, Jain.** s.l. : Asian Journal of Information Technology, 2008.

9. *Web-based searching and browsing of multimedia data.* **Niblack, W., et al.** 2000. IEEE International Conference of the Multimedia and Expo.