Dynamic Threshold in Clip Analysis and Retrieval

Satishkumar L Varma varmasl@yahoo.co.in
Department of Computer Engineering
PIIT, Mumbai University
Mumbai, 410206, India

Sanjay N Talbar sntalbar@yahoo.com
Department of Electronics and Telecommunication
SGGS IE&T,
Nanded, 431606, India

Abstract

Key frame extraction can be helpful in video summarization, analysis, indexing, browsing, and retrieval. Clip analysis of key frame sequences is an open research issues. The paper deals with identification and extraction of key frames using dynamic threshold followed by video retrieval. The number of key frames to be extracted for each shot depends on the activity details of the shot. This system uses the statistics of comparison between the successive frames within a level extracted on the basis of color histograms and dynamic threshold. Two program interfaces are linked for clip analysis and video indexing and retrieval using entropy. The results using proposed system on few video sequences are tested and the extracted key frames and retrieved results are shown.

Keywords: Video Clip, Key Frames Extraction, Entropy, Indexing, Retrieval.

1. INTRODUCTION

The growing prevalence of digital images and videos increases the need for effective and efficient searching techniques. In traditional databases the stored data is searched through alphanumeric matching. Each entry in the database has several key fields by which a query is matched. However images and video cannot be characterized by alphanumeric strings.

To save these video records we need large memory space. The videos are consists of key frames. So, instead of storing the whole video we can save only the key frames in which significant change is noted. This will save the memory space. Also accessing them will be easier. Key-frames are still images which best represent the content of the video sequence in an abstracted manner, and may be either extracted or reconstructed from original video data. Each key frame is usually described by low level features such as color, texture, sketch, shape, spatial constraints, annotation etc. A combination of features is always needed because there is no single best feature that gives accurate description in any general setting. In video databases each key frame is stored with its corresponding features. The features are chosen in the hope of capturing salient semantically information about the video key frame.

The research work done on video summarization methods has only focused on the uncompressed domain [1]-[6]. The shot detection techniques are also mentioned in [7][8]-[10]. The TRECVID Rushes task and data are described in Over et al. [9]. The scene analysis [11] and video content based frame and scene extraction [12] have been already worked out. Systems for summarizing unedited BBC footage were developed in TRECVID which was organized by NIST [13] [14]. A number of algorithms for extraction of spatio-temporal features from shots are reported in the literature [15]-[19]. Alternative approach was carried out by Vermaak et al. [20] to find the optimal set of key frames such that the frames are maximally distinct. Han and Yoon [21], describes a technique for key frame extraction using temporal sampling. In case of content based video indexing and retrieval, the various systems have been developed and reported, These systems include: the iMATCH [22], the IRMoment [23], the QBIC system [24], the Retrieval Ware system [25], the Virage system [26], the VisualSEEK and WebSEEK system [27], the Blobworld system [28], the Photobook system [29], the Mars system [30], the Video-Q [31], and the USC system [32].

The clip analysis is overviewed in section 2. In section 3, we describe the system of content-based video indexing and retrieval using entropy. The video retrieval interface and experimental results are discussed in section 4. The summary and conclusion of the paper is presented in Section 5.

2. OVERVIEW OF CLIP ANALYSIS

In this paper we describe the user interface that tries to compare frames in the Video and extract the key frames as well as do video segmentation by segmenting the video into shots using dynamic threshold depending on the number of key frames extracted. Further these key frames are used for indexing and retrieval of videos in the given database. The prepared system architecture is given in Fig. 1.

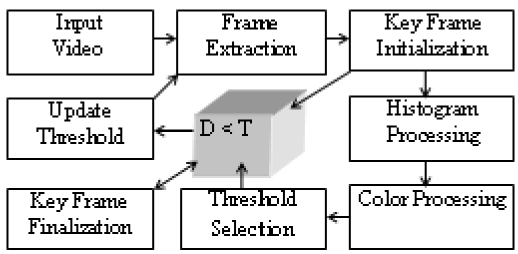

FIGURE 1: Key Frame Extraction Using Dynamic Threshold.

Shot segmentation: A shot is defined as the consecutive frames from the start to the end of recording in a camera. It shows a continuous action in an image sequences. There are two different types of transitions that can occur between shots, abrupt (discontinuous) also referred as cut, or gradual (continuous) such as fades, dissolves and wipes.

The cut boundaries show an abrupt change in image intensity or color, while those of fades or dissolves show gradual changes between frames. These transitions are defined as follows [33]:
• A cut is an instantaneous transition from one scene to the next and it occur over two frames
• A fade is a gradual transition between a scene and a constant image (fade out) or between a constant image and a scene (fade in)
• A dissolve is a gradual transition from one scene to another, in which the first scene fades out and the second fades in
• A wipe occurs as a line moves across the screen, with the new scene appearing behind the line.
There are different approaches used to detect the shot in a video and some are outlined here.

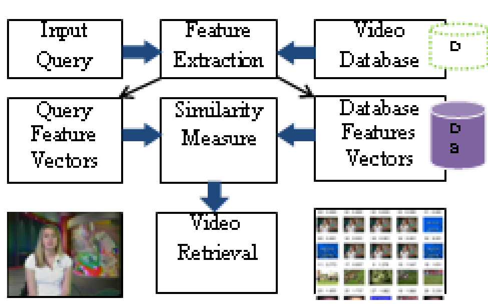

FIGURE 2: Key Frame Indexing and Retrieval.

3. VIDEO INDEXING AND RETRIEVAL

The indexing is carried out using color feature and texture entropy feature. Key frame indexing is carried out using color feature. The HSV color space is used to indexing the key frames to generate feature vectors. The block diagram is shown in Fig. 2. The indexing is also carried out using entropy. Entropy is a scalar value representing a statistical measure of randomness that can be used to characterize the texture of the input images. The value of entropy is also an invariant that is neither affected by rotation nor scaling. Entropy is defined as

$$H = -\sum_{i=1}^{\Omega} P_i \log(P_i) \tag{1}$$

3.1 Video Retrieval Measure

The video retrieval system is evaluated using two common measures, recall and precision. The recall measures also known as the true positive function or sensitivity, which corresponds to the ratio of correct experimental detections over the number of all true detections. It measures the ability of a system to present all relevant items:

$$recall = \frac{number\ \ of\ \ r}{number\ \ of\ \ rel} \tag{2}$$

The Precision measure defined as the ratio of correct experimental detections over the number of all experimental detections. It measures the ability of a system to present only relevant items.

$$precision = \frac{number\ of\ r}{total\ number} \tag{3}$$

3.2 Key Frame Retrieval

The query key frame is provided as input to the system. The query feature vector is compared with every key frame feature vector in the database under similarity measure process as shown in Fig. 2. The similarity measure is sorted in ascending or descending order. The close matches are found. The key frames also called as candidate key frames are displayed using user interface. The retrieved key frame belongs to a video in the database. So the respective video is retrieved.

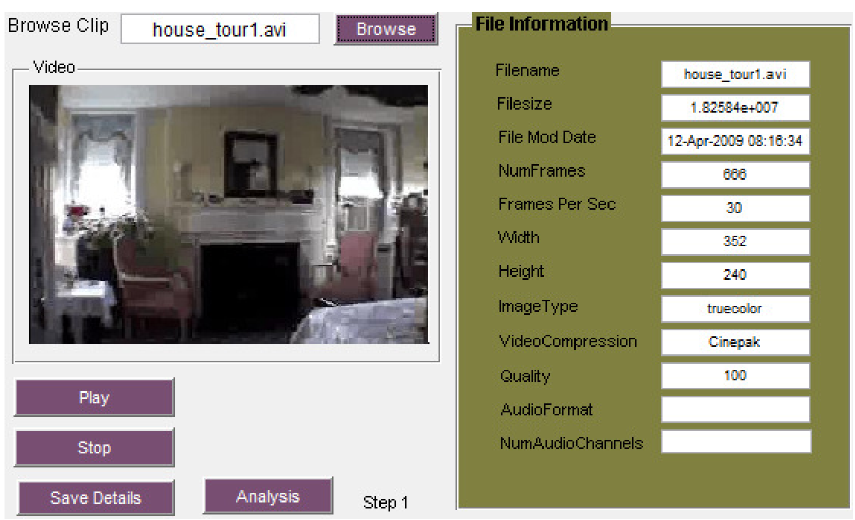

FIGURE 3: User Interface for Key Frame Extraction.

4. EXPERIMENTAL RESULTS

There are two different user interfaces. The user interface helps for clip analysis. The program interface is shown in Fig. 3. The interface also displays the information about the video being browsed in the display area. The second user interface shown in Fig. 4 is used for video indexing and retrieval. The first number above the frame is video number. Next to the video number, the distance measure is displayed. The program interface shows the provision of first 20 candidate video frames to be displayed.

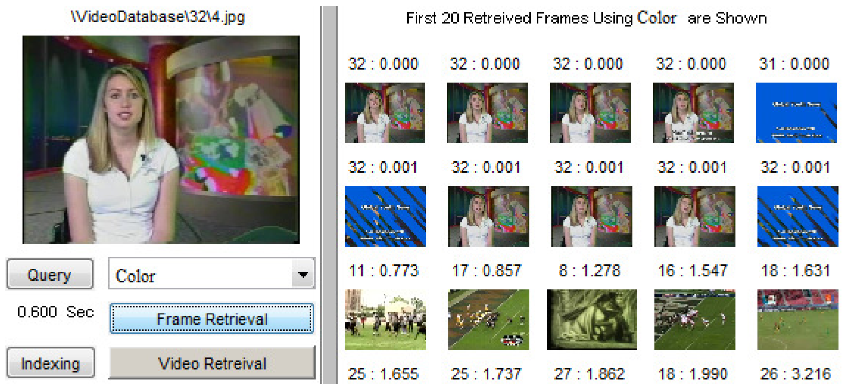

FIGURE 4: User Interface for Video Indexing and Retrieval.

4.1 Clip Analysis

The key frames are then packed into one summary file called shots.avi in this file; the size of the video frames is determined by its relative length of the shots, in another word, determined by how important that shot is. The output file format for segmented video is .avi. The file format of extracted key frames is .jpg. The system stores extracted key frames as key0.avi, key1.avi, and so on. The extracted key frames using user interface is shown in Fig. 5.

FIGURE 5: Interface Extracted Key Frames (Scene.avi).

Clip Details	Duration (sec)	Frames	Key Frames	Position of Key Frames	Tolerance
BladeRunner	3	103	8	1,4,20,25,33,36,37,41	2.064
Rollcall2	8	194	9	1,10,20,27,33,40,176,183,189	1.950
Scene	6	92	4	1,17,41,73	2.476
Vip	5	46	6	1,2,5,11,26,45	2.476
house_tour	91	666	14	1,2,14,31,39,41,73,77,96,121, 153,211,229,250,419,447	2.125

TABLE 1: Clip Analysis.

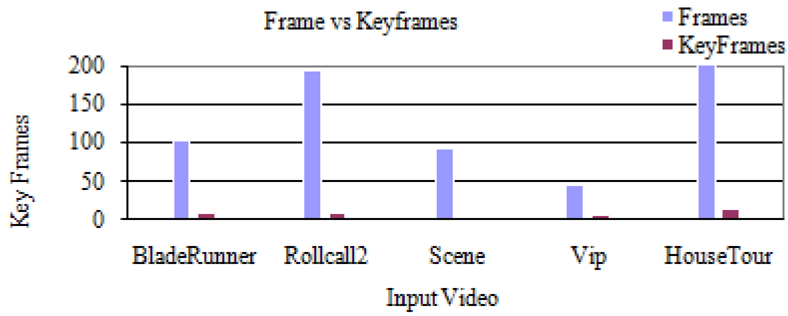

FIGURE 6: Number of Key Frames in Videos.

The number of key frames with respect to given frames in a video is shown in Fig. 6. The first and last frame of any given video is always a key frame in the given video. The detail of the key frames is stored in text file. Different length of input videos is used here to compare the results. The positions of the key frames extracted are shown in Table 1. The dynamic threshold is automatically selected based on the difference between current and next key frame of a given videos. The tolerance and key frames are also shown in Table 1.

4.2 Video Retrieval Results

The key frames retrieved using HSV and Entropy are shown in Fig. 7 and Fig. 8 respectively.

FIGURE 7: Video Retrieval Using HSV Color Planes (a) Space (b) TV Show.

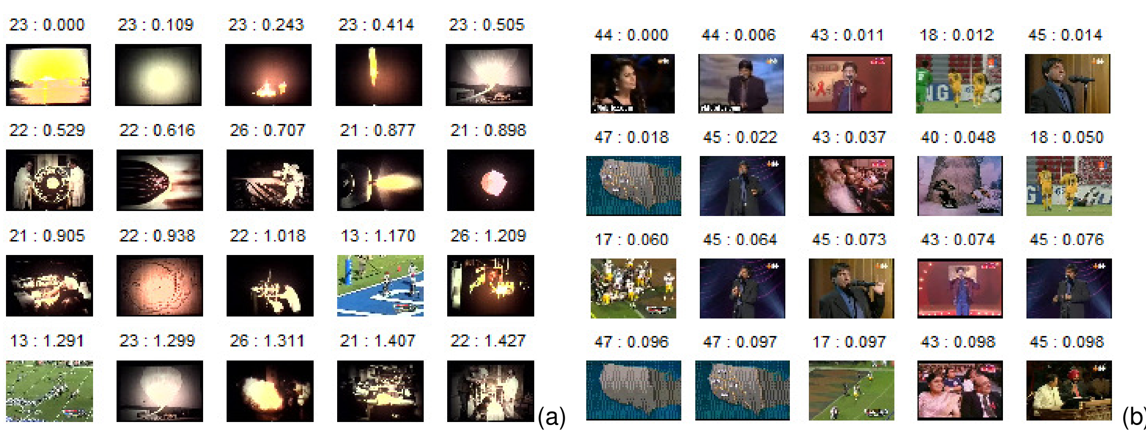

FIGURE 8: Video Retrieval Using Entropy Features (a) Space (b) TV Show.

In this result class of input videos is space and TV Shows. The result is measured and compared using precision and recall curve. The estimated precision recall curve using color features and entropy features for video retrieval are shown in Fig. 9.

It is observed that the video retrieval results using entropy features are better than the color features results.

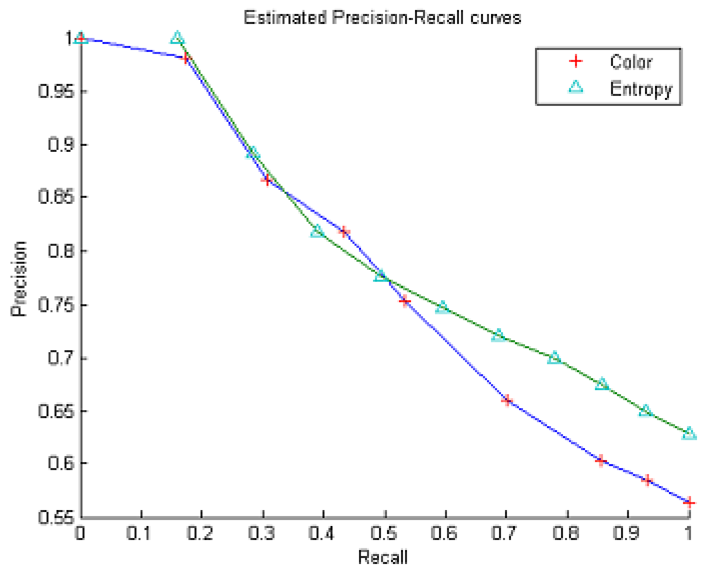

FIGURE 9: Estimated Precision-Recall curve.

5. REFERENCES

[1] F. Idris and S. Panchanathan. “Review of Image and Video Indexing Techniques,” Journal of Visual Communication and Image Representation. 8(2). 146-166. 1997.

[2] N. Dimitrova, T. McGee and H. Elenbaas. “Philips Research,” 345 Scarborough Rd. Briarclii Manor NY, 10510.

[3] J. Oh, A. Kien and Hua. “An Efficient Technique for Summarizing Videos Using Visual Contents,” Multimedia and Expo, ICME 2000. IEEE International Conference, Vol. 2, pg 1167 – 1170, Jul 2000.

[4] Y. Gong and X. Liu. “Generating Optimal Video Summaries,” Multimedia and Expo, ICME 2000, IEEE International Conference, Vol. 3, pg 1559 – 1562, Jul 2000.

[5] D. DeMenthon, V. Kobla and D. Doermann. “Video Summarization by Curve Simplification,” Technical Report LAMP-TR-018, CS-TR-3916, University of Maryland, College Park, 1998.

[6] R. Leinhart, S. Pfeiffer and W. Effelsberg. “Video Abstracting,” Communications of the ACM, Vol. 40, No. 12, Dec 1997.

[7] A. Girgensohn and J. Boreczky. “Time-Constrained Keyframe Selection Technique,” in *IEEE Multimedia Systems “99, IEEE Computer Society*, vol. 1, pp. 756-761, 1999.

[8] N. Gamaz, X. Huang, and S. Panchanathan. “Scene Change Detection in MPEG Domain,” Image Analysis and Interpretation, IEEE Southwest Symposium, pg. 12-17, 1998.

[9] D. M. Ali and M. Ghanbari. “Clear Scene Cut Detection Directly from MPEG Bit Streams,” IEEE IP and its Applications, No.465, Vol. 1, pg. 285-289, Jul 1999.

[10] J. Nang, S. Hong and Y. Ihm. “An Efficient Video Segmentation Scheme for MPEG Video Stream using Macroblock Information,” 7th ACM international conference on Multimedia, pg 23 – 26, Oct 1999.

[11] B. L. Yeo and B. Liu. “Rapid scene analysis on compressed videos,” IEEE Trans. Circuits Systems Video Technol. 5, 1995, 533–544.

[12] N. Doulamis, A. Doulamis, Y. Avrithis and S. Kollias. “Video content representation using optimal extraction of frames and scenes,” in Proc. of IEEE Int. Conference on Image Processing (ICIP), Chicago USA, Oct. 1998.

[13] F. Chen, M. Cooper and J. Adcock. “Video summarization preserving dynamic content,” proceeding of the TRECVID video summarization, pages 40-44. Germany 2007.

[14] P. Over, A. F. Smeaton and P. Kelly. “The TRECVID 2007 BBC rushes summarization evaluation pilot,” In Proceedings of the TRECVID Workshop on Video Summarization (TVS”07), pages 1-15, New York, NY, September 2007, ACM Press.

[15] X. F. Yang, Q. Tian and P. Xue. “Short Video Repeat Identification With Application to News Video Structure Analysis,” IEEE Transaction on Multimedia, 9(3)(2007), pp. 600-609.

[16] N. M. Loccoz, E. Bruno, and S. M. Maillet. “Interactive Retrieval of Video Sequences from Local Feature Dynamics, Lecture Notes in Computer Science, 3877(2006), pp. 128-140.

[17] H. Lu, B. C. Ooi, H. T. Shen, and X. Xue. “Hierarchical Indexing Structure for E.cient Similarity Search in Video Retrieval,” IEEE Transaction on Knowledge and Data Engineering, 18(11), 2006, pp. 1544-1559.

[18] J. Shao, Z. Huang, H. T. Shen, X. Zhou, E. P. Lim and Y. Li. “Batch Nearest Neighbor Search for Video Retrieval,” IEEE Transaction on Multimedia, 10(3)(2008), pp. 409-420.

[19] C. G. M. Snoek, B. Huurnink, L. Hollink, M. D. Rijke, G. Schreiber and M.Worring. “Adding Semantics to Detectors for Video Retrieval,” IEEE Transaction on Multimedia, 9(5)(2007), pp. 975-986.

[20] T. T. Sato, T. Kanade, E. K Hughes and M. A. Smith. “Video OCR for digital news archive,” Proceedings of IEEE International Workshop on Content-Based Access of Image and Video Databases, Jan. 3rd, pp. 52-60, 1998.

[21] S. H. Han, K. Yoon and I. S. Kweon. “A new technique for shot detection and key frames selection in histogram space,” 12th Workshop on Image Processing and Image Understanding, pp. 475-479, 2000.

[22] S. L. Varma and S. N. Talbar. “iMATCH Image Matching and Retrieval for Digital Image Libraries,” ICETET10, pp. 196-201, December 2009.

[23] S. L. Varma and S. N. Talbar. “IRMoment Image Indexing and Retrieval by Combining Moments,” IET Digest, Volume 2009, Issue 1, pp. 38, 2009.

[24] W. Niblack, R. Berber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic and P. Yanker. “The QBIC project. querying images by content using color, texture and shape,” SPIE Storage and Retrieval for Image and Video Database II, pp. 173-187, 1993.

[25] J. Dowe. "Content-based retrieval in multimedia imaging," in SPIE Storage and Retrieval for Image and Video Databases II, pp.164-167, 1993.

[26] J. R. Bach, C. Fuller, A. Gupta, A. Hampapur, B. Horowitz, R. Humphrey, R. Jain and C. F. Shu. "The Virage image search engine. an open framework for image management," in SPIE Storage and Retrieval for Image and Video Databases V, pp 76-87, 1996.

[27] J. R. Smith and S. F. Chang. "An image and video search engine for the World-Wide Web," in Proc. of SPIE, vol. 3022, pp 85-95, 1997.

[28] C. Carson, S. Belongie, H. Greenspan and J. Malik. "Region-based image querying," IEEE CVPR"97 Workshop on Content-Based Access of Image and Video Libraries, pp. 42-49, 1997.

[29] T. P. Minka and R. W. Picard. "Interactive learning with a society of models," in Pattern Recognition, 30(4), pp. 565-581, Apr. 1997.

[30] Y. Rui, T. Huang and S. Mehrotra. "Content-based image retrieval with relevance feedback in MARS," IEEE International Conference on Image Processing, pp. 815-818, Oct. 1997.

[31] S. F. Chang, W. Chen, H. J. Meng, H. Sundaram and D. Zhong. "A fully automated content-based video search engine supporting spatiotemporal queries," IEEE Trans. Circuits Syst. Video Technol., vol. 8, no. 5, pp. 602–615, Sep. 1998.

[32] Z. Yang, X. Wan and C. C. J. Kuo. "Interactive image retrieval. concept, procedure and tools," in IEEE 32nd Asilomar Conference, Montery, CA, pp. 261–265, Nov. 1998.

[33] R. Zabih, J. Miller and K. Mai. "A feature based algorithm for detecting and classifying scene breaks," Proceedings of the 3rd ACM International Conference on Multimedia, pp. 189-200, 1995.