Elysium Journal of Engineering Research & Management
P-ISSN: 2347-4408
E-ISSN: 2347-4734

VIDEO SUMMARIZATION USING COLOR FEATURES AND GLOBAL THRESHOLDING

[1]Nishant Kumar [2]Amit Phadikar
[1]Department of Computer Science & Engineering
[2]Department of Information Technology
MCKV Institute of Engineering, Liluah, Howrah, India
[1]nishant89.sheo@gmail.com
[2]amitphadikar@rediffmail.com

Abstract*— Compact representations of video data can enable efficient video browsing. Such representations provide the user with information about the content of the particular sequence being examined. Most of the methods for video summarization relay on complicated clustering algorithms that makes them too computationally complex for real time applications. This paper presents an efficient approach for video summary generation that does not relay on complex clustering algorithms and does not require frame length as a parameter. Our method combines color feature with global thresholding to detect key frame. For each shot a key frame is extracted and similar key frames are eliminated in a simple manner.

Index Terms — Video Summarization, YC_bC_r Color Space, Color Histogram.

I. INTRODUCTION

Enormous popularity of the Internet video repository sites like YouTube or Yahoo Video caused increasing amount of the video content available over the Internet. In such a scenario, it is necessary to have efficient tools that allow fast video browsing. These tools should provide concise representation of the video content as a sequence of still or moving pictures - i.e. video summary. There are two main categories of video summarization [1]: static video summary and dynamic video skimming. In static video summary methods, a number of representative frames, often called keyframes, are selected from the source video sequence and are presented to the user. Dynamic video summary methods generate a new, much shorter video sequence from the source video. Since static video summaries are the most common technique used in practical video browsing applications, we focused our research on static video summarization. Most of the existing work on static video summarization is performed by clustering similar frames and selecting representatives per clusters [2-6]. A variety of clustering algorithms were applied such as: Delaunay Triangulation [2], k-medoids [3], k-means [4], Furthest Point First [5] and [6] etc. Although they produce acceptable visual quality, the most of these methods relay on complicated clustering algorithms, applied directly on features extracted from sampled frames. It makes them too computationally complex for real-time applications. Another restriction of these approaches is that they require the number of clusters i.e. representative frames to be set a priori.

The contribution of this paper is to propose a fast and effective approach for video summary generation that does not relay on complicated clustering algorithms and does not require length (number of summary frames) as a parameter. The proposed model is based upon color histogram in YC_bC_r color space.

The rest of the paper is outlined as: In section II, color space has been discussed. Section III discusses the proposed work. Performance evaluation is discussed in section IV. Finally, section V discusses the conclusion.

II. COLOR SPACE

RGB Color Space: This is an additive color system based on tri-chromatic theory. Often found in systems that use a CRT (Cathode Ray Tube) to display images. It is device dependent and specification of colors is semi–intuitive. RGB (Red, Blue & Green) is very common, being used in virtually every computer system as well as television etc.

YC_bC_r Color Space: The difference between YC_bC_r and RGB is that YC_bC_r represents color as brightness and two color difference signals, while RGB represents color as red, green and blue. In YC_bC_r, the Y is the brightness (luma), C_b is blue minus luma (B-Y) and C_r is red minus luma (R-Y). This color space exploits the properties of the human eye. The eye is more sensitive to light intensity changes and less sensitive to hue changes. When the amount of information is to be minimized, the intensity component can be stored with higher accuracy than the C_b and C_r components. The JPEG (Joint Photographers Engineering Group) file format makes use of this color space to throw away unimportant information [7]. RGB images can be converted to YC_bC_r Color Space using following conversion process given in matrix form in Eq: 1. Y component is luminance, C_b is blue chromaticity and C_r is red chromaticity.

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.2989 & 0.5866 & 0.1145 \\ -0.1688 & -0.3312 & 0.5000 \\ 0.5000 & -0.4184 & -0.0816 \end{bmatrix} * \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

P-ISSN: 2347-4408
E-ISSN: 2347-4734

III. OVERVIEW OF THE PROPOSED METHOD

We propose an approach which is based on several efficient video processing procedures. At first, video frames are sampled in order to reduce further computational burden. Then, Color feature is extracted on pre-sampled video frames and the Euclidean distance measure is used to measure the similarity between the frames. These features are deployed for key frames detection using a threshold approach. Based on the preset threshold, key frame is said to be detected at places where the frame difference is maximal and larger than the global threshold. Then, a representative key frame is extracted and similar key frames are eliminated in a simple manner. As a final result the most informative key frames are selected as a video summary. In the rest, detailed description of every step of the method is presented.

STEP 1: Frame Sampling: The video is sampled at 24 frames per second. This sampling may contain redundant frames.

STEP 2: Frame Feature Extraction: Frame feature extraction is a crucial part of a key frame extraction algorithm which directly affects performances of the algorithm.

Color: Several methods for retrieving images on the basis of color feature have been described in the literature. Color feature is easy and simple to compute. The color histogram is one of the most commonly used color feature representation in image retrieval as it is invariant to scaling and rotation. Color histogram of an image in the Y (Luminance), C_b (Chrominance of blue), and C_r (Chrominance of Red) color space are calculated. Color histogram is very effective for color based image analysis. They are especially important for classification of images based on color.

STEP 3: Dissimilarity Measure: The next important step is similarity measures. Similarity measure is playing important role in the system. It compares the image feature vector of a frame with the feature vectors of previous image. It actually calculates the distance between them. Images at high distance are tagged as key frame and will be selected finally.

Euclidean distance measure is used to find the histogram difference. If this distance between the two histograms is above a threshold, a key frame is assumed The dissimilarity between frames, f_i and f_{i+1} is computed as the Euclidian distance between feature vector of f_i and feature vector of f_{i+1}

Euclidean Distance is represented as:

$$dm(f_i, f_{i+1}) = \sqrt{\sum(F_i(j,k)^2 - F_{i+1}(j,k)^2)} \quad (2)$$

where, F_i and F_{i+1}: feature vector containing components of Y, C_b and C_r channels of frames.

STEP 4: Threshold Selection: The problem of choosing the appropriate threshold is a key issue in the key frame algorithms. Here, we have chosen global thresholds as an appropriate method. The threshold is calculated from average value of distance of all frames.

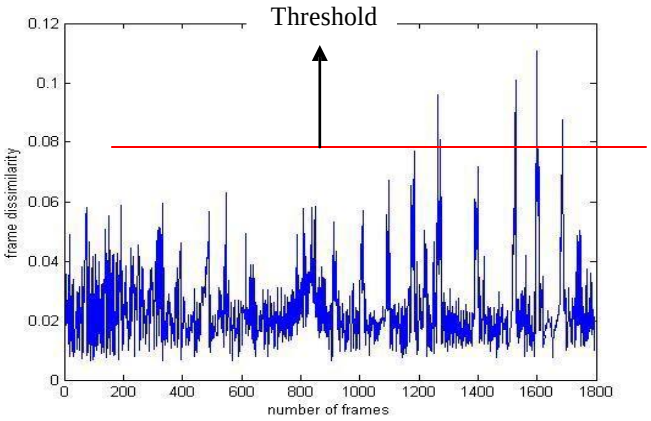


Fig. 1. Threshold value on frame difference (video 2).

The above figure that is Figure 1: shows key frames being detected. The bar which crossed the threshold value is selected as the key frames. For example, frame numbers like 9, 154, 176,195, 257, etc have been selected as the key frames because they have crossed the threshold value in the experimental data set (video 2).

STEP 5: Detection of Key frames: The proposed model is based on color histogram. Given a video which contains many frames, the color histogram for each frame is computed and the Euclidean distance measure is used to measure the dissimilarities between the frames. Based on the predefined threshold, a key frame is said to be detected if the dissimilarity between the frames is higher than the threshold value.

IV. PERFORMANCE EVALUATION

This section presents the results of the experiments conducted to corroborate the success of the proposed model. The experimentation is conducted on set of YOUTUBE- videos and The Open Video Project- videos.

The performance of the proposed model is evaluated using precision and recall as evaluation metrics. The precision measure is defined as the ratio of number of correctly detected keyframe to the sum of correctly detected & falsely detected keyframe of a video data and recall is defined as the ratio of number of detected keyframe to the sum of detected & undetected keyframe. These parameters were obtained for the proposed model on three different video samples.

$$Precision = \frac{Number\ of\ relevant\ frames\ retrieved}{To\ 换\ al\ number\ of\ frames\ retrieved} \quad (3)$$

P-ISSN: 2347-4408
E-ISSN: 2347-4734

$$Recall = \frac{Number\ of\ relevant\ frames\ retrieved}{\quad}$$ (8)

Fig. 3. Preview of generated summaries of test videos: (a) Wildlife, (b) New Horizon 1 & (c) New Horizon 2.

TABLE 1: METRICS OF THE PROPOSED WORK.

	Size	No. of frames tested	Key frame detection performance of proposed work	
			Precision	*Recall*
Video 1	7.89 MB	901	95.72%	80.00%
Video 2	8.81 MB	1813	89.90%	87.10%
Video 3	8.73 MB	1795	91.40%	91.80%

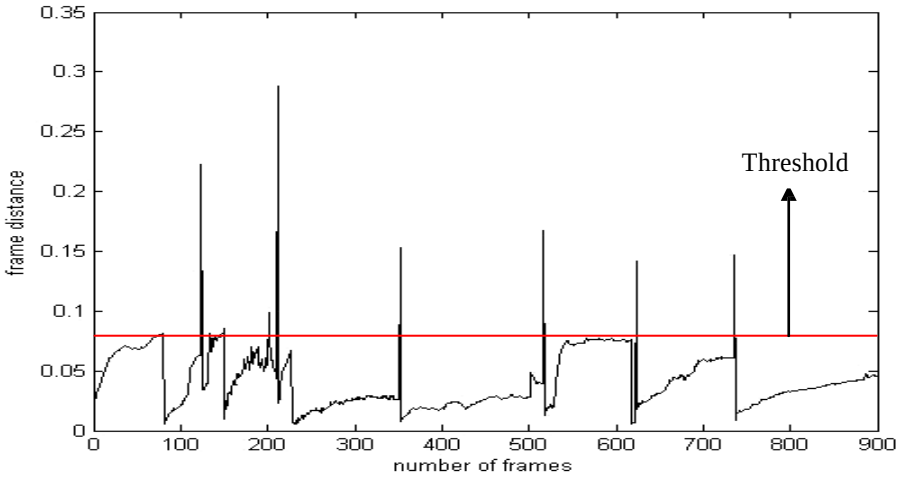
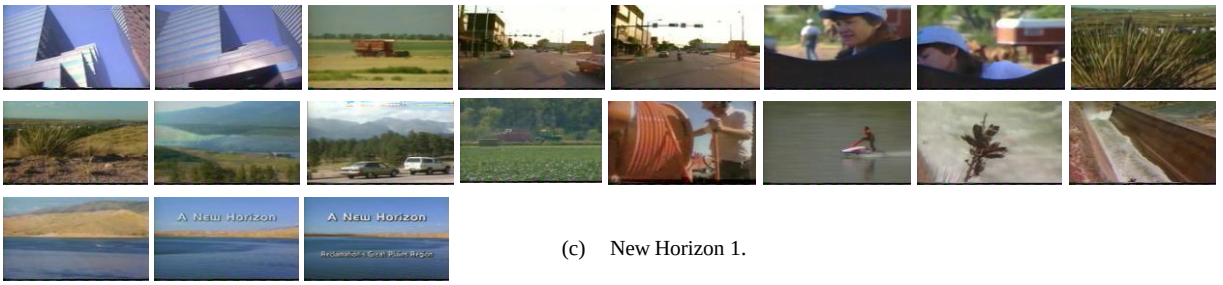


Fig. 2. Plot of frame dissimilarity for video 1.

Video 1

(a) Wildlife.

Video 2



(c) New Horizon 1.

Video 3

(e) New Horizon 2.

P-ISSN: 2347-4408
E-ISSN: 2347-4734

The results for three test videos randomly selected from YOUTUBE and The Open Video Project- videos action data is presented in Table 1. The Figure 2 is the plot of frame dissimilarity of video 1. The frame numbers that have crossed the threshold value have been selected as key frames. Figure 3 presents results of our method, preview of generated summaries of test videos, Video 1, Video 2, and Video 3 respectively.

The precision and recall comparisons between our method and Angadi et al. [9] are shown in Table 2. It is found that our method offers nearly similar result like S. A. Angadi and Vilas Naik [9]. Moreover, it is to be noted that the method proposed by S. A. Angadi and Vilas Naik [9] had high computation complicity as the scheme used color moments. However, our scheme has low computational complicity as it uses simple color histogram.

TABLE 2. PRECISION AND RECALL COMPARISONS.

S. A. Angadi and Vilas Naik [9]		Proposed	
Precision	***Recall***	***Precision***	***Recall***
90.66%	95.23%	92.34%	91.80%

V. CONCLUSION

In this paper, we proposed an efficient method for video summary generation. Every color histogram computed for an image in YC_bC_r color space is used to find difference between two frames in a video. The difference between consecutive frames to detect similarity/dissimilarity is computed as Euclidian distance between feature vector containing color of Y(Luminance), C_b (Chrominance of blue), C_r (Chrominance of red) values of frame. The key frames are detected wherever difference value is more than predefined threshold. Experimental results on standard YOUTUBE videos and on The Open Video Project-videos, data reveal that the proposed model is robust and generates video summary efficiently.

Future work will focus on further performance improvement of the proposed scheme by selecting adaptive threshold based on genetic algorithm (GA) and combination of motion, edge and color to increase the efficiency of key frame detection.

REFERENCES

[1] B. T. Truong and S. Venkatesh, "Video abstraction: A systematic review and classification," ACM Transactions on Multimedia Computing Communications and Applications, vol. 3, pp. 1-37, 2007.

[2] P. Mundur, Y. Rao and Y. Yesha, "Keyframe-based video summarization using Delaunay clustering," International Journal on Digital Libraries, vol. 6, pp. 219–232, 2006.

[3] Y. Hadi, F. Essannouni and R. O. H. Thami, "Video summarization by k-medoid clustering," in Proceedings of the ACM Symposium on Applied Computing, New York, p. 1400–1401, 2006.

[4] S. E. F. De Avila, A. P. B. Lopes, A. Luz and A. Albuquerque Araújo, "VSUMM: A mechanism designed to produce static video summaries and a novel evaluation method,"Pattern Recognition Letters, vol. 32, pp. 56–68, 2011.

[5] M. Furini, F. Geraci, M. Montangero and M. Pellegrini, "VISTO: visual storyboard for web video browsing", in Proceedings of the ACM International Conference on Image and Video Retrieval, p. 635–642, 2007.

[6] M. Furini, F. Geraci, M. Montangero and M. Pellegrini, "STIMO: Still and moving video storyboard for the web scenario," Multimedia Tools and Applications, pp. 47–69, 2007.

[7] H. B. Kekre, S. D. Thepade, and R. Chaturvedi, "Walsh, Sine, Haar & Cosine Transform With Various Color Spaces for „Color to Gray and Back," International Journal of Image Processing, vol. 6, pp. 349-356, 2012.

[8] S. Cvetkovic, M. Jelenkovic, and S. V. Nikolic, "Video summarization using color features and efficient adaptive threshold technique," Przegląd Elektrotechniczny, R. 89NR 2a, pp. 274-250, 2013.

[9] S. A. Angadi and Vilas Naik., "A shot boundary detection technique based on local color moments in YC_bC_r color space," Computer Science and Information Technology, vol. 2, pp. 57-65, 2012.