Automatic Video Summarization of Sport Archives using Visual Features

D. S. Pandya[1], M. A. Zaveri[2]

[1]U. V. Patel College of Engineering, Ganpat university, Mehsana, India

[2]Sardar Vallabhbhai National Institute of Technology, Surat, India

Abstract- *This paper presents an effective and fast approach for video summarization. Recent years have witnessed very large video volume that creates stiff challenge for users to quickly browse or retrieve these videos. Video summarization techniques help in creating short summary clip which is meaningful and able to convey original video. In this paper we have proposed a two step simple key frame extraction algorithm based on low level features, color and motion. In the proposed algorithm, Video is first segmented into shots using unsupervised learning approach. The algorithm emphasizes on the changes in the motion magnitude to identify the key frames from every shot. Simulations have been performed on large number of video datasets. Generated summary persists the temporal continuity and important events of the video.*

Keywords- key frames, unsupervised learning, motion magnitude

1 Introduction

Rapid revolution in digital video has brought many applications at home in affordable cost. The volume of digital data has been increasing rapidly due to the wide usage of multimedia applications in the areas of education, entertainment, business, medicines etc. Video summarization helps to meet these needs by developing condensed version of full length video [1, 2].

Raw video is an unstructured data stream, physically consisting of a sequence of video shots. A video shot is composed of a number of frames and its visual content can be represented by key-frames. Video summarization is defined as a collection of key-frames extracted from a video. This summary can be a sequence of stationary images or moving images (video skims) In general, content-based video summarization can be thought as a two-step process. The First step is partitioning the video into physical shots, called video segmentation or video shot boundary detection. The second step is to find these representative frames. Thus, video can be organized as video, shot and key frames.

In [3] different solutions to video summarization have been broadly described. Shot based key frame extraction techniques have dragged the huge attention of researcher's community. The principal methodology of shot boundary detection is to extract one or more features from the frames in a video sequence and then difference between two consecutive frames is computed. If the difference is more than a certain threshold value, a shot boundary is detected. There are number of techniques have been attempted by the researchers to segment the video [4].

Color and motion are most important features used for video frames. There are different color histograms based approaches in which consecutive frames are compared to decide the selection of key frames. In [5, 6], HSV color space is used to measure interframe difference. HSV color space has outperformed RGB color space due to its perceptual uniformity. RGB mutual information and joint entropy of adjacent two frames have been used in [7]. A color histogram is insensitive to camera and object motion. Therefore color based key frame selection may not be enough to render the visual contents of a shot. Optical flow components are extracted and motion function is computed between two frames. There are several well known methods to compute optical flow among which Lucas & Kanade is popular and also faster [8, 9, 10]. In last recent years, machine learning based approaches have dragged the attention. A machine learning system is developed that learns to predict video transitions based on feature information derived from frames. In supervised learning, low level features are employed to train the system that can predict transitions on unseen data [11, 12]. Various unsupervised learning approaches have been attempted by the researchers to extract key frames and found faster as they do not require training [13]. Several keen observations have been inferred from these studies: some approaches emphasize only single characteristics of video contents like color histogram or motion. While some approaches based on clustering selects key frames which are nearer to the cluster centers [14, 15]. Our proposed approach has been described in section 2. Section 3 discusses experimental results. Section 4 concludes the paper.

2 Proposed Approach

A good video summary should be compact and meaningful hence generating video summary requires detail understanding of video structure and it's semantic. In this paper we presented two step video summarization algorithms for sports specifically for soccer game. Flowchart of our algorithm has been shown in Fig 1, in which first step segments the video into shots using clustering approach and selection of key frames is performed in second step.

2.1 Feature Extraction

Sport videos (soccer) are dominated by ground, audiences and players hence we extract the frame transitional features in terms of mean of green color and hue. Each frame is divided into 4 equal blocks which has been shown in Fig 2. From these blocks, we pertain only block 2 and 3 for extracting these features.

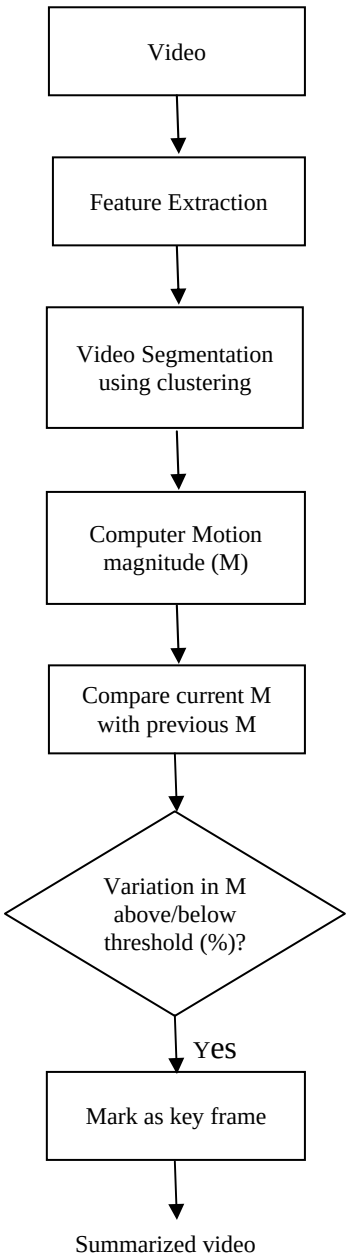

Fig. 1 Flow chart of Video Summarization.

Following features are used to segment the video using clustering approach.

2.1.1 Dominant color:

Game videos are always dominated by the ground and soccer game has the maximum contribution of the grass filed. We pick up green histogram of the blocks and find the mean of green color.

$$m_{green} = \sum_{i=0}^{L-1} r_i p(r_i)$$

2.1.2 Hue:

HSV color space outperforms RGB color space in many applications because of its perceptual uniformity. Hue is very important attribute because it represents dominant wavelength. We find the mean of hue for every frame of the video.

2.2 Video Segmentation

Before generating video summary, it is required to identify the shot boundary of the video. In soccer games there are well known shots (view) like long view, close-up, audience etc. Long view is dominated by the ground and also indicates that play is on, while close-up shot focuses on individual player/ referee. Audience shot may be close or long view but it has no dominance of grass ground hence it can be easily detected by the color features.

Fig. 2 Frame is divided into 4 equal regions

It is quite evident that color features are good descriptors to clearly detect the boundaries among the long, close-up and audience shots. K-means has been the center choice and also faster method to cluster the data. Hue and green color mean have been collectively used as a feature for classifying the video frames into three classes (long, close-up and audience) using k-means. Fig 3 depicts the mean of green and Hue component of the various shot. Many segmentation approaches rely on strictly defining threshold to declare shot transition however in these approaches for different videos it becomes necessary to redefine threshold. Clustering

approach is advantageous because it eliminates the need of statically defining the threshold for shot transition.

2.3 Summary generation

After the video has been segmented, it is required to select key frames which are able to convey the original video. Games video genre has the intrinsic highly dynamic nature hence it possesses frequent and intensive motion. Between every two frames, optical flow components are computed. Lucas and Kanade is a widely used differential method for optical flow estimation in computer vision. It is also less sensitive to image noise.

Field view	**Close-up**	**Audience view**
Green mean:125.4	Green mean:97.4	Green mean:69.8
Hue mean:0.2740	Hue mean:0.3498	Hue mean:0.4663

Fig. 3 Field, Close-up, Audience shot with their values

Lucas-Kanade method assumes that displacement of the image contents is approximately constant within a neighbourhood (window) of the pixel under consideration. Velocity vector (V_x, V_y) must satisfy

$$Av = b \tag{1}$$

where

$$A = \begin{bmatrix} I_x(q_1) & I_y(q_1) \\ I_x(q_2) & I_y(q_2) \\ . & . \\ I_x(q_n) & I_y(q_n) \end{bmatrix}, \quad b = \begin{bmatrix} -I_t(q_1) \\ -I_t(q_2) \\ . \\ -I_t(q_n) \end{bmatrix} \quad v = \begin{bmatrix} V_x \\ V_y \end{bmatrix}$$

q_1, q_2,…q_n are pixels inside the window, and $I_x(q_i)$, $I_y(q_i)$, $I_t(q_i)$ are the partial derivatives of the image I with respect to position x, y, and t evaluated at pixel q_i and at the current time. It obtains the solution of equation (1) by least square method. Finally it computes

$$v = \begin{bmatrix} \sum_i I_x(q_i)^2 & \sum_i I_x(q_i)I_y(q_i) \\ \sum_i I_x(q_i)I_y(q_i) & \sum_i I_y(q_i)^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_i I_x(q_i)I_t(q_i) \\ -\sum_i I_x(q_i)I_t(q_i) \end{bmatrix}$$

In the experiments, window size has been kept 3. For the group of N+1 frame, N optical fields (F_1, F_2, …., F_n) will be computed by the algorithm. For every optical field, Motion function is computed as follows.

$$M(i) = \frac{\sum_{(x,y)\in F_i} \sqrt{{V_x}^2(x,y)+{V_y}^2(x,y)}}{P \max_{(x,y)\in F_i} \sqrt{{V_x}^2(x,y)+{V_y}^2(x,y)}} \tag{2}$$

Key frame selection algorithm is described below.

1. Select key frame from the shot if the motion magnitude undergoes the change of half or twice compared to the magnitude of previous frame.
2. Select every first frame as a key frame of shot transition.

Motion magnitude of the three consecutive frames belonging to long filed view has been shown in Fig. 4. Even though all these three frames look similar, motion magnitude undergoes huge change in (a) and (b). Images from (a) to (c) observe change in position of football and various players which eventually gives rise to the motion magnitude.

(a)	**(b)**	**(c)**
Magnitude:350.5	Magnitude:2019.2	Magnitude:1838.9

Fig. 4 Long field view with motion magnitude

3 Simulation and results

We experimented 9 soccer videos with total length of more than 2 hours and 20 minutes. Well known European leagues like Barclays premier league, Ligabbva, LaLiga and Serie A premier league soccer videos have been used in the experimentation. All these video possess varying ground and illumination conditions. All video data sets have the 352x288 resolution at 25 frames/s. Every frame of the video is processed to generate the summary. Experimental results have been shown in Table I, where $L_{original}$ denotes full video length and $L_{summary}$ is the summarized video length in seconds.

Table I: Video Summarization Performance

Video	$L_{original}$	$L_{summary}$	C_r (%)	Im (%)
1	600	108	18	9.4
2	600	95	15.8	8.8
3	1500	226	15.1	9
4	840	147	17.5	8.8
5	840	161	19.2	9.4
6	840	141	16.8	9.2
7	1000	147	14.7	9.4
8	915	194	21.2	9.1
9	1310	213	16.3	10
Average	8445	1432	17	9.3

For the comparison between original video and summary we have used two parameters: Compression ratio (C_r) and informativeness(I_m). Compression ratio is the ratio of summary video length $L_{summary}$ divided by the original video length $L_{original}$. Informativeness evaluates how many objects/events of the original video are included in the summary video. Soccer videos are eventful and events like replays, yellow card, penalty corner/kick (goal attempts), and goal are considered to evaluate the informativeness of video. Results in Table I clearly reflect that proposed approach provides high compression and imformativeness. Informativeness has been shown in the scale of 10.

4 Conclusion

In this paper, we proposed two step automatic video summarization method using unsupervised approach for large sport archives. Sports videos are dominated by color and motion feature. First step of algorithm succeeds to segment the different shots of videos based on color features using k-means and second step succeeds to select key frames using motion feature. Experimental results clearly show that our algorithm achieves very good compression and informativeness in spite of varying ground and illumination conditions of video datasets. On average, we achieved 17% compression ratio and 9.3 informative score. Generated summary maintains temporal sequence and does not deteriorate the enjoyability. In the future, we will extend the work to model the events.

5 References

[1] Ying Li, Tong Zhang, Daniel Tretter "An overview of video abstraction techniques" Technical *Report.* HP-2001-191, HP Laboratory, July 2001.

[2] M. Roach, J. Mason, L.-Q. Xu, and F. Stentiford. "Recent trends in video analysis: a taxonomy of video classification problems. *In Proceedings of the International Conference on Internet and Multimedia Systems and Applications, IASTED*, pp. 348-353, Aug 2002.

[3] Arthur G Money, Harry Angius, "Video summarization: A conceptual framework and survey of the state of the art" *Journal of visual communication & image represent* ", vol. 19, pp. 121-143, 2008.

[4] S Carrato, I Koprinska "Temporal video segmentation: A Survey", *Signal Processing: Image Communication*, vol. 16, no. 5, pp. 477-500, 2001.

[5] Yue Gao, Hai Dai "Shot based similarity measure for content based video summarization", *IEEE Transactions on Image Processing*, pp. 2512-2515, 2008

[6] Chen, Hu, Zeng and Li "Indexing and matching of video shots based on motion and color analysis, "IEEE *Trans. on ICARCV*", pp. 1-6, 2006.

[7] Wei, Shen, Jiang "A novel algorithm for video retrieval using video metadata information". *IIEEE international workshop on Education technology and Computer science*, vol. 2, pp. 1059-1062, 2009.

[8] Wolf, W "Key frame selection by motion analysis" , *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Atlanta, GA, pp. 1228-1231, 1996

[9] B Lucas, T. Kanade "An iterative image registration technique with an application to stereo vision", *Seventh International joint conference on Artificial Intelligence*, Vancouver, Canada, pp. 674-679, 1981.

[10] Ling Shao, Ling Ji "Motion Histogram analysis based key frame extraction for human activity representation", *Canadian Conference on computer and robot vision*, pp. 88-92, 2009

[11] Ren, Yuesheng Zhu "A Video Summarization Approach Based on Machine Learning" *International conference on intelligent information hiding and multimedia signal processing IIHMSP*, pp. 450-453, Aug. 2008.

[12] Basak, J, Luthra, V. Chaudhury, S. "Video summarization with supervised learning" *19th international conference on pattern recognition ICPR*, pp. 1-4, Dec. 2008.

[13] Ren, K, Fernando, W.A.C, Calic, J, "Optimising video summaries using unsupervised clustering" *50th International Symposium ELMAR*, vol. 2, pp. 451-454, Sept. 2008.

[14] Z. Yueting, R. Yong, T. S. Huang, and S. Mehrotra, "Adaptive key frame extraction using unsupervised clustering," in *Proc. ICIP*, vol. 1, pp. 866–870, 1998.

[15] Hammoud, R, Mohr, R. "A probabilistic framework of selecting effective key frames from video browsing and indexing", *Proceedings of International Workshop on Real-Time Image Sequence Analysis*, pp. 79-88, Oulu, Finland, Aug 2000.