Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 23 (2011) 713 – 717

Procedia Engineering

www.elsevier.com/locate/procedia

PEEA 2011

Genetic Keyframe Extraction for Soccer Video

Xue Yang, Zhicheng Wei*

College of Information Technology, Hebei Normal University, Shijiazhuang, 050016, China

Abstract

This paper presents a genetic algorithm based model for soccer video summarization. The model first introduces audio features to improve fitness function which is used to calculate the relative differences among all the selected frames. Then the model employs crossover and mutation operators to get the meaningful summary in a video search space. Experimental results and comparisons are presented to show our model can get more reasonable and attractive frames than the traditional method on soccer video static summarization.

Selection and/or peer-review under responsibility of [name organizer]

Keywords:scooer video summarization; genetic algorithm; fitness function; audio feature

1. Introduction

Rapid increase in the amount of video data demands for various multimedia applications to effectively manage and store a huge amount of audio visual information. Hence, there is a strong demand for a mechanism to provide compact representations of video sequences that allow users to gain certain perspectives of a video without having to watch it in its entirety.

Video summarization offers a concise representation of the original video. According to [1], there are two fundamental types of video abstracts viz. static video abstract and dynamic video skimming. A static abstract is a small collection of salient images extracted from the original video sequence. A dynamic skimming consists of both the image sequences and the corresponding audio abstract. A simple approach to extracting key-frames from shot is based on frame content changes computed by features, such as color histogram [2] or motion activity [3]. Zhu proposed a hierarchical video summarization strategy that explores video content structure to provide the users with a scalable, multilevel video summary [4]. In

* Corresponding author. Tel.:+86-158-311-66767.
E-mail address: weizhicheng@hebtu.edu.cn.

doi:10.1016/j.proeng.2011.11.2570

above methods, a predefined threshold is required to control the number of keyframes. Furthermore, there are some shot independent approaches. For instance, Girgensohn proposed a time-constrained clustering method to extract keyframe [5]. Other more sophisticated methods are also proposed, including the integration of motion and spatial activity analysis with face detection technologies [6], object-based approach [7], and a progressive multi-resolution keyframe extraction technique [8].

Despite numerous efforts in generating video summarization, the results are still far from satisfactory. Those systems that neglect audio track are not able to obtain impressive results. Also, the algorithms involving over-intensive computation are normally impractical to real applications. Moreover, those methods can't suit for all kinds of videos. To overcome these drawbacks, this paper proposes to employ a Genetic Algorithm to deliver a meaningful summary (still image abstract) of soccer videos that uses multimodal features.

2. Visual Feature Extraction

A soccer game usually lasts for about two or more hours, in this image sequence, a lot of images may be similar to adjacent ones. We should give more attention to those that are not too similar, so we first subsample the video at a low rate, and call this original set of frames A. We pick out the least similar frames by measuring their differences with color histograms, and call this reduce set A'.

There is a notion that the longer shots are more important, because they can attract more attention than the shorter ones. So, we define the length factor I_L to evaluate the frames. Then, people are always interested in rare image in a video, if they have no subjective intention. So we use commonality factor I_C as a criterion to define the fitness function. Furthermore, earlier appearing frames are more heavily weighted than later ones if they are similar, so we use another factor I_{Pr} to eveluation the precedence of each frames[9].

3. Genetic Algorithm for Soccer Video summarization

In this paper, we combine visual features and audio features for defining the fitness function to select key frames. The soccer video abstract procedure is described in Fig. 1.

3.1. Audio Feature Extraction

Audio feature is an important part of feature framework. People are always attracted by the louder or sudden sound if they have no subjective intention. Therefore, in this section, we define two audio features: average sound energy and average sound peak. We also consider the response time when we calculate the audio features. Because when an attack or a goal occurs, commentators and audiences need short time, about 0.1s~0.3s, to react [10].

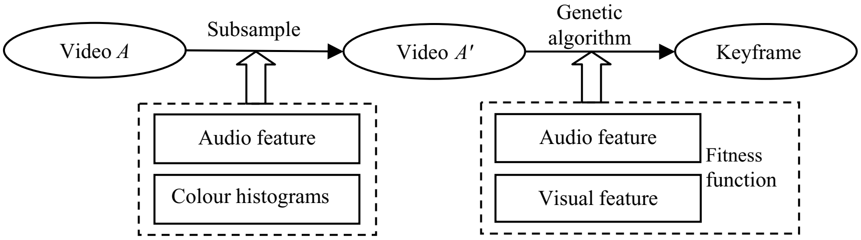


Fig. 1. The block diagram of the present model

In general, people may pay attention to an audio segment with absolute loud sound, which can be measured by average energy of an audio segment. Hence, the average sound energy is defined as:

$$I_E = E_{avr} / MaxE_{avr} \ . \tag{1}$$

Where E_{avr} denotes the average energy of an audio segment. $MaxE_{avr}$ is the maximum average energy of an entire audio segments.

On the other hand, the sudden sound effects always grab human attention. So the average sound peak is calculated as follow.

$$I_{Pe} = E_{peak} / MaxE_{peak} \tag{2}$$

We define the E_{peak} and $MaxE_{peak}$ as we defined E_{avr} and $MaxE_{avr}$ above.

3.2. Integrated Audio Feature and Visual Feature

When subsampling the original video, we not only use the differences with color histograms, but also consider the impact of audio features. So we define the difference of any two images *i* and *j* as follow.

$$g(i,j) = d(i,j) \cdot I_E \cdot I_{Pe}$$
$$gh(i) = g(i-1,i)$$

Where *d(i,j)* is the difference of any two images *i* and *j* with color histograms.

Then we put together five factors for average sound energy, average sound peak, length, commonality and precedence, we define the importance as follow.

$$I = I_E \cdot I_{Pe} \cdot \log(I_L) \cdot \log(1/I_C) \cdot I_{\mathrm{Pr}} \tag{3}$$

To take into consideration the relative distinction among keyframes, we use the important together with the difference of any two frames in set *A'* to define the fitness function as follow.

$$f(S_k) = \sum_{i,j \in S_k} g(i,j)(I_i + I_j) \tag{4}$$

Where S_k is a subset of *k* selected images in *A'*.

3.3. Genetic Algorithm

A genetic algorithm is a search method used to find exact or approximate solutions to optimization problems [11]. Due to the large number of the set *A'*, traditional algorithm is hard to effectively optimize (4). So, we use GA to look for a given number *k* of images that provide as much information as possible about what actually happened during the video.

GA used in this paper can be divided into three parts: encoding, fitness function, crossover and mutation operations. In GA approach, a population of chromosomes is randomly generated, and the evolution process is performed iteratively one generation at a time. In the end, the individual with the highest fitness is decoded to obtain the video summarization.

3.3.1. Encoding

We have chosen to encode our chromosomes with binary encoding because of its popularity and its relative simplicity. Moreover, our experiments have proved that the binary genetic algorithm is more quickly and easily to get the best fitness value than the decimal genetic algorithm when they are used in video summarization. For the binary encoding, every chromosome is a string of bits (0, 1). In our genetic solution, the bit position of a chromosome string is an index for a image in A'. The length of the chromosome is the number of frames. We use 1 to denote the selected frames, while 0 denotes the frame which doesn't be selected. The number of 1 is set to be a fixed constant by the input specification.

3.3.2. Fitness Function

In this paper, we use (4) as fitness function. We emphasize that any well-defined fitness function can also be used and will work with the genetic mechanism of the algorithm. That is one of the advantages of the proposed model.

3.3.3. The Crossover and Mutation Operators

The genetic algorithm works by randomly selecting chromosomes to reproduce, biasing selection toward individuals with higher fitness. For Standard GA (SGA) crossover operator, two chromosomes are sliced at the crossing site, and the two tail pieces are swapped and rejoined with the head pieces to produce two offsprings with crossover possibility P_c. In our GA method, instead of crossing at a random bit, we use the position of left of the 1 as the possible crossing site, all crossing sites being selectable with equal probability.

In order to maintain a fixed number of 1's in each chromosome, a mutation procedure is applied. A gene is randomly selected and the value at that position adjusted in order to make the total number of 1 closer to that originally specified. This operation is repeated until obtaining the desired number of 1.

4. Experiment and Results

We take the frames which encoded by 1 plus the first frame of the original video as the keyframe. This algorithm is applied to the video with k=5 and population size of 50 randomly chosen chromosomes. It runs over 100 generations with a mutation possibility of 0.2 and a crossover possibility of 0.8.

Video 1 is a soccer video which record shot process. Fig. 2 shows the key frames layout of three methods: uniform approach, GSA approach [9] and present approach. In a collection of equidistant frames, long scenes without much change are heavily emphasized, as shown in Fig. 2(a). Meanwhile, frames from very short shots might be missed with a uniform selection. At the same time, it can be seen in Fig. 2(b) that the keyframes using GSA approach puts emphasis on diversity and abundance. Unfortunately, it doesn't contain the most attractive frame, the goal frame. Frame 238 in video 1 is the image in which the player kicks a goal. The results using our approach are showed in Fig. 2(c). We can see that present approach can get the main information including the goal frame from the original soccer videos. Meanwhile, the two other methods obtain too much redundant information and hard for people to get the main idea.

5. Conclusion

In this paper, we have presented a universal model to generating summarizations of soccer videos. We pick out the least similar images from original video by measuring the differences with the color histograms and audio information. The novel fitness function was designed by integrating visual features and audio features together. We also consider the response time during calculating the audio features. Experimental results show that the GA based method with audio and visual features is more suitable for abstracting soccer video, and it can get the optimal fitness to get the main idea of the soccer video.

As future paths of research, we plan to add other features (motion, speech) to improve the quality of the summary. Furthermore, the larger set of soccer video keyframe extraction is also worth an investigation.

6. References

[1] Truong BT, Venkatesh S. Video abstraction: A systematic review and classification*, ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, vol. 3, no. 1, February 2007.

[2] Zhang HJ, Wu JH, Zhong D, Smoliar SW. An integrated system for content-based video retrieval and browsing, *Pattern Recognition*, vol.30, no.4, pp.643-658, 1997.

[3] Wolf W. Key frame selection by motion analysis, *Proc. Of ICASSP'96*, vol.2, pp.1228-1231, 1996.

[4] Zhu X, Wu X, Fan J, Elmagarmid AK and Aref WG. Exploring video content structure for hierarchical summarization, *Multimedia Syst*, vol.10, no. 2, pp.98-115, 2004.

[5] Girgensohn A, Boreczky J. Time-constrained key frame selection technique, *Proc. of ICMCS*, pp. 756-761, 1999.

[6] Dufaux F. Key frame selection to represent a video, *Proc. of ICME, Vancouver*, vol. 2, pp. 275-278, 2000.

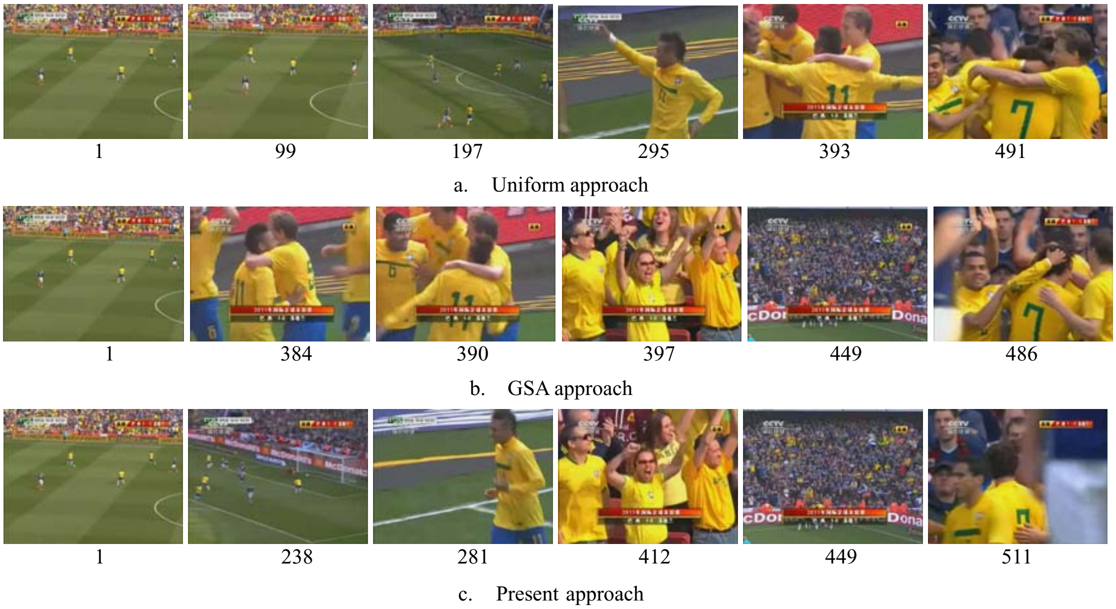
[7] Rav-Acha A, Pritch Y, Peleg S. Making a long video short: Dynamic video synopsis, In *CVPR'06*, New York, pp. 435 – 441, June 2006.

[8] Kim C, Hwang JN. An integrated scheme for object based video abstraction, *Proc. of ACM Multimedia 2000*. Los Angeles, CA, 2000.

[9] Patrick C, Andreas G, Wolf P et. al. A Genetic Algorithm for Video Segmentation and Summarization, *roceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, New York, Vol. 3, pp. 1329, 2000.

[10] Spierer D, Petersen R, Duffy K. Gender influence on response time to sensory stimuli, *Journal of Strength and Conditioning Research*, vol. 24, No. 4, pp. 957-963, 2010.

[11] Harik GR, Lobo FG, Goldberg DE. The compact genetic algorithm, *IEEE Transactions on Evolutionary Computation*, Vol. 3, No. 4, pp. 287-297, 1999.



a. Uniform approach

b. GSA approach

c. Present approach

Fig. 2. Keyframes comparison from video 1