http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 8 (3): 388-392, 2009
ISSN 1812-5638

A Novel Dynamic Video Summarization Approach Based on Rough Sets in Compressed Domain

[1,2]Li Xiang-Wei, [3]Zhang Ming-Xin, [2]Zhao Shuang-Ping and [2]Zhu Ya-Lin
[1]Department of Computer Engineering, Lanzhou Polytechnic College, Lanzhou 730050, China
[2]College of Electrical Engineering and Information Engineering,
Lanzhou University of Technology, Lanzhou 730050, China
[3]College of Mathematics and Information Science, Northwest Normal University,
Lanzhou 730070, China

Abstract: In this study, a novel dynamic video summarization approach based on Rough Sets (RS) is developed. It can not only rapidly generate video summary that minimizes the visual content redundancy for input video sequences, but also specify the number of frames to get various summary according to user demand. First, DCT coefficients and DC coefficients, the most important video visual features are extracted from raw video sequences to represent video information. Second, an Information system is constructed with DC coefficients. Third, a new and concise Information system is achieved by using the reduction theory of RS, meanwhile, the effective representation of frames and its corresponding reduced frame numbers are recorded, i.e., dynamic video summarization. Experimental results indicate that the proposed algorithm is more effective and intelligent than conventional methods in video summarization generation.

Key words: video sequences, RS, key frames

INTRODUCTION

In recent years, the development of novel technologies such as electronic technologies, broadband communication networks and storages have diversified services for the multimedia application. Vast amounts of video information are created, stored, and transmitted for education, entertainment, etc., every day. In particular, it is an urgent need to develop technologies to quickly search and browse video information from large-scale video archives, e.g., when working with large volume video data, people may only care about specific content contained in video stream. To date, great efforts have been made to solve this problem, there into, automatic video content summarization technologies become one of the most promising solutions. The objective of video summarization is to create a shorter video clip that maintains as much semantic content of the original video streams. A concise and informative video summary will enable the user to quick figure out general contents of a video and help us to decide if it is worthwhile to watch through the whole sequences. For content-based video retrieval, a video summary will dramatically save the user's time and efforts to spot the desired videos from a large volume of video collections.

Generally, the automatic video summarization technique can be divided into two basically classes, i.e., static video summary and dynamic video skimming. The static video summary is simply constructed by extracting a few key frames from original video sequences through techniques like clustering, key frame extraction etc. Unfortunately, such a scheme does not consider the temporal information among video shots. In contrast, dynamic video skimming tries to condense the original video into a much shorter sequences which demonstrates the dynamic variation of video content against time elapse. At present, compared with static summary, there are relatively few works that address dynamic video skimming. Nonetheless, due to the advance and popularity of visual capturing tools, effective techniques for dynamic video skimming are highly in demand.

Conventional methods of dynamic video summarization can be categorized about into following classes :(1) analyzing the attributes of frame according to low-level features (Ma and Zhang, 2002), such kind of methods is suitable for most videos sequences since heavy-motion frames can often attract the focus. (2) Combining low-level and high-level features to evaluate the importance of a frame or a shot (Kadir and Divakaran, 2003; Wang *et al*., 2008).

Corresponding Author: Li Xiang-Wei, Department of Computer Engineering, Lanzhou Polytechnic College, Lanzhou 730050, China

This kind of methods is a ideal scheme, while in most case, the high-level features is difficult to extract. (3) Selecting shots according to results of shot classification and event detection (Ekin *et al*., 2003). Such kind of methods can be often found in analyzing highlights of video. (4) Determining the importance of a frame based on user preferences (Fonseca and Pereira, 2004), which are specified via querying interface. For this kind of methods, semantic event analysis is usually required since low-level features are not sufficient to identify frames or shots that match the user preferences.

The major limitations of above technologies are that many algorithms need preliminary knowledge or training model, and the selected frames are not scientific. Moreover, the generated summarization is stationary which can not alter according to user demand. RS is a novel and powerful tool for data analysis. It has successfully been used in many application domains, such as machine learning, expert system and pattern classification (Kotlowski *et al*., 2008; Mac-Parthalain *et al*., 2008). The main advantage of rough sets is that it does not need any preliminary or additional information about data, like probability in statistics, or basic probability assignment in Dempster-Shafer theory and grade of membership or the value of possibility in fuzzy sets. Especially, the attributes reduction theory has widely used in many application domain (Wang *et al*., 2008; Yao *et al*., 2008). In this study, by using the advantage of RS theory, a RS based video summarization extraction algorithm is developed.

BASIC THEORY OF ROUGH SETS

Rough sets is a mathematical formulation of the concept of approximative equality of sets in a given approximation space. An approximation space is the pair (U, R), where U is the universe and R is an indiscernibility relation on U. The main important concepts are described as following.

Indiscernibility relation: Let $U \neq \phi$ be a universe of discourse and X be a subset of U. An equivalence relation, R, classifies U into a set of subsets $U/R = \{X_1, X_2, X_3,..., X_n\}$ in which the following conditions are satisfied:

- $X_i \subseteq U, X_i \neq \phi$ for any i
- $X_i \cap X_j \neq \phi$ for any i, j
- $\cup_{I = 1, 2, ..., n}, X_i = U$

Any subset X_i, which called a category, class or granule, represents an equivalence class of R. A category in R containing an object $x \in U$ is denoted by $[x]_R$. For a family of equivalence relations $P \subseteq R$, an indiscernibility relation over P is denoted by IND(P) and is defined by Eq. 1.

$$IND(P) = \bigcap_{R \in P} IND(R) \quad (1)$$

Lower and upper approximations: The set X can be divided according to the basic sets of R, namely a lower approximation set and upper approximation set. Approximation is used to represent the roughness of the knowledge. Suppose a set $X \subseteq U$ represents a vague concept, then the R-lower and R-upper approximations of X are defined by Eq. 2 and 3.

$$\underline{R}X = \{x \in U : [x]_R \subseteq X\} \quad (2)$$

Equation 4 is the subset of X, such that X belongs to X in R, is the lower approximation of X.

$$\overline{R}X = \{x \in U : [x]_R \cap X \neq \phi\} \quad (3)$$

Equation 5 is the subsets of all X that possibly belong to X in R, thereby meaning that X may or may not belong to X in R and the upper approximation $\overline{R}$ contains sets that are possibly included in X. R-positive, R-negative and R-boundary regions of X are defined respectively by Eq. 4-6.

$$POS_R(X) = \underline{R}X \quad (4)$$

$$NEG_R(X) = U - \overline{R}X \quad (5)$$

$$BNR(X) = \overline{R}X - \underline{R}X \quad (6)$$

Attributes reduction and core: In RS theory, an Information table is used for describing the object of universe, it consists of two dimensions, each row is an object, and each column is an attribute. RS classifies the attributes into two types according to their roles for Information table. Core attributes and redundant attributes. Here, the minimum condition attribute set can be received, which is called reduction. One Information table might have several different reductions simultaneously. The intersection of the reductions is the Core of the Information table and the Core attribute are the important attribute that influences attribute classification.

A subset B of a set of attributes C is a reduction of C with respect to R if and only if:

- $POS_B(R) = POS_C(R)$, and
- $POS_{B-\{a\}}(R) \neq POS_C(R)$, for any $a \in B$

And, the Core can be defined by Eq. 7

$$CORE_C(R) = \{c \in C \mid \forall c \in C, POS_{C-\{c\}}(R) \neq POS_C(R)\} \quad (7)$$

THE PROPOSED ALGORITHM

The proposed algorithm can be described as following six steps:

Step 1: Extract the DCT coefficients from video sequences
Step 2: Extract the DC coefficients from DCT coefficients achieved in step1
Step 3: Construct the Information System using DC coefficients
Step 4: Generate Core of Information System by attribute reduction theory of RS and record the redundant frames
Step 5: Sort the frame in the Core according to the redundant frame number
Step 6: Generate summarization according to user demand

Extraction of DCT coefficients: In term of MPEG national standard, the video sequences in compressed domain consist of I, P and B frame. The I frame is the base of video sequences, which use DCT to compress in spatial, so the DCT coefficients can represent the full video information and be easily extracted from video sequences directly. We can represent this process as following:

$$\psi(P(x),t) \xrightarrow{\text{extract}} \text{DCT coefficients}$$

where, $\psi(P(x),t)$ denotes the video sequences. Table 1 shows part of 8×8 block DCT coefficients extracted from video sequences.

Extraction of DC coefficients: The DCT coefficients are made of DC coefficients and AC coefficients, DC coefficients denote the average and most important information in video frame. So we can utilize the DC coefficients to represent the video frame. This process can be described as following:

$$\text{DCT coefficients} \xrightarrow{\text{repreces sin g}} \text{DC coefficients}$$

Table 2 shows part of DC coefficients extracted from DCT coefficients

Construct information system with coefficients: We have got the DC coefficients of each frame, so we can construct an Information system with it. Each row is a DC coefficient, and each column is a frame. This process can be described as following.

$$\text{DC} \xrightarrow{\text{construct}} \text{information table } S=\{U, A, V, f\}$$

where, U is sets, denotes all the object of Information System, A is also a sets, denotes all attributes in Information system, V is the sets of attributes value, f is a function denotes the relations between objects and attributes.

By using above process, we can get Information System as Table 3.

Table 1: Part of DCT coefficients extracted from video sequences

A 8*8 block DCT coefficients							
0.35355	0.35355	0.35355	0.35355	0.35355	0.35355	0.35355	0.35355
0.49039	0.41573	0.27779	0.097545	-0.097545	-0.27779	-0.41573	-0.49039
0.46194	0.19134	-0.19134	-0.46194	-0.46194	-0.19134	0.19134	0.46194
0.41573	-0.097545	-0.49039	-0.27779	0.27779	0.49039	0.097545	-0.41573
0.35355	-0.35355	-0.35355	0.35355	0.35355	-0.35355	-0.35355	0.35355
0.27779	-0.49039	0.097545	0.41573	-0.41573	-0.097545	0.49039	-0.27779
0.19134	-0.46194	0.46194	-0.19134	-0.19134	0.46194	-0.46194	0.19134
0.097545	-0.27779	0.41573	-0.49039	0.49039	-0.41573	0.27779	-0.097545

Table 2: DC coefficients extracted from DCT coefficients

DC coefficients extracted from DCT coefficients							
3.0456	2.2260	2.1020	2.3544	2.4456	2.2279	2.0706	1.9765
2.0397	2.1328	2.4147	4.4863	4.3309	4.3603	4.3294	4.3304
4.3941	4.4235	4.3912	4.3608	4.4549	4.2657	2.7294	2.2902
2.2275	2.1348	2.2265	2.0078	2.8245	2.2902	3.2632	2.5098
1.9784	2.4779	2.6422	2.7539	2.1426	1.7578	1.6324	1.6309
1.3799	1.2554	1.7250	4.1382	4.4809	4.1711	4.2363	4.2368
4.2353	4.2662	4.2990	4.2667	4.1078	4.3275	2.6990	2.3225
2.7917	2.1343	2.8230	1.3946	2.7294	2.3191	3.1054	2.2583

Table 3: Information system constructed using DC coefficients

	Frame							
Coefficients	1	2	3	4	5	6	7	8
DC1	0.35355	0.35355	0.27771	0.35355	0.41572	0.35355	0.35355	0.35355
DC2	0.35355	0.35355	0.27779	0.35355	0.41573	0.35354	0.35355	0.35355
DC3	0.35355	0.35355	0.27779	0.35356	0.19134	0.35355	0.35355	0.35355
DC4	0.35356	0.35356	0.27779	0.35355	0.19134	0.35351	0.35355	0.35355
DC5	0.35355	0.35355	0.27779	0.35355	0.19134	0.35336	0.35355	0.35355
DC6	0.27779	0.27779	0.27779	0.35356	0.19134	0.35345	0.35355	0.35355
DC7	0.27779	0.27779	0.27779	0.35355	0.19134	0.35353	0.35355	0.35355
DC8	0.23761	0.23761	0.27779	0.35355	0.35355	0.35355	0.35355	0.35355

Table 4: The reduced information system

	Frame					
Coefficients	1	3	4	5	6	7
DC1	0.35355	0.27771	0.35355	0.41572	0.35355	0.35355
DC2	0.35355	0.27779	0.35355	0.41573	0.35354	0.35355
DC3	0.35355	0.27779	0.35356	0.19134	0.35355	0.35355
DC4	0.35356	0.27779	0.35355	0.19134	0.35351	0.35355
DC5	0.35355	0.27779	0.35355	0.19134	0.35336	0.35355
DC6	0.27779	0.27779	0.35356	0.19134	0.35345	0.35355
DC7	0.27779	0.27779	0.35355	0.19134	0.35353	0.35355
DC8	0.23761	0.27779	0.35355	0.35355	0.35355	0.35355

Reduction of information system with RS theory: The attributes in the Information Table can be divided into two types according to their roles: Core attributes and redundant attributes. From Table 3, we can see that the frame 2 and 8 can be reduced; the reduced Information System is showed as Table 4. If we introduce a threshold, more attributes can be reduced.

The processing of reduction of attributes can be described as following:

```
  CORESET (A):=m_{i,j}                          // initiation of
CORESET(A)
 For (i=0;i<n;i++)
 {
                For (j=0;j<n;j++)
      {
                If (|m_{i,j}-m_{i,(j-1)}|>d)
                 {
                 CORESET: =CORESET (A) ∪ (m_{i,(j-1)})
                 Count (i)++;        //count the reduced
frame number of each core
                 }
                }
     }
```

The CORESET represents Core of Information System. Because the CORESET is the frame that can not be reduced, so it is the salient and interesting content in video sequences.

Generation of core in Information system and its reduced frame: The core is the most important attributes in information system that can not be reduced, which represents the interesting information of shots in video sequences, since Core eliminate many redundant frames of video sequences, the volume of video shortened dramatically. Core is the discrete frame, so it is the static summary and sacrificing the temporal evolution. In Information System, reduced frame represent the redundant information, the more frame is reduced; indicate the part content is more important in original video sequences. Video summarization should also emphasize these part contents, so the reduced frame number can represent the measure of salient content. Sorted according to the reduced number of each Core frame, and display the core and its reduced frame according user demand, the algorithm can extract the summarization automatically and intelligently.

EXPERIMENTATION AND RESULTS

A group of MPEG video sequences are selected to examine the performance of proposed algorithm. Figure 1 shows the first frame of summarization results on part sports video program. Table 5 shows the detailed evaluation results of five different video sequences.

The experiment demonstrates that proposed approach can effective and scientific generates more informative video skimming. This is because the approach process in the compressed domain, the amount of data reduced dramatically, only the one-eighth original data is considered in whole processing. Moreover, RS can reduce information system without any preliminary knowledge or training model.

Fig. 1: The first frame of summarization results on part sports video program

Table 5: Detailed evaluation results of five different video sequences

Video contents	Frames	Shots	Time length	Summary length
Gym	112	7	5	2
Animation	174	13	14	4
Scenery	317	23	18	5
Story	126	13	12	7
News	153	14	17	8

CONCLUSION

Conventional video summarization extraction technologies have deficiency of low efficiency and redundant data. In this study, a novel automatic approach for video summarization is developed. DCT coefficients and DC coefficients are extracted from raw video sequences. Then RS theory is introduced to data reduction. Users can also capable of setting time or frame number constraint for video summary. Present experiments have shown that better results are obtained compared to conventional technologies.

ACKNOWLEDGMENT

The study is supported by the Gansu Natural Science Foundation of China under Grant (No. 3ZS051-A25-047)

REFERENCES

Ekin, A., A.M. Tekalp and R. Mehrotra, 2003. Automatic soccer video analysis and summarization. IEEE Trans. Image Process., 12: 796-807.

Fonseca, P.M. and F. Pereira, 2004. Automatic video summarization based on MPEG-7 descriptions. Signal Process. Image Commun., 19: 685-699.

Kadir, P.A. and A. Divakaran, 2003. An extended framework for adaptive playback-based video summarization. Internet Multimedia Manage. Syst., 5242: 26-33.

Kotlowski, W., K. Dembczynski, S. Greco and R. Slowinski, 2008. Stochastic dominance-based rough set model for ordinal classification. Inform. Sci., 178: 4019-4037.

Ma, Y.F. and H.J. Zhang, 2002. A model of motion attention for video skimming. Proceeding of IEEE International Conference on Image Process, 2002 USA., pp: I-129-I-132.

Mac-Parthalain, N. and S. Qiang, 2008. Exploring the boundary region of tolerance rough sets for feature selection. Pattern Recognit. (In Press). 10.1016/j.patcog.2008.08.029

Wang, C., C. Wu and D. Chen, 2008. A systematic study on attribute reduction with rough sets based general binary relations. Inform. Sci., 178: 2237-2261.

Yao, Y. and Y. Zhao, 2008. Attribute reduction in decision-theoretic rough set models. Inform. Sci., 178: 3356-3373.