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 7 (7): 1043-1048, 2008
ISSN 1812-5638

A Novel Rough Sets Based Key Frame Extraction Algorithm in Compressed-Domain

[1,2]Li Xiang-wei, [1]Li Zhan-ming, [3]Zhang Ming-xin, [1]Wei Zhe and [1]Zhang Guo-quan
[1]College of Electrical Engineering and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China
[2]Department of Software Engineering, Lanzhou Polytechnic College, Lanzhou 730050, China
[3]College of Mathematics and Information Science, Northwest Normal University, Lanzhou 730070, China

Abstract: In this study, we propose a novel key frame extraction algorithm based on Rough Sets (RS) in Discrete Cosine Transform (DCT) compressed-domain. Firstly, we extract DCT coefficients in compressed-domain, select and preprocess the DC coefficients that derived from DCT coefficients. Secondly, We construct Information System with DC coefficients. Finally, we reduce Information System using attributes reduced theory of RS and obtained the representation of the video frames by reduced DC coefficients. Experimental results show that the proposed algorithm is fast and effective. Compared to conventional algorithm, our algorithm enjoys the following advantages: (1) the numbers of the key frame extracted using our algorithm become more scientific; (2) the algorithm can avoid the expensive computations in decompression processes.

Key words: Key frame, rough sets, discrete cosine transform, compressed-domain

INTRODUCTION

The amount of digital video content has grown extensively during recent years, resulting in a rising need for the development of systems for automatic index, summarization and semantic analysis (Bruyne *et al*., 2008). Shot boundary detection and key frame extraction are two bases for video indexing, browsing and retrieval. After shots are segmented, key frames can be extracted from each shot to represent the salient contents of the shot (Shuping and Xinggang, 2005). Key frame is still images extracted from the video stream and presented in temporal order (Money and Agius, 2007). Key frames provide a suitable abstraction and framework for video indexing, browsing and retrieval (Aigrain *et al*., 1996). Use of key frames greatly reduces the amount of data required in video indexing and provides an organizational framework for dealing with video content.

There have been many considerable work reported on key frame extraction in recent years. Different novel methods and technologies are developed. In sum, existing techniques for key frame extraction may be categorized into following classes: Shot boundary-based approaches select a key frame from a fixed position in the scene or several frames separated by a fixed distance; the visual content-based approach uses multiple visual criteria to extract key frames (Sze *et al*., 2005; Lee *et al*., 2006; Cooper *et al*., 2007); the motion analysis-based approach selects key frames at local minima of motion (Wu *et al*., 2005; Liu *et al*., 2003); the method based on the cluster, the video object, or the compressed domain, which can directly extract key frame in video sequence without decompression (Shuping and Xinggang, 2005; Fan and Ji, 2004; Jens-Raine *et al*., 2001).

The above methods have their merits respectively, but all exist below deficiency: (1) more above methods processed in uncompressed domain, so it need expensive computation to decoding. (2) The key frames can be simply selected from predetermined temporal locations such as the first, middle or last frame, but all these methods can still not provide an optimal and scientific representation of the video shots concerned.

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. In this study, we make use of the advantage of compressed domain and introduce the novel RS theory, propose a novel algorithm of key frame extraction. It can overcome the deficiency of common algorithm above and can get a satisfy effect.

VIDEO COMPRESSION PRINCIPLE AND ROUGH SET

DCT based video compression principle: According to MPEG national standard, the frame of video is firstly divided into many non-overlapping blocks whose size is 8×8. For each block, the number 128 is subtracted from

Corresponding Author: Li Xiang-wei, College of Electrical Engineering and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China

each pixel value before it is transformed by DCT. The transformed block is called a DCT block. Next, a weighted quantification table, called a Q table, is used to quantify the DCT coefficients for each block, preceded by the procedure for entropy encoding. The DCT block must be transformed into a one-dimensional (1D) array using a zigzag scan and the quantized coefficients in the 1D array are converted into the form by variable length coding. After that, these results are put through entropy encoding with the output the compressed data stream.

Each DCT block has its own DCT coefficients, each of which falls in the interval (0, 0) is called DC coefficients. The other coefficients are called AC coefficients. Discrete Cosine Transforms (DCT) defined by Eq. 1 is the heart of this compression scheme (Khayam, 2003).

$$C(u,v) = a(u)a(v)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}\cos\frac{(2x+1)u\pi}{2N}\cos\frac{(2y+1)u\pi}{2N}f(x,y) \quad (1)$$

For u, v = 0, 1, 2…, N-1 and a(u) a(v) are defined following:

$$a(u) = \begin{cases}\sqrt{\frac{1}{N}}\\ \sqrt{\frac{2}{N}}\end{cases}$$

The inverse transform is defined by Eq. 2.

$$F(x,y) = a(u)a(v)c(u,v)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}\cos\frac{(2x+1)u\pi}{2N}\cos\frac{(2y+1)u\pi}{2N}f(x,y) \quad (2)$$

Basic concepts of rough set theory: RS is a powerful and intellectual tool for data analysis. It can subdivide object and reduce attribute without any prior knowledge. The equivalence relation and equivalence class of its fundamental concept are described as following.

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 = $\{X_1, X_2,..., X_n\}$ 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.
- $\bigcup_{i=1,2,...,n} X_i = U$.

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. 3.

$$IND(P) = \bigcap_{R\in P} IND(R) \quad (3)$$

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 $X \subseteq R$ represents a vague concept, then the R-lower and R-upper approximations of X are defined by Eq. 4 and 5.

$$\underline{R}X = \{x \in U : [x]_R \subseteq X\} \quad (4)$$

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

$$\overline{R}X = \{x \in U : [x]_R \cap X \neq \phi\} \quad (5)$$

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. 6, 7 and 8.

$$POS_R(X) = \underline{R}X \quad (6)$$

$$NEG_R(X) = U - \overline{R}X \quad (7)$$

$$BNR(X) = \overline{R}X - \underline{R}X \quad (8)$$

Attributes reduction and core: In RS theory, a Information table is used for describing the object of universe, it consists of two dimension, each row is an object and each column is an attribute. RS classifies the attributes into three types according to their roles for Information table: Core attributes, reduced attributes and superfluous 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 $a \in B$

The Core can be defined by Eq. 9:

$$CORE_C(R) = \{c \in C \mid \forall c \in C, POS_{C-\{c\}}(R) \neq POS_C(R)\} \quad (9)$$

THE PRESENTATION OF PROPOSED ALGORITHM

The algorithm can be described as following five steps.

Extract DCT coefficients: 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 be extract from video sequences directly. It can describe 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 DCT coefficients extracted from video sequences.

Extract DC coefficients: The DCT coefficients are consist of DC coefficients and AC coefficients, the DC coefficients denotes the average and most important Information. So it can utilize the DC coefficients to represent the video frame. It can describe as following:

$$\text{DCT coefficients} \xrightarrow{\text{reprecesing}} \text{DC coefficients}$$

Table 2 demonstrates part of DC coefficients extracted from DCT coefficients.

Construct information system: We have got the DC coefficients of each frame, so we can construct an Information table use them. Each row is a DC coefficient and each column is the number of frame. The process can be describe as following.

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

where, U is sets, denotes all the elements 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 elements and attributes.

With above method, we can construct the Information System by DC coefficients, Table 3 is the part of Information System.

Reduct information system: Now, we have constructed a Information System with DC coefficients in terms of attributes reduced theory of RS many

Table 1: Part DCT coefficients extracted from video sequences

a 8*8 block DCT coefficients							
0.35355	1.73E-17	-1.73E-17	1.73E-17	1.04E-17	-8.67E-18	-9.54E-17	1.50E-16
2.45E-18	2.89E-34	-1.93E-34	3.37E-34	9.63E-35	-1.44E-34	-1.88E-33	2.94E-33
-1.23E-17	-7.70E-34	1.54E-33	-1.54E-33	-5.78E-34	5.78E-34	9.53E-33	-1.49E-32
1.47E-17	3.85E-34	-1.16E-33	1.93E-33	3.85E-34	-1.35E-33	-1.14E-32	1.73E-32
4.91E-18	5.78E-34	-3.85E-34	6.74E-34	1.93E-34	-2.89E-34	-3.76E-33	5.87E-33
-7.36E-18	-1.93E-34	5.78E-34	-9.63E-34	-1.93E-34	6.74E-34	5.68E-33	-8.67E-33
-9.57E-17	-3.08E-33	9.24E-33	-1.23E-32	-3.08E-33	4.62E-33	7.40E-32	-1.14E-31
1.53E-16	1.85E-32	-1.23E-32	2.16E-32	6.16E-33	-9.24E-33	-1.19E-31	1.86E-31

Table 2: DC coefficients extracted from DCT coefficients

DC coefficients we extract from 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	-0.09755	-0.27779	-0.41573	-0.49039
0.46194	0.19134	-0.19134	-0.46194	-0.46194	-0.19134	0.19134	0.46194
0.41573	-0.09755	-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	-0.09755	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	-0.09755

Table 3: The Information System constructed using DC coefficients

DC	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

redundant attributes can be reduced and it can get the Core of Information System, the process can be described as following:

```
CORESET(A):=mij // mij is a element of Core
  For(i=0;i<n;i++)
    {for(j=0;j<n;j++)
      {if(|mij-mi(j-1)|>d)
            CORESET:=CORESET(A)∪(mi(j-1))
        }
      }
```

Output the core of information system: Core is the most important attributes in Information System, which can not be reduced. In parallel with a shot, it is the most important frame, i.e., key frame.

RESULTS AND DISCUSSIONS

Various MPEG video sequences are selected to examine the performance of the proposed approach (here we segmented test sequences into video shots manually). Figure 1 illustrates some experimental results, we can see obviously that the condensed and succinct representations of the content is obtained by key frame extraction from video sequences and these frame can represent the focus of original video. Table 4 shows the evaluating result of key frame extraction by various video sequences. By comparison, the proposed algorithm can satisfy the need of key frame extraction, but the efficiency is increased dramaticly. Figure 2 shows the interface of comparison between before extraction and after extraction. The shot contain many

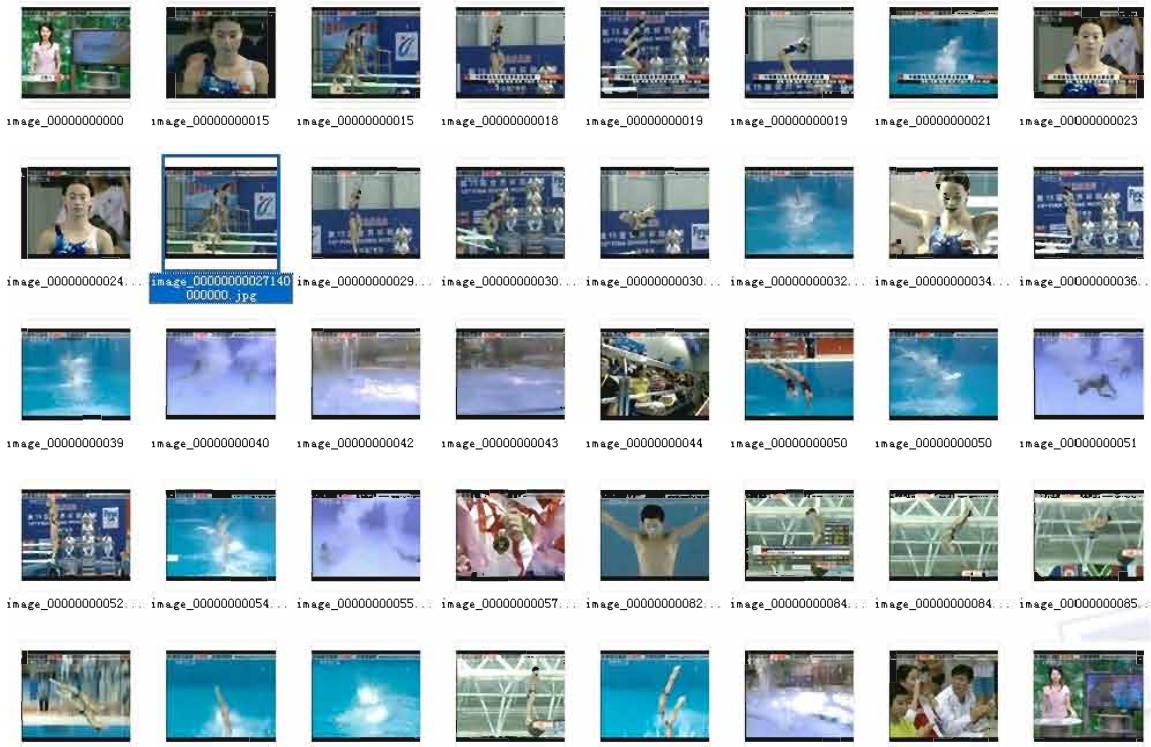

Fig. 1: The results of key frame extracted from sports shot

Table 4 Evaluating result of key frame extraction by various video sequences

Video type	Total frame No	Frame No extracted by system	Frame No extracted manually	Error extracted frame No	Loosed frame No	Conclusion
Gym	112	7	5	2	0	Excellent
Animation	174	13	14	0	1	Excellent
Scenery	317	23	18	5	0	Good
Story	126	13	12	1	0	Excellent
News	153	14	17	3	3	Good

(a) Before extraction (b) After extraction

Fig. 2: The interface of key frame extraction

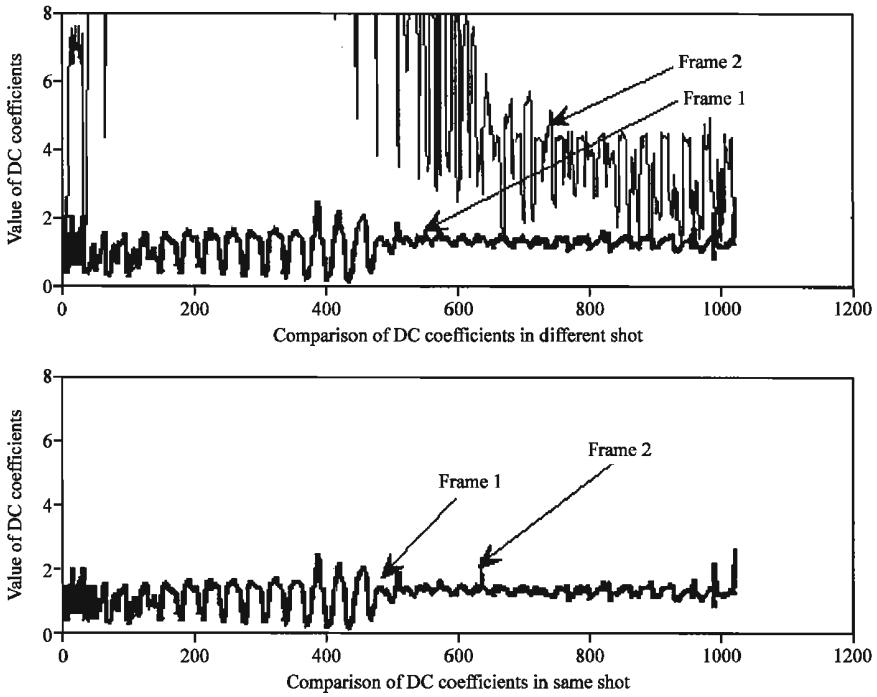

Fig. 3: Comparison of DC coefficients of two frames

same frame before key frame extraction. To validate the DC coefficients in different shot and in same shot, we also compared the DC data of two frame, the results are illustrated by Fig. 3.

The proposed algorithm applied to extract the key frames from real world video shots. Experimental results show high robustness in general case, similar to the standard test sequences. However, the results are still sensitive to strong lighting changes. In some extent, we can be improved by adjusting the threshold value in our approach.

CONCLUSIONS

In this study, a novel algorithm of key frame extraction based on RS is proposed. The algorithm directly operates the DCT and DC coefficients of video sequences without full-frame decompression, it occupies only a small fraction of the original date size while retaining most of the essential global Information. At the same time, the algorithm uses the attributes reduction theory of RS to get the Core of Information System, it can scientific represents the content of a shot. Applied to

various types of video sequences, It can be seen from the experiment results that the algorithm can effectively extract the key frame automatically from the MPEG compressed domain, which can avoid the expensive computation of DCT decoding and eliminate the redundant DC coefficients.

ACKNOWLEDGMENT

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

REFERENCES

Aigrain, P., H. Zhang and D. Petkovic, 1996. Content-based representation and retrieval of visual media: A state-of-art review. Multimedia Tools Applic., 3: 179-202.

Bruyne, S.D., D.V. Deursen and J.D. Cock, 2008. A compressed-domain approach for shot boundary detection on H.264/AVC bit streams. Signal Processing: Image Commun., 10.1016/j.image. 2008.04.012.

Cooper, M., T. Liu and E. Rieffel, 2007. Video segmentation via temporal pattern classification. IEEE. Trans. Multimedia, 9: 610-619.

Fan, J. and Y. Ji, 2004. Automatic moving object extraction toward content-based video representation and indexing. J. Visual Commun. Image Represent., 12: 306-347.

Jens-Rainer, M.O., V.V. Vasudevan and A. Yamada, 2001. Clolor and texture descriptors. IEEE. Trans. Cir. Syst. For Vidoe Technol., 11: 703-716.

Khayam, S.A., 2003. The discrete cosine transform (DCT): Theor. Applic., Inform. Theor. Coding. ECE 802-602, March 10th.

Lee, M.H., H.W. Yoo and D.S. Jang, 2006. Video scene change detection using neural network: Improved ART2. Expert Syst. Applic., 31: 13-25.

Liu, T., H.J. Zhang and F. Qi, 2003. A novel video key-frame-extraction algorithm based on perceived motion energy model. IEEE Trans. Cir. Syst. Technol., 13: 1006-1014.

Money, A.G. and H. Agius, 2007. Video summarisation: Aconceptual framework and survey of the state of the art. J. Visual Image Represent., 19: 121-143.

Shuping, Y. and L. Xinggang, 2005. Key frame extraction using unsupervised clustering based on a statistical model. Tsinghua Sci. Technol., 10: 169-172.

Sze, K.W., K.M.L. Lam and G. Qiu, 2005. A new key frame representation for video segment retrieval. IEEE. Trans. Cir. Syst Technol., 15: 1148-1155.

Wu, Q.Z., H.Y. Cheng and B.S. Jeng, 2005. Motion detection via change-point detection for cumulative histograms of ratio images. Patt. Recog. Lett., 26: 555-563.