Research Journal of

Information Technology

ISSN 1815-7432

Academic Journals Inc.

www.academicjournals.com

Research Journal of Information Technology 1 (2): 70-78, 2009
ISSN 1815-7432

Rapid Shot Boundary Detection Algorithm Based on Rough Sets in Video Compressed Domain

[1,2]L. Xiang-Wei, [1]L. Zhan-Ming, [3]Z. Ming-Xin, [2]Z. Ya-lLn and [1]W. Wei-Yi
[1]College of Electrical Engineering and Information Engineering,
University of Technology, Lanzhou 730050, China
[2]Department of Computer Engineering,
Lanzhou Polytechnic College, Lanzhou 730050, China
[3]College of Mathematics and Information Science,
Northwest Normal University, Lanzhou 730070, China

Abstract: Based on rough sets (RS), a novel rapid shot boundary detection algorithm was proposed in discrete cosine transform (DCT) compressed domain. First, DCT coefficients and DC coefficients are extracted from video sequences, so an information system is constructed with DC coefficient. Second, information system is reduced by ruduction theory of RS, the representation of the video frame is obtained by reduced DC coefficients. Finally, by introducing subdividing theory of RS, the frames of video are segmented objectively. The experimental results show that the algorithm can achieve higher performance. Compared to conventional algorithm, the algorithm enjoys many advantages. Firstly, only a subset of frames needs to be considered during analysis, allowing the reduction of the computational complexity, so the algorithm can avoid the expensive computations in decompression processes. Secondly, the relativity of segmentation of video shot becomes more scientific than earlier methods. Its robustness and effectiveness are validated by experiments with various kinds of video sequences.

Key words: Shot boundary detection, rough sets, discrete cosine transform

INTRODUCTION

Video has become a major interactive medium of communication means in our daily life. The sheer volume of video makes it extremely difficult to browse and find the focus in huge video information. So, it is necessary to develop a system to analyze, process and index video data automatically. Typically, the first step in automating system of video analysis is shot boundary detection which breaks the massive volume of video into smaller chunks called shots. Each shot represents an event or continuous sequence of actions. It can be used as a unit to browse or index, the boundaries between video shots are commonly known as shot boundary and the act of segmenting a video into shots is called shot boundary detection.

Many theories and technologies are presented for shot boundary detection, especially in uncompressed domain, i.e., all video must be decompressed before methods are processed. Since operations on uncompressed video do not permit rapid processing because of the account of data, so it is very time consuming. At the same time, more and more videos are in compressed forms according to MPEG national standard, either in storage or in communication. So, it is urgent need to develop algorithms to operate directly in compressed-domain.

Corresponding Author: L. Xiang-Wei, College of Electrical Engineering and Information Engineering, University of Technology, Lanzhou 730050, China

This study proposes a rapid shot boundary detection algorithm that operate directly in compressed video for temporal segmentation and boundary detection. The algorithm overcomes the limitations of earlier approaches and increases the efficiency of shot boundary detection.

PREVIOUS APPROACHES

There have been many considerable study reported on shot boundary detection. Different methods and technologies are used to examine the changes between successive frames and determine whether changes have taken place. At present, approaches may be categorized into following classes: (1) visual feature difference based technology, include difference of gray-level sums, sum of gray-level differences, difference of color histograms, colored template matching, difference of color histograms and x^2 comparison of color histograms (Chi-Chun and Wang, 2001), (2) motion analysis and fuzzy clustering based technology, it can use motion information to detect the discontinuities of frame and draw a conclusion by using threshold (Haoran *et al*., 2006), (3) model classification or neural network based technology, by constructing model or neural network train and capture the different type of shot transitions (Kyungnam *et al*., 2005; Lee *et al*., 2006), (4) key frame and video summarization based technology, by optimal key frame representation scheme or video summarization simplify video shot boundary detection (Sze *et al*., 2005; Money *et al*., 2007), (5) object based technology, employing the coding scheme of MPEG-4, extract the object of video to examine the boundary of shots such as localised and global lighting changes, variations in object size, occlusions and complex object motion and so on (Bangjun and Xn, 2006), (6) background based technology, exploits the fact that shots belonging to one particular scene often have similar backgrounds, although part of the video frame is covered by foreground objects (Liang-Hua *et al*., 2008), (7) event based technology, use of dimensionality reduction for video event detection without explicitly using motion estimation or object tracking (Ioannis *et al*., 2008) and (8) compressed domain based technology, instead of working on the original image sequences, technique to detect changes directly on intra-frame JPEG coded compressed data is develop. A fixed number of connected regions is selected and some predetermined collections of AC coefficients from the 8×8 DCT blocks in the regions are used to form a vector (Sarah De *et al*., 2008; Suchendra and Chandrasekaran, 2004).

The above methods have their merits, respectively, however, the earlier mentioned methods all suffer from following limitations: (1) more above methods processed in uncompressed domain, so it need expensive computation to decoding. This is very time consuming, (2) the shot detection algorithm is subjective or need some additional condition or threshold. These conditions or thresholds are difficulty to determined generally.

THE PROPOSED ALGORITHM BASED ON RS IN COMPRESSED DOMAIN

Basic Concepts of Rough Set

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 $U|R = \{X_1, X_2, \ldots, X_n\}$ in which the following conditions are satisfied:

$$X_i \subseteq U, X_i \neq \phi \text{ for any i.} \tag{1}$$

$$X_i \cap X_j \neq \phi \text{ for any i, j.} \tag{2}$$

$$U_{i=1,2,\ldots,n}, X_i = U \tag{3}$$

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 (4).

$$IND(P) = \bigcap_{R \in P} IND(R) \quad (4)$$

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. 5 and 6.

$$\underline{R}X = \{x \in U : [x]_R \subseteq X\} \quad (5)$$

Equation 5 is the subset of X, such that X belongs to X in R, is the lower approximation of X.

$$\overline{R}X = \{x \in U : [x]_R \cap X \neq \phi\} \quad (6)$$

Equation 6 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. 7-9.

$$POS_R(X) = \underline{R}X \quad (7)$$

$$NEG_R(X) = U - \overline{R}X \quad (8)$$

$$BNR(X) = \overline{R}X - \underline{R}X \quad (9)$$

Attributes Reduction and Core

In RS theory, an 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 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. 10

$$CORE_C(R) = \{c \in C \mid \forall c \in C, POS_{C-\{c\}}(R) \neq POS_C(R)\} \quad (10)$$

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	2.226	2.102	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	1.725	4.1382	4.4809	4.1711	4.2363	4.2368
4.2353	4.2662	4.299	4.2667	4.1078	4.3275	2.699	2.3225
2.7917	2.1343	2.823	1.3946	2.7294	2.3191	3.1054	2.2583

The Proposed Algorithm

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 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.

Extract DCT and 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:

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

Table 2 shows 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 system with it. Each row is a DC coefficient and each column is the frame. This process can be described as following:

$$\text{DC} \xrightarrow{\text{Construct}} \text{information table } S = \{U, A, V, F\}$$

Table 3: Information System constructed using DC coefficients

	Frame							
	1	2	3	4	5	6	7	8
DC_1	0.35355	0.35355	0.27771	0.35355	0.41572	0.35355	0.35355	0.35355
DC_2	0.35355	0.35355	0.27779	0.35355	0.41573	0.35354	0.35355	0.35355
DC_3	0.35355	0.35355	0.27779	0.35356	0.19134	0.35355	0.35355	0.35355
DC_4	0.35356	0.35356	0.27779	0.35355	0.19134	0.35351	0.35355	0.35355
DC_5	0.35355	0.35355	0.27779	0.35355	0.19134	0.35336	0.35355	0.35355
DC_6	0.27779	0.27779	0.27779	0.35356	0.19134	0.35345	0.35355	0.35355
DC_7	0.27779	0.27779	0.27779	0.35355	0.19134	0.35353	0.35355	0.35355
DC_8	0.23761	0.23761	0.27779	0.35355	0.35355	0.35355	0.35355	0.35355

Table 4: The reduced information system

	Frame					
	1	3	4	5	6	7
DC_1	0.35355	0.27771	0.35355	0.41572	0.35355	0.35355
DC_2	0.35355	0.27779	0.35355	0.41573	0.35354	0.35355
DC_3	0.35355	0.27779	0.35356	0.19134	0.35355	0.35355
DC_4	0.35356	0.27779	0.35355	0.19134	0.35351	0.35355
DC_5	0.35355	0.27779	0.35355	0.19134	0.35336	0.35355
DC_6	0.27779	0.27779	0.35356	0.19134	0.35345	0.35355
DC_7	0.27779	0.27779	0.35355	0.19134	0.35353	0.35355
DC_8	0.23761	0.27779	0.35355	0.35355	0.35355	0.35355

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.

Reduct Information System

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 frame 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))
}
}
```

The CORESET represents core of information system. Because the CORESET is the frame that can not to reduce, so it is the salient content in video sequences.

Construct New Information System

After the frame is reduced, the information system are consists of most important frames, we invert the row and column of information table, that is, each row is a frame and each column is the DC coefficients, we can get a new information system as Table 5.

Table 5: The inverted information system

Frame	DC_1	DC_2	DC_3	DC_4	DC_5	DC_6	DC_7	DC_8
1	0.35355	0.35355	0.35355	0.35356	0.35355	0.27779	0.27779	0.23761
3	0.27771	0.27779	0.27779	0.27779	0.27779	0.27779	0.27779	0.27779
4	0.35355	0.35355	0.35356	0.35355	0.35355	0.35356	0.35355	0.35355
5	0.41572	0.41573	0.19134	0.19134	0.19134	0.19134	0.19134	0.35355
6	0.35355	0.35354	0.35355	0.35351	0.35336	0.35345	0.35353	0.35355
7	0.35355	0.35355	0.35355	0.35355	0.35355	0.35355	0.35355	0.35355

Table 6: Evaluating result of shot boundary detection by various video sequences

Video type	Frames	Shots	Manually detected	Falsely detected	Loose detected	Evaluation
Gym	112	7	5	2	0	Excellent
Animation	174	13	14	0	1	Excellent
Senary	317	23	18	5	0	Good
Story	126	13	12	1	0	Excellent
News	153	14	17	3	3	Good

Table 7: Total evaluation of index

Video type	Loose detected rate (%)	Falsely detected rate (%)	Full-detected rate (%)	Accurate rate (%)
Gym	0	40	100	71
Animation	7	0	93	92
Senary	0	27	100	78
Story	0	8	100	92
News	17	17	82	82

Construct Model and Subdivide New Information System

we can also classify all frame into some parts, the model used during classification is defined by Eq. 11.

$$D(l_i, l_{i+1}) = \frac{1}{1024}\sum_{k=1}^{1024}\frac{|c(l_i,k) - c(l_{i+1},k)|}{\max\{c(l_i,k), c(l_{i+1},k)\}} \tag{11}$$

l_i and l_{i+1} represent the i and i+1 frame, c (l_i, k) and $c(l_{i+1}, k)$ are the k block DC coefficients of successive frame. $D(l_i, l)$ is the difference of successive frame.

EXPERIMENTAL RESULTS

Various MPEG video sequences are selected to examine the performance of the proposed algorithm. Figure 1 shown some frames in different shot segmented by proposed algorithm, from it, we can see obviously that all frames are different in visual feature. Figure 2 shows some frames in same shot. we can see clearly that all frames are similar. Table 6 shows the evaluating results of shot boundary detection by various video sequences. By comparison, we can see the proposed algorithm can achieve satisfied results. Table 7 shows the total evaluation of index. To validate the DC coefficients in different shot and in same shot, we also compared the DC data of two frame, the results are shown by Fig. 3. In Fig. 3, (1) and (3) are from same shot, the feature curves are similar, while (2) and (4) come from different shot, the feature curves are different.

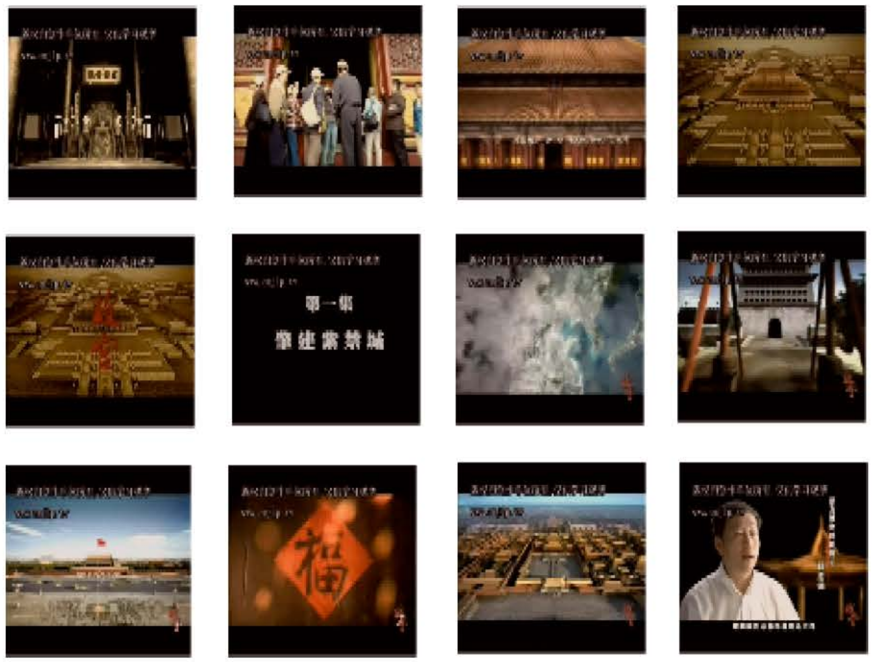

Fig. 1: Frames in different shot

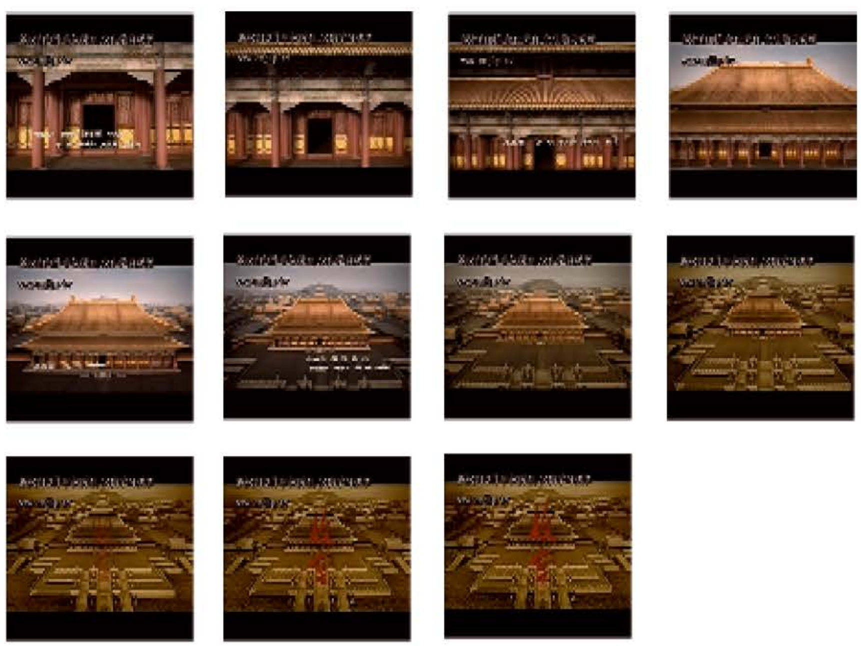

Fig. 2: Frames in same shot

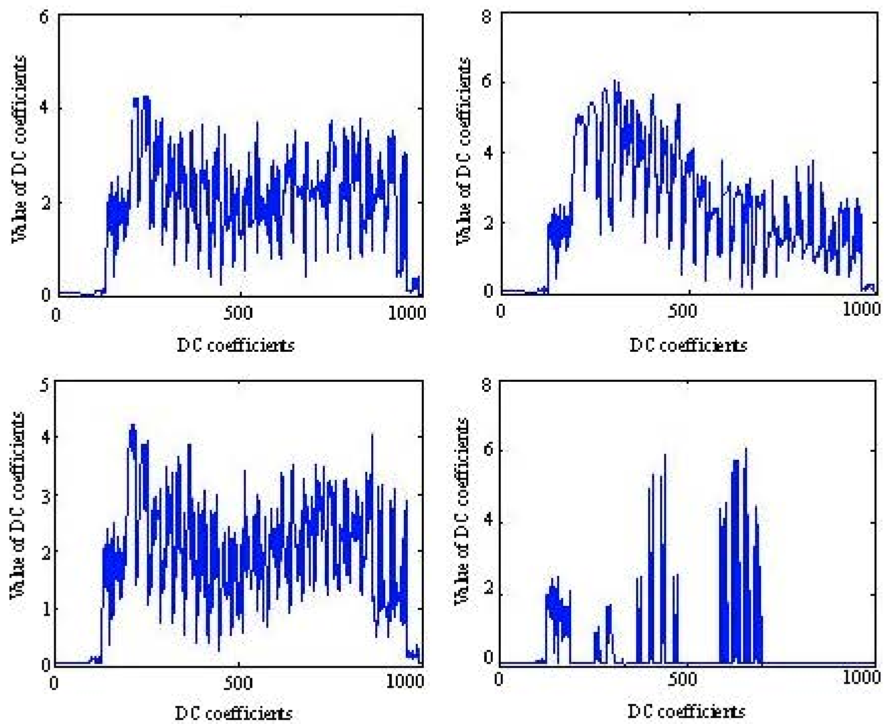

Fig. 3: Comparison of DC coefficients of both different shot and same shot

CONCLUSION

Shot boundary detection is a prerequisite for semantic video analysis, this study proposed a novel algorithm for shot boundary detection. To increase the detection efficiency and make the relativity of shots more scientific, the algorithm introduces theory of RS and reduces the redundant information of shot, then classifies the frame into shot without any prior_ knowledge. The experimental results proved the validity and feasibility of present proposed algorithm.

REFERENCES

Bangjun, L. and L.Q. Xn, 2006. Real-time outdoor video surveillance with robust foreground extraction and object tracking via multi-state transition management. Pattern Recog. Lett., 27: 1816-1825.

Chi-Chun, L. and S.J. Wang, 2001. Video segmentation using a histogram-based fuzzy c-means clustering algorithm. Comput. Stand. Interfac., 23: 429-438.

Haoran, Y., D. Rajan and L.T. Chia, 2006. A motion-based scene tree for browsing and retrieval of compressed videos. Inform. Syst., 31: 638-658.

Ioannis, T., A. Cavallaro and L.Q. Xu, 2008. Video event segmentation and visualisation in non-linear subspace. Pattern Recong. Letters.

Kyungnam, K., T.H. Chalidabhongse and D. Harwood, 2005. Real-time foreground-background segmentation using codebook model. Real-Time Imag., 11: 172-185.

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.

Liang-Hua, C., Y.C. Lai and M.L. Hong-Yuan, 2008. Movie scene segmentation using background information. Pattern Recog., 41: 1056-1065.

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.

Sarah De, B., D.V. Deursen, J.D. Cock, W.D. Neve, P. Lambert and R.Van de Walle, 2008. A compressed-domain approach for shot boundary detection on H.264/AVC bit streams. Image Commun., 23: 473-489.

Suchendra, M.B. and S.R. Chandrasekaran, 2004. Parallel parsing of MPEG video on a shared-memory symmetric multiprocessor. Parallel Comput., 30: 1233-1276.

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.