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 10 (5): 1056-1060, 2011
ISSN 1812-5638 / DOI: 10.3923/itj.2011.1056.1060

A Novel Rough Sets Based Video Shot Clustering Algorithm

[1]Wei Zhe, [1]Li Zhan-Ming, [2]Qi Yan-Fang and [2]Zhao Li-Dong
[1]College of Electrical Engineering and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China
[2]Department of Computer Engineering, Lanzhou Polytechnic College, Lanzhou 730050, China

Abstract: A novel image clustering algorithm was proposed in compressed domain. Firstly, DCT coefficients and DC coefficients are extracted from image consequences and an Information System is achieved by using DC coefficients. Secondly, Information System is reduced by introducing reduction theory of Rough Sets (RS), so the concise representation of the image is obtained by reduced DC coefficients. Finally, by introducing subdividing theory of RS, the image is clustered objectively. Based on the experimental results obtained in this study, the proposed algorithm enjoys following advantages. (1) the image is processed in compressed domain, the computing complexity is decrease. (2) RS is introduced to pre-prepare the image data, the effectiveness is enhanced.

Key words: Video clustering, compressed domain, rough sets, DCT coefficients, DC cpeffocoemts

INTRODUCTION

Contend-based image retrieval has become a very active research area since the introduction of automatic video retrieval system (Ferman and Tekalp, 1998). Advances in communications, multimedia and computer technologies have made video information producing too easily in our daily life, whoever, the sheer volume of video makes it extremely difficult to use and analysis. So it is necessary to develop a efficient technology to analyze, process and index video data automatically. Video mining is typically and effectively method to solve it. Video Shot clustering is the key technology in the video mining which breaks the massive volume of video consequence into smaller chunks. Each shot represents an event of actions.

Many theories and technologies are presented for shot clustering, 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 clustering directly in compressed domain.

The study proposes a novel video shot clustering algorithm that operate directly in compressed video for shot clustering. The algorithm overcomes the limitations of previous approaches and increases the efficiency obviously.

CONVENTIONAL ALGORITHM OF VIDEO CLUSTERING

There have been many considerable work reported on shot clustering. Different methods and technologies are used to examine the changes between successive frames and determine whether changes have taken place (Antani *et al*., 2002). 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 χ^2 comparison of color histograms (Lo 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 (Yi *et al*., 2006; Dhillon *et al*., 2009); (3) model classification or neural network based technology, by constructing model or neural network train and capture the different type of shot transitions(Kim *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 and Agius, 2007); (5) object based technology,

Corresponding Author: Wei Zhe, College of Electrical Engineering and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China

employing the coding scheme of MPEG-4, extract the object of video to examine the boundary of shots such as localized and global lighting changes, variations in object size, occlusions and complex object motion and so on (Lei and Xu, 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 (Chen *et al*., 2008); (7) event based technology, use of dimensionality reduction for video event detection without explicitly using motion estimation or object tracking (Tziakos *et al*., 2008; Choi *et al*., 2002) 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 are selected and some predetermined collections of AC coefficients from the 8×8 DCT blocks in the regions are used to form a vector (De Bruyne *et al*., 2008; Bhandarkar and Chandrasekaran, 2004).

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

Rough Sets is a powerful and intelligent tool for data analysis, it has successfully been used in many applications such as Machine Learning, Expert Systems and Pattern Classification· it can classify object without any prior knowledge. So it can detect shot boundary easily and efficiently.

THE PROPOSED ALGORITHM

Rough sets theory used in this algorithm: 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 $U / R = \{X1, X2, ..., X_n\}$ 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
- $U_i = 1,2,..., n, X_i = U$

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 byIND (P):

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

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. 4 and 5:

$$RX = \{x \in U : [x]_R \subseteq X\} \tag{2}$$

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

$$\bar{R}X = \{x \in U : [x]_R \cap X \neq \phi\} \tag{3}$$

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 approximate $\bar{R}$ on contains sets that are possibly included in X. R-positive, R-negative and R-boundary regions of X are defined respectively by Eq.6-8:

$$POS_R(X) = RX \tag{4}$$

$$NEG_R(X) = U - \bar{R}X \tag{5}$$

$$BNR(X) = \bar{R}X - RX \tag{6}$$

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:

Table 1: Information system constructed using DC coefficients

	Frame							
Coefficient	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

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

And the Core can be defined by Eq. 9:

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

Main steps of algorithm

Extraction of DCT coefficients from DCT domain: 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, the processing of DCT and IDCT are showed as expressions (8) and expression (9), so the DCT coefficients can be easily extracted from video sequences directly. We can represent this process as representation Eq. 10:

$$F(u,v)=\frac{2}{N}c(v)\sum_{i=0}^{N-1}\sum_{i=0}^{N-1}f(i,j)\cos\left[\frac{(2i+1)u\pi}{2N}\right]\cos\left[\frac{(2j+1)v\pi}{2N}\right] \qquad (8)$$

$$f(i,j)=\frac{2}{N}\sum\sum c(u)c(v)F(u,v)\cos\frac{(2j+1)v\pi}{2N}\cos\frac{(2j+1)v\pi}{2N} \qquad (9)$$

$$\psi(P(x),t)\xrightarrow{\text{extract}}\text{DCT coefficients} \qquad (10)$$

where, Ψ (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 representation (Eq. 11):

$$\text{DCT coefficients}\xrightarrow{\text{reprecessing}}\text{DC coefficients} \qquad (11)$$

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 representation (Eq. 12):

$$DC\xrightarrow{\text{construct}}\text{information table } S=\{U,A,V,f\} \qquad (12)$$

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 1.

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 1, we can see that the frame 2 and frame 8 can be reduced; the reduced Information System is showed as Table 2. if we introduce a threshold, more attributes can be reduced.

The reduced frame called redundant attributes; they have no effect as to subdividing of DC coefficients, such as frame 2 and frame 8. The attributes that can not reduce called CORESET in information system, which represents the salient content in video sequences.

After the process the volume of video data is reduced dramatically. While the main information of video is remained, so the efficiency of following processing is increased.

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 3.

Construct model and subdivide new information system: Use the theory 2.3, we can also classify all frame into some parts, the model used during classification is defined by Eq. 13:

Table 2: The reduced information system

	Frame					
Coefficient	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

Table 3: The inverted information system

	DC							
Frame	1	2	3	4	5	6	7	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 4: Evaluating result of shot cluster by various video sequences

Video type	Frames	Auto cluster	Manually cluster	Falsely cluster	Loose cluster	Recall (%)
Gym	65	7	4	1	2	66
Animation	83	19	14	2	3	82
Senary	142	26	18	5	3	85
Story	102	16	12	3	1	92
News	126	23	17	3	3	85

Table 5: Total evaluation

Video type	Loose detected rate	Falsely detected rate	Precision (%)
Gym	28	14	80
Animation	15	10	87
Senary	19	19	78
Story	18	18	80
News	13	13	85

$$D(I_i, I_i) = \frac{1}{1024} \sum_{k=1}^{1024} \frac{|c(I_i,k) - c(I_{i+1},k)}{\max\{c(I_i,k), c(I_{i+1},k)\}} \quad (13)$$

I_i and I_{i+1} represents the Ith and i+1 th frame, c (I_i k) and c (I_{i+1}, k)are kth block DC coefficients of successive frame. D (I_i, I) is the difference of successive frame.

EXPERIMENTAL RESULTS

A consistent evaluation criterion is crucial to shot clustering. So we briefly discuss the evaluation criteria we have employed in following experiments.

A good clustering method should minimize the number of false detections while maximizing the number of correctly identified shot boundaries, originally introduced to assess the performance of information retrieval systems, recall and precision have also been used, perhaps for lack of better alternatives, as evaluation criteria for shot clustering methods:

$$\text{Recal} = \frac{\text{Correct}}{\text{Correct} + \text{Missed}} \quad (14)$$

$$\text{Precision} = \frac{\text{Correct}}{\text{Correct} + \text{falsealarms}} \quad (15)$$

Various MPEG video sequences are selected to examine the performance of the proposed algorithm. Table 5 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 4 and 5 show the total evaluation of proposed algorithm.

CONCLUSIONS

Video shot clustering is a prerequisite for semantic video analysis, the study proposed a novel algorithm for shot clustering in compressed domain. To increase the clustering efficiency and make the results more scientific, the algorithm introduces theory of RS and reduces the redundant information of shot, then classifies the remain frame into cluster without any prior-knowledge. The experimental results proved the validity and feasibility of our proposed algorithm.

REFERENCES

Antani, S., R. Kasturi and R. Jain, 2002. A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video. Pattern Recog., 35: 945-965.

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

Chen, L.H., Y.C. Lai and H.Y. Mark-Liao, 2008. Movie scene segmentation using background information. Pattern Recog., 41: 1056-1065.

Choi, Y., S.J. Kim and S. Lee, 2002. Hierarchical shot clustering for video summarization. Proceedings of International Conference on Computational Science, April 21-24, Springer-Verlag, London, UK., pp: 1100-1107.

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

Dhillon, P.S., S. Nowozin and C.H. Lampert, 2009. Combining appearance and motion for human action classification in videos. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, June 20-25, Miami, USA., pp: 935-942.

Ferman, A.M. and A.M. Tekalp, 1998. Efficient filtering and clustering methods for temporal video segmentation and visual summarization. J. Visual Commun. Image Represent., 9: 336-351.

Kim, 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. Exp. Syst. Appl., 31: 13-25.

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

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

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

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

Tziakos, I., A. Cavallaro and L.Q. Xu, 2008. Video event segmentation and visualisation in non-linear subspace. Pattern Recong. Lett., 30: 123-131.

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