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 (4): 610-614, 2009
ISSN 1812-5638

Rough Sets based Temporal-spatial Color Descriptor Extraction Algorithm in Compressed Domain for Video Retrieval

[1,2]Li Xiang-Wei, [1]Li Zhan-Ming, [3]Zhang Ming-Xin, [2]Wang Yi-Ju and [2]Zhang Zhi-Xun
[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
[3]College of Mathematics and Information Science, Northwest Normal University, Lanzhou 730070, China

Abstract: In this study, based on Rough Sets (RS), a compact and efficient temporal-spatial color descriptor extraction algorithm is developed in compressed domain. Firstly, Discrete Cosine Transform (DCT) coefficients and Direct Current (DC) coefficients, the most important video visual features are extracted from raw video sequences to represent video information. Secondly, an information system table is constructed using DC coefficients. Thirdly, a novel and concise information system table is achieved by using the reduction theory of RS, i.e., core of information system. The core contains major visual color information and eliminates the redundant video information. Furthmore, DC coefficients also contain important spatial information of each frame, so the core of information system can regard as effective temporal-spatial color descriptor for video retrieval. Compared to existing technologies, the proposed algorithm enjoyed the following three advantages: the extracted descriptor consider not only visual color feature and temporal information, but also spatial information of each frame, the algorithm introduced attributes reduction theory of RS and the more redundant video information are eliminated and the whole process accomplished in compressed domain, so the volume of video data also decreased dramatically. Effectiveness is documented by experimental results.

Key words: Rough sets, audio-video object, descriptor, MPEG-7

INTRODUCTION

There has been a greatly increase in the volume of video information in past few years after the national standard of MPEG-1 and MPEG-2 are developed (Dorairaj and Namuduri, 2004). MPEG-4 introduced the concept of Audio-Video (AV) object, which enhanced the storage efficiency of video data. At the same time, video information were efficiently indexed and retrieved by various content based indexing and retrieval algorithm until MPEG-7 (Sikora, 2001). The major objective of MPEG-7 standard is to allow interoperable searching, indexing, filtering and access of AV content by enabling interoperability among devices and applications that deal with AV content description. MPEG-7 describes specific features of AV content, as well as information related to AV content management (Salembier and Smith, 2001).

For the convenience of video retrieval, there are several visual descriptors defined by MPEG-7 standards (Manjunath *et al*., 2001). Among the low level features, color is the most expressive and effective visual content description, especially, the use of low-level visual color features to find interesting information from video database or data warehouse has drawn much attention in recent years. More detailed information regarding the color descriptors in MPEG-7 may be found in the references and other related MPEG document (Manjunath *et al*., 2001). Color histogram is also the most widely used color descriptor in content based video retrieval. A color histogram captures global color distribution in a frame of video sequences, but the performance is quit limited due to the lack of spatial color distribution information. Among various MPEG-7 color visual descriptors, Color Structure Descriptor (CSD) always gives the satisfying retrieval rate. The major reason is it can use spatial color distribution information in CSD. However, CSD always has redundancy information as it uses a fixed color space for the histogram representation (Wong *et al*., 2007). To solve this problem in this study, a novel data analysis tool is introduced, i.e., attributes reduction theory of RS. It can effectively overcome the redundant information and preserve the temporal spatial information of video frame sequences.

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

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 and Qiang, 2008). The main advantage of RS 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. (Wang *et al*., 2008; Yao and Zhao, 2008). The developed algorithm used theory of RS and effectively overcome the deficiency of earlier technologies in compressed domain.

THE FUNDAMENTAL THEORY OF ROUGH SETS AND ATTRIBUTES REDUCTION

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 $U/R = \{X_1, X_2,...X_n\}$ 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. 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 by Eq. 4-6, respectively.

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

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

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

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 consists of the following five steps. Figure 1 show the flow chart of the 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 information in sequence, which uses 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:

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

where, Ψ(P (x), t) denotes the video sequences and P(x) is the frame of video. Table 1 shows part of 8×8 block DCT coefficients extracted from video sequences.

Table 1: Part of DCT coefficients extracted from video sequences

Video							
1	2	3	4	5	6	7	8
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

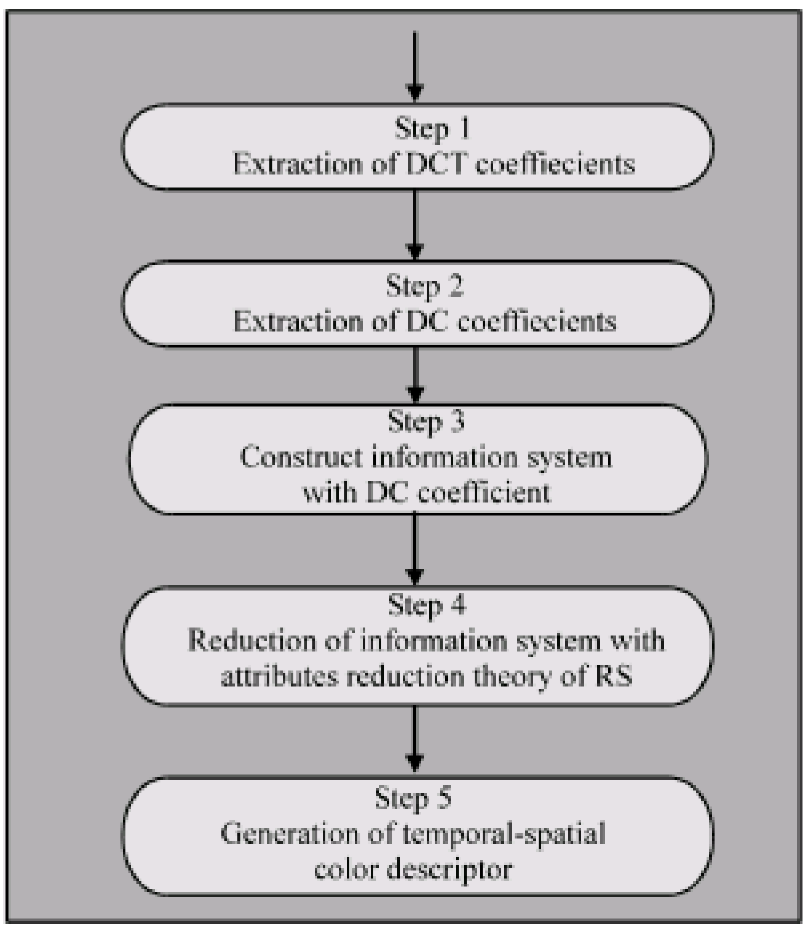

Fig. 1: The flow chart of developed algorithm

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

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

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

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

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

where, U is a set which denotes all object of information system. A is a set which denotes all attributes in information system, V is the set of attributes value and f is a function denotes the relations between objects and attributes.

Table 2: 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

By using above process, an information system is achieved as Table 3.

Reduction of information system with attributes reduction of 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 shown in Table 4. If we introduce a threshold, more attributes can be reduced.

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

algorithm: Generation of_Coresets (GC)
input: Original information table
output: Reduced attributes sets (Corsets)
processing

Step 1:

For i = 1 to n // n is the number of information table recorder
{
for j = 1 to m // m is the number of information table attributes
{
If $p_{ik} = p_{jk}$ //k = 1,2…n
$m_{ij} = \Phi$
Else
$m_{ij} = \{a \in A | f(x_i, a) \neq f(x_j, a)\}$
i, j = 1,…n //m_{ij} is the sets of all attributes that can discern x_i and x_j
}
}

Table 3: Information system constructed using DC coefficients

DC coefficient	Frame 1	2	3	4	5	6	7	8
1	0.35355	0.35355	0.27771	0.35355	0.41572	0.35355	0.35355	0.35355
2	0.35355	0.35355	0.27779	0.35355	0.41573	0.35354	0.35355	0.35355
3	0.35355	0.35355	0.27779	0.35356	0.19134	0.35355	0.35355	0.35355
4	0.35356	0.35356	0.27779	0.35355	0.19134	0.35351	0.35355	0.35355
5	0.35355	0.35355	0.27779	0.35355	0.19134	0.35336	0.35355	0.35355
6	0.27779	0.27779	0.27779	0.35356	0.19134	0.35345	0.35355	0.35355
7	0.27779	0.27779	0.27779	0.35355	0.19134	0.35353	0.35355	0.35355
8	0.23761	0.23761	0.27779	0.35355	0.35355	0.35355	0.35355	0.35355

Table 4: The reduced information system

DC coefficient	Frame 1	3	4	5	6	7
1	0.35355	0.27771	0.35355	0.41572	0.35355	0.35355
2	0.35355	0.27779	0.35355	0.41573	0.35354	0.35355
3	0.35355	0.27779	0.35356	0.19134	0.35355	0.35355
4	0.35356	0.27779	0.35355	0.19134	0.35351	0.35355
5	0.35355	0.27779	0.35355	0.19134	0.35336	0.35355
6	0.27779	0.27779	0.35356	0.19134	0.35345	0.35355
7	0.27779	0.27779	0.35355	0.19134	0.35353	0.35355
8	0.23761	0.27779	0.35355	0.35355	0.35355	0.35355

Step 2:

```
For i = 1 to n
  {
   for j = 1 to m
   {
     If and [x_i] A∩ ⌊x_j⌋ A = Φ
        Output A
   }
}
```

A is the corsets of information system. The corsets represent core of information system. Because the corsets are the frames that can not be reduced, so it is the salient and interesting content in video sequences.

Generation of temporal-spatial color descriptor: Core is the most important attributes in information system which can not be reduced. It represents the most important and interesting visual color information in video sequences. Since, core eliminate many redundant frames, the volume of video shortened dramatically. At the same time, the DC coefficients of row in reduced information system contain not only the color feature, but also the spatial information of each frame. In addition, the temporal information is also considered in core. So, the core of information system table can regard as effective and feasible visual descriptor for video retrieval.

EXPERIMENTATION

To validate the proposed algorithm, we constructed a large video library manually and selected five types of ordinary MPEG-2 video sequences. Some of video sequences are selected to examine the performance of proposed algorithm. Figure 2 shows port of temporal-spatial color descriptor generated by proposed algorithm. Table 5 shows the detailed evaluation results of four different video shots.

Table 5: Detailed evaluation results of four sports video

Video shot No.	Frames in original shot	Frame in core	Evaluation
1	112	17	Good
2	174	33	Good
3	317	42	Excellent
4	126	13	Good

The experiment results demonstrate that proposed algorithm can effective and scientific generates temporal-spatial color descriptor in compressed domain. It considered not only temporal information by using a series of frame, but also spatial information by using AC coefficients in each frame. It resolves the limitation of previous algorithm effectively. In addition, all operation accomplished in the compressed domain, it does not need decompress of the original video, the amount of data reduced dramatically, so the algorithm is very efficient and scientific.

CONCLUSIONS

In this study, a novel RS based temporal-spatial color descriptor extraction algorithm is developed in compressed domain. DCT coefficients and DC coefficients are extracted directly from raw video sequences and an information system constructed to represent video shot. Then attributes reduction theory of RS is introduced for redundant data processing. At last, a concise and important core of information system generated to represent temporal-spatial color descriptor. Since, the generated descriptor considered three main factor of color, spatial and temporal, it can effective used in video retrieval and many applications. Present experiments have shown that better results are achieved using our testing video.

Fig. 2: Temporal-spatial color descriptor generated by proposed algorithm

ACKNOWLEDGMENT

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

REFERENCES

Dorairaj, R. and K.R. Namuduri, 2004. Compact combination of MPEG-7 color and texture descriptors for image retrieval. Proceedings of the 38th Asilomar Conference on Signals, Systems and Computers, 2004, pp: 387-391.

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

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.

Manjunath, B.S., J.R. Ohm and V.V. Vasudevan, 2001. Color and texture descriptors. IEEE Trans. Circuits Syst. Video Technol., 11: 703-715.

Salembier, P. and J.R. Smith, 2001. MPEG-7 multimedia description schemes. IEEE Trans. Circuits Syst. Video Technol., 11: 748-759.

Sikora, T., 2001. The MPEG-7 visual standard for content description-an overview. IEEE Trans. Circuits Syst. Video Technol., 11: 696-702.

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.

Wong, K.M., L.M. Po and K.W. Cheung, 2007. Dominant color structure descriptor for image retrieval. Proceedings of IEEE International Conference on Image Processing, San Antonio, TX, Sept. 16-Oct. 19, IEEE Computer Society USA., VI-365-VI-368.

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