International Journal of Information Technology, Vol. 11, No. 7, 2005

A Shot Boundary Detection Method for News Video Based on Rough-Fuzzy Sets [1]

Bing Han[1], Xinbo Gao[2], and Hongbing Ji[3]

School of Electronic Engineering, Xidian Univ., Xi'an 710071, China
[1] hanbing@lab202.xidian.edu.cn, [2] xbgao@lab202.xidian.edu.cn,
[3] hbji@xidian.edu.cn

Abstract

With the rapid growing amount of multimedia, content-based information retrieval has become more and more important. As a crucial step in content-based news video indexing and retrieval system, shot boundary detection attracts much more research interests in recent years. To partition news video into shots, many metrics were constructed to measure the similarity among video frames based on all the available video features. However, too many features will reduce the efficiency of the shot boundary detection. Therefore, it is necessary to perform feature reduction for every decision of shot boundary. For this purpose, the rough-fuzzy operator based on rough-fuzzy sets for feature reduction and the dissimilarity function is proposed. According to the characteristics of news scenes, shot transition can be divided into three types: *cut transition*, *gradual transition* and *no transition*. The efficacy of the proposed method is extensively tested on more than 2 *h* of news programs and 98.0% recall with 96.6% precision have been achieved.

Key words: Shot boundary detection, rough sets, rough-fuzzy sets, news video, automatic threshold.

1 Introduction

With the increasing proliferation of digital video contents, efficient techniques for analysis, indexing, and retrieval of videos according to their contents have become evermore important. A common first step for most content-based video analysis techniques available is to segment a video into elementary shots, each comprising a continuous in time and space. These elementary shots are composed to form a video sequence during video sorting or editing with either cut transitions or gradual transitions of visual effects such as fades, dissolves, and wipes.

In recent years, a large number of metrics have been proposed to segment a video into shots by measuring the dissimilarity, or distance, between two or a short se-

[1] This work was supported by National Natural Science Foundation of China (No.60202004) and the Key Project of Chinese Ministry of Education (No.104173)

quence of adjacent frames [1] – [3]. These metrics make use of such frames or video features as pixel values, statistic features, intensity and color histogram and *etc*. If the measured dissimilarity is greater than some predetermined threshold, the shot boundary is assumed. How to adequately use the features available is becoming the hot topic on shot boundary detection. To improve the detection efficiency with keeping the detection accuracy, it is necessary to perform feature reduction for every decision of shot boundary. To this end, the rough-fuzzy set based feature reduction and the dissimilarity function is presented in this paper. First, features of video sequences used as conditional attributes are extracted and the decision attributes are given according to the coarse detection [4], by calculating the correlation between conditional attributes, the importance of conditional attributes in the rough set can be obtained. Due to the set of feature values are fuzziness, the class precision of rough-fuzzy is given. Then, we define the new importance of conditional attributes as rough-fuzzy operator which is the product of the importance of conditional attributes in the rough set and the class precision of rough-fuzzy. According to the proportion of each feature, the top k features can be obtained. The dissimilarity function is generated for by weighting these important features in term of their proportion in the whole feature. The preliminary experimental results with real news videos demonstrate the effectiveness of the proposed scheme.

The rest of this paper is organized as follows. First, the basic theory of rough sets is introduced in Section 2. Section 3 describes the proposed shot boundary detection method based on Rough-Fuzzy Set. The experimental results and analysis are presented in Section 4. Finally, conclusions and future works are given in Section 5.

2 Basic Concepts of Rough Sets and Rough-Fuzzy Set

The rough sets theory introduced by Pawlak in the early 1980s[5-10] is an effective mathematical analysis tool to deal with vagueness and uncertainty in the areas of machine learning, knowledge acquisition, decision analysis, knowledge discovery from database, expert systems, decision support systems, inductive reasoning, and pattern recognition.

In this section, the basic notions of rough sets theory will be first introduced.

Definition 1: Let R be an equivalence relation on a universal set X. Moreover, let X/R denote the family of all equivalence classes introduced on X by R. One such equivalence class in X/R, which contains $x \in X$, is designated by $[x]_R$. For any output class $A \subseteq X$, we can define the lower and upper approximations, denoted as $\underline{R}(A)$ and $\overline{R}(A)$, which approach A as closely as possibly from inside and outside, respectively. Here, the union of all equivalence classes in X/R that are contained in A and the union of all equivalence classes in X/R that overlap with each other in A are defined respectively as

$$\underline{R}(A) = \bigcup \{[x]_R \mid [x]_R \subseteq A, x \in X\} . \quad \textbf{(1)}$$

$$\overline{R}(A) = \bigcup \{[x]_R \mid [x]_R \cap A \neq \phi, x \in X\} . \quad (2)$$

A rough set $R(A) = \left\langle \underline{R}(A), \overline{R}(A) \right\rangle$ is a representation of the given set *A* by $\underline{R}(A)$ and $\overline{R}(A)$. The set *BN*(*A*)= $\overline{R}(A) - \underline{R}(A)$ is a rough description of the boundary of *A* by the equivalence classes of *X*/*R*. The approximation is rough uncertainty free if $\overline{R} = \underline{R}$. Thus, when all the patterns from an equivalence class do not carry out the same output class label, rough ambiguity is generated as a manifestation of the one-to-many relationship between that equivalence class and the output class labels.

Definition 2: Let *U* be a finite set of *objects* called the universe. *A* is a finite set of *attributes*, and *V* is a set of attribute values, where $V = \bigcup_{a \in A} V_a$, V_a is called the domain of *a*. *f* is an *information function*, $f : U \times A \to V$, where for every $x \in U$ and $a \in A$, $f(x,a) \in V_a$. By the *information system* we will understand a quadruple $S = < U, V, f, A >$. Then, a decision table is defined as an information system, $A = (U,V, f, C \cup D)$. The positive region of *C* to *D* is defined as

$$\mathrm{POS}_C(D) = \bigcup_{X \in U/IND(D)} \underline{C}(X) . \quad (3)$$

An instance belongs to $\mathrm{POS}_C(D)$ only if its decision value vector can be completely predicted according to its corresponding condition attribute value vectors, that is, $\mathrm{POS}_C(D)$ is the deterministic part of the universe of the system.

In condition attributes, some attributes play an important role for classification results, while the others may not be more effective on classification results. So the importance or dependence of attributes can be used to measure the importance of classification results and the attributes can be deducted.

Definition 3: Let *P* and *Q* be the attribute sets (such as condition attributes and decision attributes). $P \subseteq R, Q \subseteq R : U/P = \{X_1, X_2, \cdots, X_l\}$, $U/Q = \{Y_1, Y_2, \cdots, Y_l\}$. We say that *Q* depends on *P* in the degree of *k* on *P* if

$$k = \gamma_P(Q) = \frac{|\mathrm{POS}_P(Q)|}{|U|} . \quad (4)$$

where $|\cdot|$ denotes the cardinality of a given set. Thus, the coefficient *k* expresses the ratio of all elements of the universe which can be properly classified into blocks of the partition *U*/*I*(*Q*), employing attributes *P*. It can be dealt with consistency of information.

Definition 4: Let *X* be a set, *R* be an equivalence relation defined on *X* and the output class *A* be a fuzzy set. A rough-fuzzy set is a tuple ($\underline{A}$, $\overline{A}$) where the lower approximation $\underline{A}$ and the upper approximation $\overline{A}$ of *A* are fuzzy sets of *X*/*R*, with membership functions defined by

$$\underline{A}(x) = \inf\{A(y) \mid y \in [x]_R\}, \quad x \in U . \quad (5)$$

$$\overline{A}(x) = \sup\{A(y) \mid y \in [x]_R\}, \quad x \in U \ . \qquad \textbf{(6)}$$

$\underline{A}$ can be comprehended as the membership degree of object *x* which must belong to fuzzy set *A*, while $\overline{A}$ is the membership degree of object *x* which may be belong to fuzzy set *A*.

Definition 5： Let X be a set, R be an equivalence relation defined on X， $A \in F(U)$ ， the definition of class accuracy $\eta_R(A)$ of *A* is

$$\eta_R(A) = {|\underline{A}|}\big/{|\overline{A}|} \ . \qquad \textbf{(7)}$$

where $|\underline{A}| = 0$， then $\eta_R(A) = 1$.

3 Shot Boundary Detection Scheme Based on Rough-Fuzzy Set

As mentioned in literatures, the selected low-level features are essential to achieve high accuracy for shot boundary detection. But there are too many features available in the frame or video, such as pixel values of different color channels, statistic features, intensity and color histogram *etc*. By choosing the most appropriate features to represent a shot or video, the computational burden will be reduced and the efficiency will be improved. However, it is not certain that which set of features is most effectiveness for detecting shot boundary. For this purpose, the feature optimal choice method based rough sets is introduced in this section.

To detect the video shot boundaries, 12 candidate features, classified into 5 types, are usually extracted for common use [11] – [13]. The first is the RGB space model, the changes of three colors during shot transition can be measured; The second is HSV space model, the component of which can be measured to the changes of hue, saturation and value between adjacent frames. In computation, we compute the mean of every component of each frame in the RGB or HSV model. The histogram features is categorized into two types: grey histogram and color histogram, which are our third and forth types of features. Finally, the statistic feature is considered as the fifth. The mean, variance and skewness of lightness component *V* in each frame are computed. The details are given as follows:

(1) The red (R), green (G) or blue (B) component in RGB model respectively;
(2) The hue (H), saturation(S) or value (V) component in HSV model respectively;
(3) Gray-histogram (G-H);
(4) Color-histogram: the color histogram of RGB model (Rgb-H) and the color histogram of HSV model (Hsv-H) respectively;
(5) Statistic features: mean (M), variance (St) and skewness (P).

Here, the extracted features from news videos are served as condition attribute set $P = \{c_1, c_2, \cdots, c_n\}$ and the corresponding transition types of shots is defined as decision attribute set, *Q* = *d*, *n* is the number of all feature. By the **definition 3** and definition 5, the new importance of attributes is defined as follows

$$\kappa_i = \eta_{P_i} * \lambda_{P_i}(Q), \ i = 1, 2, \cdots, n \ . \qquad \textbf{(8)}$$

The κ_i is changing to by the order from the maximum to the minimum. If $\sum_{j=1}^{k}\kappa_j \Big/ \sum_{j=1}^{n}\kappa_j \geq T$ ($j=1,2,\cdots,\ k$), then the top k attributes are of importance.

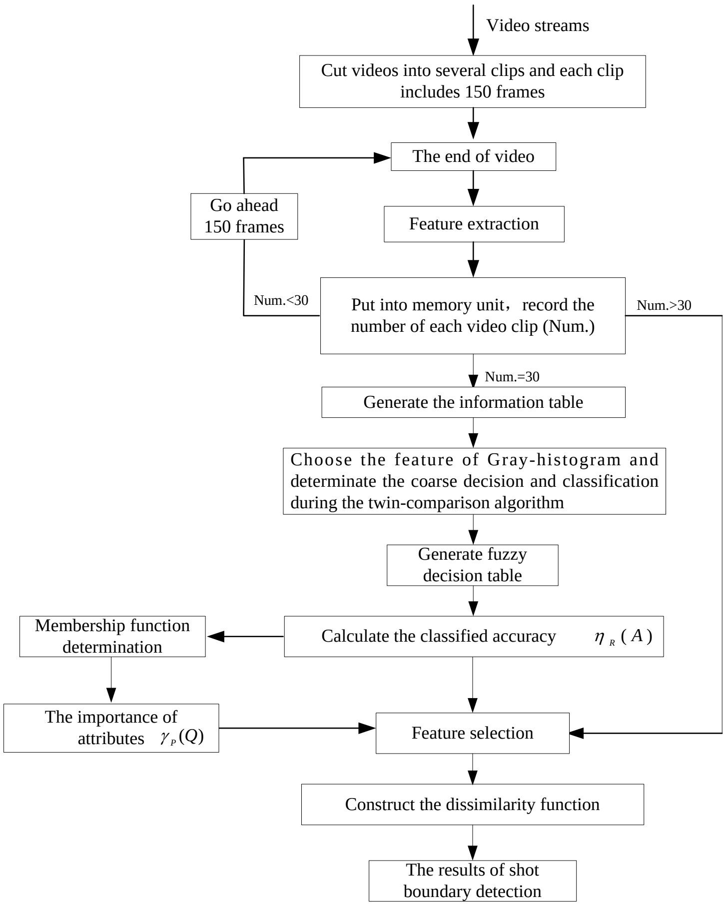


Fig.1. The detection procedure for shot boundary detection

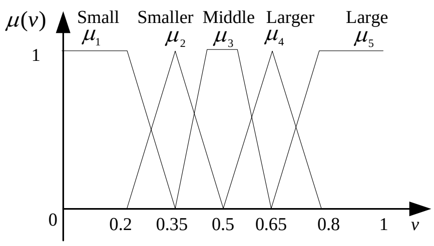


Fig. 2. Definition of the membership function

According to the analysis above, the dissimilarity function for shot boundary detection is defined as

$$S = \omega_i V_j \,. \tag{9}$$

where V_j is the j^{th} attribute value and weighting coefficients is $\omega_i = \kappa_j \Big/ \sum_{j=1}^{k} \kappa_j$, $j = 1, 2, \cdots,\ k$.

Due to the great capacity of video data, the computer cannot deal with a lot of data once. So, the video is partitioned into several clips. During a mount of observation and experiments, a little news unit often lasts less than 150 seconds and the shot transition is no more than 5 seconds. Therefore, we select 150 frames in length and deal with 30 units video clips first, then the information and decision table is generated. The detection procedure is shown in the Fig. 1 in detail where the membership function is shown in Fig. 2.

4 Experimental Results and Analysis

The method described above is applied to 5 news programs from CCTV-1, whose frame size is 352×240 and frame rate is 30 frames per second, which include cut, fade and dissolve, as well as zoom, pan and other camera motions and object motions. The detail information is shown in the Table 1.

Table 1 The detailed information on the news video programs

News video	1	2	3	4	5	Total
Duration	34’55”	30’2”	36’26”	24’47”	15’44”	141’54”
Video shots	374	323	396	247	187	1527
Cut shots	339	315	390	237	178	1459
Gradual shots	35	8	6	10	9	68

Table 2 The comparison of our method with the twin comparison method [14]

Program video		News 1	News 2	News 3	News 4	News 5	Aver-
Proposed method	Hits	372	321	390	246	182	1511
	Misses	12	10	18	7	8	55
	False alarms	10	8	12	6	3	39
	Recall	97.3%	97.6%	97.0%	97.5%	98.3%	97.4.%
	Precision	96.8%	96.9%	95.4%	97.1%	95.7%	96.3%
The twin comparison method[14]	Hits	380	350	400	250	190	1570
	Misses	30	15	46	22	17	130
	False alarms	36	42	50	25	20	173
	Recall	91.8%	95.1%	88.2%	91.0%	90.8%	91.4%
	Precision	90.3%	88.1%	87.2%	90.2%	89.3%	89.0%

Note: Rights = Hits-False alarms.

To verify the effectiveness of our method, we conduct an experiment with the twin comparison method [14] on the same video clips. The experimental results are summarized in Table 2. To evaluate the performance of the proposed scheme of gradual transition segmentation, we use the standard *recall* and *precision* criteria as follows,

$$\text{Recall} = \frac{\text{number of hits}}{\text{number of hits} + \text{number of false alarms}}. \quad \textbf{(10)}$$

$$\text{Precision} = \frac{\text{number of hits}}{\text{number of hits} + \text{number of misses}}. \quad \textbf{(11)}$$

Based on the detected 1,527 shot boundaries, we achieve 97.4% recall with 96.3% in precision in which there are 53 misses and 31 false alarms. And there are two types of false detections in videos. One results from the existence of irregular camera operations during the gradual transitions. The other is due to a lot of flash effects in a shot. The misses are mainly due to the small content changes between the frame pairs at some shot boundaries. During the experiments, we find that the false detection in the coarse detection will affect on the following feature extraction and shot boundary detection while the missed detection have less effects because the rough-fuzzy calculator weaken the mistakes in coarse detection stage, the dissimilarity function is more fit for varies video.

Taking news 3 as an example to analyze the performance of the proposed method, there exist several false alarms when the Twin Comparison method[14] is used to detect the first 30 video clips, that is , there are the false classifications in the decision attributes. In news 3, there is a cut transition in No.8 news clips which is false detected as No Transition because of the similarity structure of histogram during transition. A dissolve transition in No.14 news is false detected as a No Transition. No.25 news is a zoom of camera and it is a No Transition, but it is determined as a gradual transition. Therefore, there are many errors if the importance of attributes is calculated directly. And the accuracy of detection results will decrease if the dissimilarity function is defined only by the importance of attributes.

Now, we analyze the effect of Twin Comparison method on the proposed method. Firstly, according to the membership of all attributes value, the importance of every conditional attributes can be calculated viz. $\lambda'_{P_1} = 11/30$, $\lambda'_{P_2} = 11/30$, $\lambda'_{P_3} = 8/30$, $\lambda'_{P_4} = 16/30$, $\lambda'_{P_5} = 16/30$, $\lambda'_{P_6} = 11/30$, $\lambda'_{P_7} = 8/30$, $\lambda'_{P_8} = 11/30$, $\lambda'_{P_9} = 16/30$, $\lambda'_{P_{10}} = 3/30$, $\lambda'_{P_{11}} = 11/30$, $\lambda'_{P_{12}} = 23/30$. According to $\sum_{j=1}^{k} \lambda_{P_j} \Big/ \sum_{j=1}^{n} \lambda_{P_j} \geq T$, threshold $T = 1/2 \cdot \sum_{j=1}^{n} \lambda_{P_j}$, the important attributes can be obtained, *P*, *H*, *Hsv-H* and *S*. If these four features are only used to detect the shot boundary, we only achieve 93.8.0% recall with 92.2% in precision. Assuming the right result obtained by the Twin Comparison method, that is, we can receive the right original decision. The importance of each conditional attributes is given as follows, $\lambda_{P_1} = 19/30$, $\lambda_{P_2} = 19/30$, $\lambda_{P_3} = 18/30$, $\lambda_{P_4} = 20/30$, $\lambda_{P_5} = 10/30$, $\lambda_{P_6} = 20/30$, $\lambda_{P_7} = 19/30$, $\lambda_{P_8} = 21/30$, $\lambda_{P_9} = 22/30$, $\lambda_{P_{10}} = 15/30$, $\lambda_{P_{11}} = 17/30$, $\lambda_{P_{12}} = 27/30$. The important attributes obtained are *P*, *HSV*, *RGB*, *V*, *H*. Due to the same classification ability of *V* and *H*. *V* is more sensitive than *H* in human vision.

So, the final choice of features are *P*, *Hsv-H*, *Rgb-H*, *V*. we can achieve 95.2% recall and 93.4% precision by these four features. Obviously, the efficiency will be decreased by using improper features. This results from the important feature *S* (saturation) in the false decision by Twin Comparison method. So it is not fit for our experience. Namely, the changes of saturation (*S*) are small while the changes of value (*V*) are great during the shot transition. So, the original decision by the Twin Comparison method cannot be used to calculate the importance of attributes and reduct the attributes directly. For this purpose, the fuzzy-rough operator is defined. According to the analysis above, fuzzy classification accuracy can be obtained by the fuzzy membership as follows, $\eta_{P_1} = 0.77$, $\eta_{P_2} = 0.49$, $\eta_{P_3} = 0.76$, $\eta_{P_4} = 0.78$, $\eta_{P_5} = 0.75$, $\eta_{P_6} = 0.86$, $\eta_{P_7} = 0.84$, $\eta_{P_8} = 0.81$, $\eta_{P_9} = 0.78$, $\eta_{P_{10}} = 0.79$, $\eta_{P_{11}} = 0.83$, $\eta_{P_{12}} = 0.81$. Then the rough-fuzzy operators are $\kappa_1{=}0.28$, $\kappa_2{=}0.18$, $\kappa_3{=}0.20$, $\kappa_4{=}0.42$, $\kappa_5{=}0.30$, $\kappa_6{=}0.32$, $\kappa_7{=}0.22$, $\kappa_8{=}0.30$, $\kappa_9{=}0.42$, $\kappa_{10}{=}0.08$, $\kappa_{11}{=}0.30$, $\kappa_1{=}0.62$. The final important attributes are *P*, *H*, *Hsv-H*, *V*. The dissimilarity function structured by these four features is used to detect the following video clips, as a result of 97.0% recall and 95.4% precision are achieved. It can be seen that the rough-fuzzy operator and the dissimilarity function proposed in this paper are both effective and robust.

The results of feature selection of 5 news video are given in the Table 3. And the common rules for shot boundary detection are obtained further.

Table 3 The results of feature selection of 5 news video

Video clips	Important attributes
News 1	*P, Hsv-H, Rgb-H, V*
News 2	*P, Hsv-H, G-H*
News 3	*P, H, Hsv-H, V*
News 4	*P, V, Rgb-H*
News 5	*P, St, G-H, B*

From this table, we find that the features *P*, *Hsv-H* and *V* are more important than others, which is very fit for our experience. In News 5, the feature *B* (the blue component) is an important feature because the video is the start part of CCTV news and the anchorpersons are introducing the content abstraction of news, where the blue component occupy the chief part in the scene. So its main color is blue, which will effect on the following detection. But its weigh coefficient is small; therefore, its affect can be accepted.

The common rules for shot boundary detection can be obtained through the above analysis.

If $V = \mu_5$ or $V = \mu_4$ and $Hsv - H = \mu_5$ or $Hsv - H = \mu_4$ and $P = \mu_5$ or $P = \mu_4$,Then N;

If $V = \mu_1$ or $V = \mu_2$ and $Hsv - H = \mu_1$ or $Hsv - H = \mu_2$ and $P = \mu_1$ or $P = \mu_2$,Then C;

If $V = \mu_3$ and $Hsv - H = \mu_3$ and $P = \mu_3$, Then G.

where, N, C and G denote No Transition, Cut Transition and Gradual Transition respectively. This is only a common rule. Owing to the variation of video clips, the real situation should be dealt with by the detection procedure shown in Fig. 1 and the more accurate results can be obtained.

This paper presents an adaptive method for determining the threshold viz. $T = 1/2 \cdot \sum_{j=1}^{n} \kappa_j$. Here, c=1/2， and it is called a critical threshold.

Now, we discuss the threshold selection and taking news 3 as an example.

Firstly, the threshold is increased, that is, the value of c is increased. Then, the number of features is increased. Let c＝3/5, two features among *S*, *St* and *Rgb-H* will be added. According to our experience, the feature *St* and *Rgb-H* are selected because human eyes are not sensitive to the feature *S*. Then the recall and precision increase 0.2 and 0.3 percent respectively. The two features are added into the dissimilarity function for shot boundary detection with the computational complexity increased.

If the threshold is decreased, for example c＝2/5, features selected are decreased viz. *P*, *H*, *Hsv-H*. Then the recall and precision will decrease 2 and 3 percent respectively. Compared with the two results, the threshold $T = 1/2 \cdot \sum_{j=1}^{n} \kappa_j$ is very appropriate. That is to say, it is a critical threshold.

5 Conclusions

This paper presents a feature reduction method based on Rough-Fuzzy Set, by which the dissimilarity function for shot boundary detection is obtained. By calculating the correlation between conditional attributes, the importance of conditional attributes in the rough set can be obtained. Due to the set of feature values is ambiguous, the class precision of rough-fuzzy is given. Then, we define the new importance of conditional attributes as the rough-fuzzy operator by the product of the importance of conditional attributes in the rough set and the class precision of rough-fuzzy. According to the proportion of each feature, the top *k* features can be obtained. The dissimilarity function is generated for by weighting these important features in term of their proportion in the whole feature. Experimental results on five real news videos have shown that the proposed method not only are both similarity and effective but also can decrease data dimensions and preserve the information of original video farthest.

In the following, we will extend this work to content-based news video indexing and retrieval by incorporating other information and apply to other type of videos.

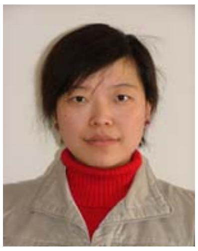
References

[1] John S. Boreczky, Lawrence A. Rowe: Comparison of video shot boundary detection techniques. In SPIE Conf. Storage & Retrieval for Image & Video Databases, 1996 (2670), pp. 170-179.

[2] Gargi, U., Kasturi R., Strayer S.H.: Performance characterization of video-shot-change detection methods. IEEE Trans. Circuits Syst. Video Technol., Vol. 10 (1) (2000), pp. 1-13.

[3] Ford R.M., C. Robson, D. Temple, M. Gerlach: Metrics for shot boundary detection in digital video sequences. Multimedia Syst., (2000), pp. 37-46.

[4] H. J. Zhang, S. W. Smoliar: Developing power tools for video indexing and retrieval. Proceedings of the SPIE: Storage and Retrieval for Image and Video databases II, San Jose, CA, Vol. 2185 (1994), pp. 140-149.
[5] Z. Pawlak: Rough Set. International Journal of Computer and Information Science. Vol. 11 (5) (1982), pp. 341-356.
[6] Z. Pawlak: Vagueness and Uncertainty: A Rough Set Perspective. ICS Research Reports 19, Warsaw Univ. of Technology, (1994).
[7] Z. Pawlak, J. Grzymala-Busse, R. Slowinski, W. Ziarko: Rough Sets. Comm. ACM. Vol. 38 (11) (1995), pp. 89-95.
[8] Guo-Yin Wang, Jun Zhao, Jiu-Jiang An, Yu Wu: Theoretical study on attribute reduction of rough set theory: comparison of algebra and information views. ICCI, (2004), pp. 148-155.
[9] D. Dubois, H. Prade: Rough fuzzy sets and fuzzy rough sets. International Journal of General Systems. Vol. 17 (1990), pp. 191-209.
[10] Sarkar, M., Yegnanarayana, B.: Rough-fuzzy membership functions, Fuzzy Systems Proceedings, IEEE World Congress on Computational Intelligence, Vol. 1 (1998), pp. 796-801.
[11] Xinbo Gao, Xiaoou Tang: Unsupervised model-free news video segmentation. IEEE Trans. on Circuits and Systems for Video Technology, Vol. 12 (9) (2002), pp. 765-776.
[12] Xinbo Gao , Xiaoou Tang: Automatic parsing of news video based on cluster analysis. Proceedings of 2000 Asia Pacific Conference on Multimedia Technology and Applications, Kaohsiung, Taiwai, China, (2000), pp. 17-19.
[13] HanBing, Gao Xin-bo, Ji Hong-bing: An efficient algorithm of gradual transition for shot boundary segmentation. SPIE on MIPPR, Vol. 5286 (2) (2003), pp. 956-961.
[14] H. J. Zhang, A Kankanhalli, S. W. Smoliar: Automatic partitioning of full motion video. Multimedia Systems, Vol. 1 (1) (1993), pp. 10-28.



Bing Han, received the B.S. in automatic control, MS degree in signal and information processing from Xidian University, Xi'an, China in 2001 and 2004, respectively. Since 2004, she has been a PH.D student in pattern recognition and intelligence systems from Xidian University, Xi'an, China. Her current interests include video retrieval and browsing, Rough Sets theory and its applications and pattern recognition.

Xinbo Gao received the B.S. in electronic engineering, the M.S. and the Ph.D. degrees both in signal and information processing from Xidian University, Xi'an, China, in 1994, 1997, and 1999, respectively. From 1997 to 1999, he was with the Computer Games Research Institute of Shizuoka University as a research fellow. From 2000 to 2001, he also worked at Multimedia Lab of the Chinese University of Hong Kong as a research Associate. He is currently a professor in the School of Electronic Engineering, Xidian University, China. His research interests include video processing, pattern recognition, and artificial intelligence.

Hongbing Ji, received the B.S. in radar engineering, the M.S. and the Ph.D. degrees both in signal and information processing from Xidian University, Xi'an, China, in 1983, 1989, and 1999, respectively. He is currently a professor in the School of Electronic Engineering, Xidian University, China. His research interests include radar target recognition, optical electronic signal processing and pattern recognition.