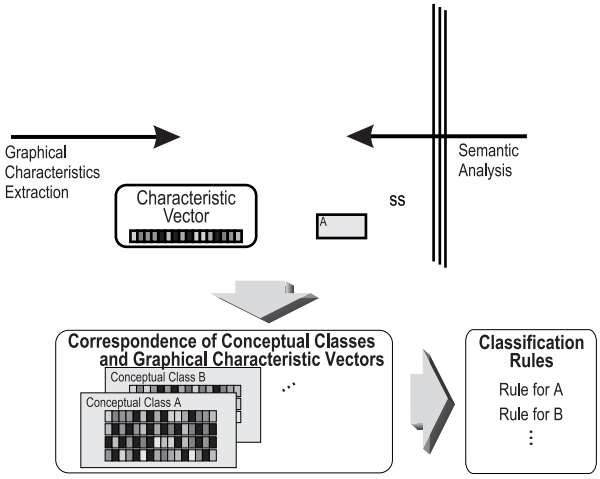

Figure 1: Learning Classification Rules from Supervisory Video Data

Figure 1 shows the outline of the learning process. As the first step, shots with captions are analyzed to learn graphical classification rules. Graphical characteristic vectors are derived from a shot, and are classified according to the semantic attributes –practically, node IDs of conceptual classes– of accompanying captions. Although the term, "rules" is used, they are actually statistic information derived from each characteristic vector in each conceptual class. After analyzing all the supervisory video data, each conceptual class should have a representative graphical characteristic vector. We are currently examining several methods such as the principal component analysis and the memory based reasoning for the acquisition of rules.

Table 1: Result of Preliminary Experiment: Correspondence between Conceptual Classes and Graphical Characteristics

Numbers of Faces	Titles of Conceptual Classes			
	Personal	Gathering	Locational	Others
None	——	'opinion, decision, investigation, acceptance'	'place name'	'counting unit'
One	'chief'; 'person's name'	——	'office, market, station'; 'place name'	——
Two	'social standing'; 'human being'	'announcement, report, rumor'	'temple, shrine, school'; 'house, inn, classroom'; 'place name'	'counting unit'; 'number'; 'principle, rule, method, custom, plan'; 'money'
Three and more	——	'speech, debate, meeting, comment, explanation'; 'promise, negotiation, approval'; 'opinion, decision, investigation, acceptance'; 'parliament'; 'gathering, presence'	'place name'	——

2.2 Preliminary Experiment

As a preliminary experiment, we have applied the learning process to 75 minutes of Japanese television news video. Only one parameter was set experimentally as an element of the graphical characteristic vector; the number of relatively large faces in the first frame of a shot. The Classified Lexical Table [5] was used to classify the captions to conceptual classes according to their semantics. Table 1 shows the correspondence of the numbers of faces and the titles of the top 30% frequent conceptual classes.

The result shows relatively good correspondence between numbers of faces and conceptual classes: (1)Conceptual classes related to human beings corresponded to one and two faces (titled 'Personal'), and (2)classes related to gatherings corresponded to two, three and more faces (titled 'Gathering'). Classes related to locations mingled in to all groups equally (titled 'Locational'), since there was no graphical characteristic parameter to classify them in this experiment. These should be classified by supplementing various characteristic parameters to the vector. The conceptual class 'opinion, decision, investigation, acceptance', related to gathering corresponded to 'no faces', since there were many 'gathering' shots taken from behind.

3 Conclusion and Future Work

We have proposed a news video classification method based on natural language information accompanying the video. The preliminary experiment showed promising results. Although, at this point, the proposed system may look somewhat similar to the Name-It system [6], further expansion of the data and graphical characteristic parameters will enhance the ability and generality.

We will also proceed with the automatic indexing phase based on the learned graphical characteristics. This should enable both advanced keyword extraction and indexing.

References

[1] "The Informedia Project"; http://www.informedia.cs.cmu.edu/.

[2] Hauptmann, A. G. and Witbrock, M. J.; "Informedia News-on-Demand: Using Speech Recognition to Create a Digital Video Library"; , pp.120-126, May 1997.

[3] Ide, I., Yamamoto, K. and Tanaka, H.; "Automatic Indexing to Video based on Shot Classification"; , Vol.2, pp.263-264, Mar 1998.

[4] Nakamura, Y. and Kanade, T.; "Semantic Analysis for Video Contents Extraction –Spotting by Association in News Video–"; , pp.393-402, Nov 1997.

[5] National Language Research Institute of Japan, The; "NLRI Natural Language Processing Data Series 5: The Classified Lexical Table (Bunrui-Goi-Hyo) [Floppy Disk Edition]"; Shuei Publishers, 1993.

[6] Satoh, S., Nakamura, Y. and Kanade, T.; "Name-It: Naming and Detecting Faces in Video by the Integration of Image and Natural Language Processing"; , pp.1488-1493, Aug 1997.