Classification of Near-Duplicate Video Segments Based on their Appearance Patterns

Ichiro Ide*†, Yuji Shamoto*‡ , Daisuke Deguchi*, Tomokazu Takahashi§ and Hiroshi Murase*
* *Graduate School of Information Science, Nagoya University*
1 Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan
Email: {ide@, yshamoto@murase.m., ddeguchi@, murase@} is.nagoya-u.ac.jp
† *National Institute of Informatics*
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
§ *Department of Economics and Information, Gifu Shotoku Gakuen University*
1-38 Naka Uzura, Gifu 500-8288, Japan
Email: ttakahashi@gifu.shotoku.ac.jp

***Abstract*—We propose a method that analyzes the structure of a large volume of general broadcast video data by the appearance patterns of near-duplicate video segments. We define six classification rules based on the appearance patterns of near-duplicate video segments according to their roles, and evaluated them over more than 1,000 hours of actual broadcast video data.**

I. Introduction

Recent advance in digital storage technology has enabled us to store a massive amount of video data in an online archive. Thanks to this trend, we have created a large-scale broadcast video archive; NII TV-RECS, which consists of the latest video data broadcast in the past several weeks on all seven terrestrial channels in the Tokyo metropolitan area. In order to make efficient use of such a large-scale archive, it is essential to analyze the structure of its contents. In this paper, we propose a video archive structuring method based on the appearance patterns of near-duplicate video segments.

Near-duplicate video segments (NDVS) are video segments that share extremely similar image features, which appear multiple times in a video stream. Recently, besides detecting advertisements [1], [2], [3], many works have focused on making use of the appearance of NDVS as a symbol that represents some sort of relationship between the structures that contain them. For example, Duygulu et al. proposed a topic tracking method that links news stories which share NDVS and a logo [4]. Yamagishi et al. implemented a browser for NDVS in news shows [5], which enables users to analyze the role and appearance patterns of NDVS in the news domain. We have proposed a cross-lingual related news event detection method that makes use of NDVS and textual correlations [6].

Since each of these works had a specific application, the detection rules were either tuned to the application, or applied simply to the specific target video data. On the contrary, the method presented in this paper is aimed so that it could analyze the structure of general broadcast video data, by handling NDVS that play various roles in a common framework.

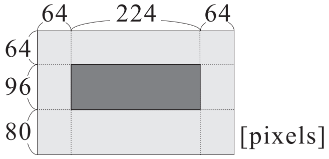


Figure 1. The region used for comparison.

II. Detection of NDVS

Various methods for NDVS detection has been proposed in recent years. Among them, we make use of the method proposed by ourselves. In this section, we first introduce the pre-processing required for the specific task discussed in this paper, and then briefly introduce the framework of our NDVS detection method.

A. Pre-Processing

Before applying the NDVS detection process to the input video frames, the following two processes are applied. Even though the NDVS detection method could naturally tolerate such effects, the pre-processing makes the detection results better.

1) *Region cropping* Since a broadcast video frequently contains captions or logos in the surrounding parts of a frame, the dark-colored region indicated in Figure 1 is cropped to compare the image features between frames.
2) *Color adjustment between channels* Since there is a considerable difference of colors between video segments obtained from different channels, the general color distribution between channels are adjusted. The

‡ Currently at Sony EMCS Corp.

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.766

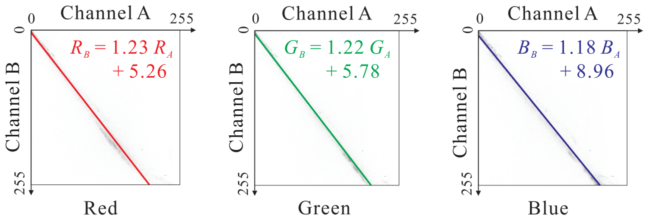


Figure 2. Example of the color correlations between channels A and B by each color component. The bold line indicates the obtained conversion function.

color conversion functions between two channels are defined as follows: In each of the R, G, B color space,

a) Create a color correlation diagram by plotting the pixel values at all pixels in the same advertisement broadcast on both channels.
b) Obtain a color conversion function by linear regression (ordinary least squares). Since extreme values of intensity are not reliable, the sections $[0, 4]$ and $[251, 255]$ were discarded.

Examples of the color correlation diagrams and conversion functions are shown in Figure 2.

B. NDVS Detection Method

We have previously proposed a method that allows efficient NDVS detection by spatio-temporal feature dimension reduction and adaptive feature space division.

The method guarantees the detection of all NDVS in a given video data set. A rough overview to the framework of the method proposed in [7] is as follows:

1) Choose the bases for feature points representation
Principal component analysis is applied to a sufficiently long video stream that could be considered to represent the nature of general broadcast video. Eigen vectors corresponding to the D largest eigen values are chosen as bases for a D-dimension feature space.
2) Detect NDVS
   a) Candidate detection in the low-dimension feature space
   Video features projected onto the D-dimension feature space are compared as low-dimension vectors, which makes each comparison fast. No pair of NDVS are overlooked at this step as long as the criterion is fixed, due to the nature of Euclidean distance.
   b) Precise detection in the original feature space
   Only the candidates detected in step 2.-(a) are compared as the original high-dimension vectors to check if they are truly near-duplicate or not.

To further reduce the computation time, we added a hierarchical feature space division process within step 2.-(a) in order to reduce the times of low-dimension feature vector comparison [8].

This is based on the idea that if a feature space is divided in two sub-feature spaces, the total times of feature points comparison (combination) is always reduced. To maximize the reduction of the times of feature points comparison at each division operation, a division boundary is set so that both of the divided sub-feature spaces contain an equal number of feature points. The division operation is applied recursively to obtain further reduction.

However, in order to guarantee that even feature points on both sides of the division border should be detected, it is necessary to set an overlap of the sub-feature spaces near their border. Since this compensation necessarily increases the number of feature points in each sub-feature space, a criterion is set so that the total times of comparison does not exceed that of the original feature space before the division operation.

Thanks to the above method, we accelerates the detection of NDVS by more than 1,000 times faster than brute-force detection on a single CPU, while maintaining the detection accuracy.

C. Experiment and Analysis

NDVS were detected from video data continuously recorded during one week for six major analog channels broadcast in the Tokyo metropolitan area, with a total length of 1,008 hours. Thanks to the NDVS detection method introduced in II-B implemented on a cluster computer with 40 CPUs, the process completed in roughly 4 days.

As a result, 3,597,942 pairs of NDVS were detected. Since the appropriateness of the detected NDVS depends highly on the application, we will not evaluate the accuracy of the results here. The NDVS formed 40,298 clusters when all the pairs were connected together (e.g. NDVS pairs $(s_a, s_b), (s_a, s_c), (s_b, s_d), \ldots$ form a cluster $s_a, s_b, s_c, s_d, \ldots)$.

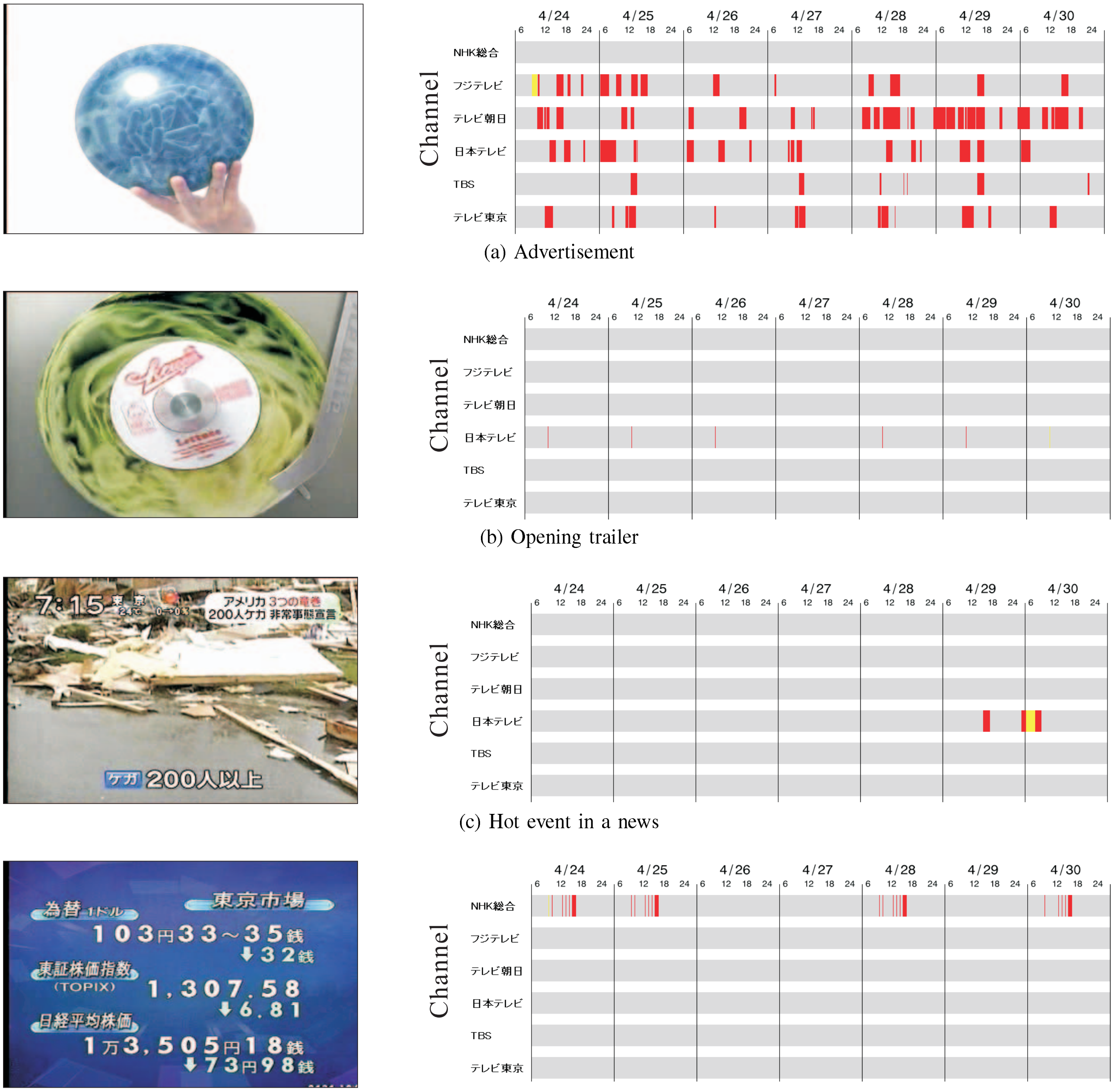
Next, we analyzed the appearance patterns and statistics of NDVS with their EPG information. Figure 3 shows examples of appearance patterns for some characteristic NDVS clusters. We can observe from (a) that an advertisement is distributed across multiple channels, while the others remain in a single channel. On the other hand, cyclic patterns are observed for an opening trailer (b) and a stocks/foreign exchange flip (d). A hot topic in a news (c) appeared concentratedly.

III. Classification of the Roles of NDVS

A. Classes and their Definition

As a result of the analysis in II-C, we assumed that NDVS could be classified into the following classes.

1) *Rebroadcast* A long video segment that contains a program broadcast multiple times. The classification could be used to remove redundant rebroadcasts of a same program.
2) *Advertisement* A short video segment that contains advertisements or announcements by the broadcaster.



(a) Advertisement

(b) Opening trailer

(c) Hot event in a news

(d) Stocks and foreign exchange

Figure 3. Example of the appearance patterns of NDVS. The distribution of NDVS differ significantly according to each role.

The classification could be used to remove redundant non-program video segments.

3) *Title* A short video segment that appears in the beginning or the ending of a program or corners. The classification could be used to detect the precise timing of a program boundary, or for the analysis of sub-program structures.

4) *Similar framing* A video segment with a very similar framing, although taken on a different timing. The classification could be used to detect anchor shots in news, or detect events in a sports match.

5) *Digest* A short video segment composed of excerpts from a longer video segment. The classification could be used to detect digests or highlights of a longer program, or summaries in the beginning of a program.

6) *News event* Video segments originally obtained from the same source. The classification could be used to extract the relation between news stories, in some cases across channels.

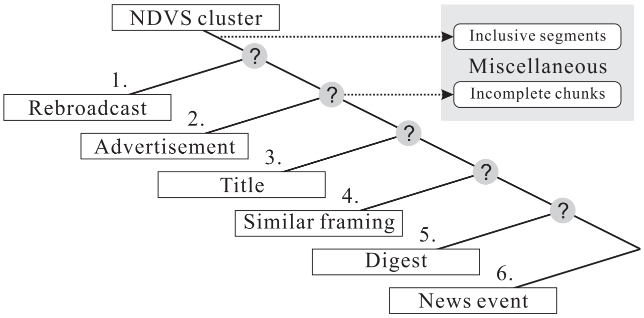


Figure 4. Classification order.

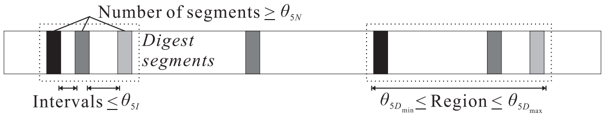

Number of segments ≥ θ_{5N}
Digest segments
Intervals ≤ θ_{5I}
$\theta_{5D_{\min}}$ ≤ Region ≤ $\theta_{5D_{\max}}$

Figure 5. Condition of ‘digest’ classification.

B. Classification Rules

The classification is performed in the order as shown in Figure 4. The order was determined considering the containment relation of the conditions. As a pre-process, pairs of NDVS that are included in pairs of longer NDVS were removed.

The classification rules for each class is as follows, which was decided based on the analysis in III-A.

1) *Rebroadcast* A NDVS cluster is labeled as a ‘rebroadcast’ when more than θ_{1P}% of the segments are longer than θ_{1L} frames.
2) *Advertisement* A NDVS cluster is labeled as an ‘advertisement’ when both of the following conditions are satisfied:
   - More than θ_{2P}% of the segments are $\theta_{2L} \pm \theta_{2E}$ frames long, based on the assumption that advertisements have a fixed length.
   - Either the segments appear during a period of more than θ_{2D} days or appear in more than θ_{2C} channels.

   As a matter of fact, some advertisements have versions with a slight difference, while some of them may be broadcast in the same sequence repeatedly. In such cases, part of an advertisement or multiple advertisements become an ‘advertisement’ segment. Such segments could be considered as noise, so we classified them as ‘miscellaneous’ and discarded them from the classification.
3) *Title* A NDVS cluster is labeled as a ‘title’ when more than θ_{3P}% of the segments appear at a certain timing of a day (fluctuations within θ_{3E} frames are allowed) for more than θ_{3D} days in a week, based on the assumption that a same program with a similar subprogram structure would be broadcast several times a week.

Table I
PARAMETERS USED IN THE EXPERIMENT.

Parameter	Value	Parameter	Value
θ_{1P}	80%	θ_{2C}	2 ch.
θ_{2P}	80%	θ_{3E}	108,000 fr.
θ_{3P}	80%	θ_{3D}	3 d.
θ_{1L}	9,000 fr.	θ_{4I}	18,000 fr.
θ_{2L}	{450, 900,	θ_{5N}	3
	1,800} fr.	θ_{5I}	150 fr.
θ_{2E}	60 fr.	$\theta_{5D_{\max}}$	21,600 fr.
θ_{2D}	1 d.	$\theta_{5D_{\min}}$	9,000 fr.

Table II
NUMBERS OF NDVS CLUSTERS CLASSIFIED TO EACH CLASS.

#	Class	Frequency	Ratio
1.	Rebroadcast	61	0.15%
2.	Advertisement	2,020	5.01%
3.	Title	533	1.32%
4.	Similar framing	14,457	35.88%
5.	Digest	3,122	7.75%
6.	News event	14,349	35.61%
–	Miscellaneous	5,756	14.28%
	Total	40,298	100.00%

4) *Similar framing* A NDVS cluster is labeled as a ‘similar framing’ when the median of the interval between all the neighboring segments is less than θ_{4I} frames, based on the assumption that NDVS with ‘similar framing’ appear concentratedly within a short period of time. The median is used in order to cover ‘similar framing’ segments that appear on different days.
5) *Digest* A NDVS cluster is labeled as a ‘digest’ when more than θ_{5N} video segments from different NDVS clusters that appear with an interval of less than θ_{5I} frames are all NDVS that appear within the range of $\theta_{5D_{\min}}$ frames and $\theta_{5D_{\max}}$ frames. Figure 5 illustrates the condition.
6) *News event* Since news events that share the same source video could have various features, all the remaining NDVS are labeled as ‘news’.

C. Experiment

The NDVS clusters detected in II-C were classified according to the rules. Parameters used in the experiment were as shown in Table I, which were empirically decided based on preliminary experiments.

Table II shows the numbers of NDVS clusters classified to each class. A portion of the NDVS clusters (61 for class 1 and 100 for classes 2–6, in total 561 clusters) were manually labeled for evaluation. Table III shows the confusion matrix of the classification results against the manually evaluation

Table III
CONFUSION MATRIX OF THE CLASSIFICATION RESULT AGAINST MANUAL CLASSIFICATION [%].

Classification result from the proposed method	Manual evaluation						
	Rebroadcast	Advertisement	Title	Similar framing	Digest	News event	Miscellaneous
Rebroadcast	**36**	0	0	0	0	0	64
Advertisement	0	**92**	2	0	0	1	5
Title	0	0	**65**	2	0	0	33
Similar framing	0	0	0	**63**	6	0	31
Digest	0	1	2	0	**35**	49	13
News event	0	1	7	5	17	**51**	19

of the results. In total, 52.5% of the NDVS clusters were correctly classified. Note that there were some clusters that were judged that they do not belong to any of the predefined classes (Indicated as 'Misc.' in the Table). These were mostly tele-shopping programs that partly seem like classes 3–5.

Next, we will discuss the results of each class. For 'rebroadcast' and 'title', most of the misclassifications were caused by tele-shopping programs. When they were excluded, the classification accuracy became 69% and 88%, respectively. For 'advertisement', we obtained good classification accuracy. If knowledge such as "a short NDVS between two advertisements is an advertisement" is introduced, the accuracy should further improve. For 'similar framing', the misclassifications were mostly due to the short replay after an advertisement break. For 'digest', the misclassifications appeared in news programs when a short video segment was repeatedly used. It was also overlooked when a digest video contained segments that do not appear in the longer segment, which occur often in movie/drama trailers. For 'news event', it was noisy since it accepted all the remaining NDVS that were not classified into other classes.

IV. CONCLUSION

In this paper, we presented a classification method of NDVS based on their appearance patterns. Results from an experiment showed that the current classification rules yield 52.5% classification accuracy. Although some classes were well classified, unexpected contents such as a tele-shopping program, caused a major misclassification for some classes. This result indicates that we need to define more classes that were not considered in this paper. In order to cope with this issue, we will consider applying learning methods for the classification to generate more complicated rules automatically. Combination with EPG information in order to analyze more precise roles (such as whether a 'title' segment is an opening or an ending trailer) is another interesting issue.

ACKNOWLEDGMENT

Parts of this work were supported by the Grants-in-Aid for Scientific Research from the Japanese Ministry of Education, Culture, Sports, Science and Technology, and by a joint research project between Nagoya University and National Institute of Informatics.

REFERENCES

[1] P. Duygulu, M.-Y. Chen, and A. Hauptmann, "Comparison and combination of two novel commercial detection methods," in *Proc. 2004 IEEE Int. Conf. on Multimedia and Expo*, June 2004, pp. 1267–1270.

[2] R. Lienhart, C. Kuhmunch, and W. Effelsberg, "On the detection and recognition of television commercials," in *Proc. IEEE Int. Conf. on Multimedia Computing and Systems 1997*, June 1997, pp. 509–516.

[3] X. Naturel and P. Gros, "Detecting repeats for video structuring," *Multimedia Tools and Applications*, vol. 38, no. 2, pp. 233–252, June 2008.

[4] P. Duygulu, J.-Y. Pan, and D. A. Forsyth, "Towards auto-documentary: Tracking the evolution of news stories," in *Proc. 12th ACM Int. Conf. on Multimedia*, Oct. 2004, pp. 820–827.

[5] F. Yamagishi, S. Satoh, and M. Sakauchi, "A news video browser using identical video segment detection," ser. Lecture Notes in Computer Science, vol. 3332. Springer-Verlag, Dec. 2004, pp. 205–212.

[6] A. Ogawa, T. Takahashi, I. Ide, and H. Murase, "Cross-lingual retrieval of identical news events using image information," ser. Lecture Notes in Computer Science, vol. 4903. Springer-Verlag, Jan. 2008, pp. 287–296.

[7] I. Ide, K. Noda, T. Takahashi, and H. Murase, "Genre-adaptive near-duplicate video segment detection," in *Proc. 2007 IEEE Int. Conf. on Multimedia and Expo*, July 2007, pp. 484–487.

[8] I. Ide, S. Suzuki, T. Takahashi, and H. Murase, "Adaptive division of feature space for rapid detection of near-duplicate video segments," in *Proc. 2009 IEEE Int. Conf. on Multimedia and Expo*, July 2009, pp. 694–697.