

Unified Approach to Detection and Identification of Commercial Films by Temporal Occurrence Pattern

Narongsak Putpuek^{2,1}, Nagul Cooharajanone^{2,1},
Chidchanok Lursinsap¹

¹Advanced Virtual and Intelligent Computing Center
Department of Mathematics, Faculty of Science,
Chulalongkorn University, Bangkok 10330, Thailand
narongsak.p@student.chula.ac.th, {nagul.c, lchidcha}@chula.ac.th

Shin'ichi Satoh²

²National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku,
Tokyo, Japan 101-8430
satoh@nii.ac.jp

Abstract—In this paper, we propose a method to detect and identify commercial films from broadcast videos by using Temporal Occurrence Pattern (TOP). Our method uses the characteristic of broadcast videos in Japan that each individual commercial film appears multiple times in broadcast stream and typically has the same duration (e.g., 15 seconds). Using this characteristic, the method can detect as well as identify individual commercial films within given video archive. Based on simple signature (global feature) for each frame image, the method first puts all frames into numbers of buckets where each bucket contains frames having the same signature, and thus they appear the same. For each bucket, TOP as a binary sequence representing the occurrence time within video archive is then generated. All buckets are then clustered using simple hierarchical clustering with similarity between TOPs allowing possible temporal offset. This clustering stage can stitch up all frames for each commercial film and identify multiple occurrence of the same commercial film at the same time. We tested our method using actual broadcast video archive and confirmed good performance in detecting and identifying commercial films.

Keywords—Temporal Occurrence Pattern(TOP); Commercial Films; Clustering; Pattern Matching;

I. INTRODUCTION

Commercial films are used for advertisement in television broadcast. Marketing research and advertising agencies are interested in monitoring commercial films to evaluate marketing plan. Since monitoring by person is very labor intensive, it is preferable to automatically detect and identify each individual commercial film.

Recently, there are mainly two research trends on commercial films detection and/or identification: shot-based [5], [1], [6], [4] and frame-based [2], [3]. Shot-based methods first extract video shots from original videos by using shot boundary detection method and then keyframes are used as representatives. On the other hand, frame-based methods use all frames in original videos as representatives. However, shot-based method depends on shot boundary detection and thus incurs lower precision, and therefore frame-based methods are preferable. [2], [3] focus on matching commercial film clips by using binary signature and simple distance

matching method. The result shows that this method has a good performance for retrieving the commercial film clips from database. However, this method is limited because it requires manually selected query videos from commercial films, and moreover it uses brute force matching which is computationally expensive. Hence, more effective methods are needed.

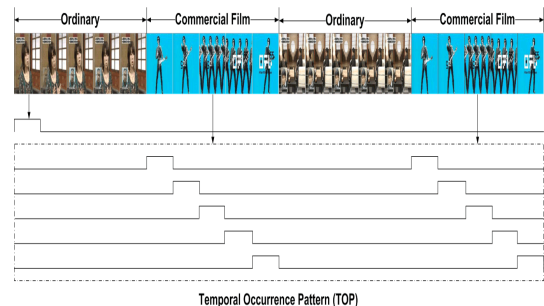


Figure 1. An example of Temporal Occurrence Pattern (TOP).

In this paper, we propose a method to detect and identify commercial films by using the characteristic of commercial films, simple signature and hierarchical clustering. Our method uses the characteristic of broadcast videos in Japan that each individual commercial film appears multiple times in broadcast stream and typically has the same duration (e.g., 15 seconds). Taking this characteristic into account, the method can detect as well as identify individual commercial films within given video archive. Different from [2], [3], our method does not require commercial film database in advance. Based on simple signature (global feature) for each frame image, the method first puts all frames into numbers of buckets where each bucket contains frames having the same signature, and thus they appear the same. For each bucket, Temporal Occurrence Pattern (TOP) as a binary sequence representing the occurrence time within video archive is then generated. Fig. 1 shows typical TOPs given frame sequence. As figure shows, TOPs corresponding to images of commercial films have couple of peaks (since they appear

multiple times), while TOPs corresponding to images of ordinary programs typically have only one peak (since they appear only once). Moreover, TOPs corresponding to an individual commercial film (but occurs multiple times in video archive) are expected to be similar each other with possible temporal offset. All buckets are clustered using simple hierarchical clustering with similarity between TOPs. This clustering stage can stitch up all frames for each commercial film and identify multiple occurrence of the same commercial film at the same time. We tested our method using actual broadcast video archive and confirmed good performance in detecting and identifying commercial films.

The rest of this paper is organized as follows: The proposed method is explained in section II. The experimental results and discussion follow in section III. Finally, in section IV the conclusion and future work are presented.

II. PROPOSED METHOD

A. Proposed Framework

The framework of this proposed method is shown schematically in Fig. 2. First, signatures of all frames of video archive are extracted. Then frames are put into buckets according to the signatures. TOP is generated for each bucket. Noise reduction is then applied to remove buckets corresponding to errors and ordinary programs. The similarity between all pairs of buckets are computed based on TOP. The hierarchical clustering algorithm is then used to generate clusters of buckets. Finally, TOP of each cluster of buckets is then generated to detect and identify commercial films.

B. Signature Extraction, Bucket Generation, and TOP Generation

We first put all frames of video archive into buckets based on signature. Each bucket is expected to contain multiple corresponding frames of an individual commercial film. For instance, if a commercial film appears multiple times in video archive, a bucket is expected to contain at least one frame for each occurrence of the commercial film, in total multiple frames in correspondence from multiple occurrences of the commercial film. Note that (possibly consecutive) frames of a commercial frame don't have to be contained in a bucket: they can be merged later at the clustering stage. Therefore signatures used to generate buckets can rather be sparse.

We found that any global feature with reasonable matching performance can be used, e.g., DCT coefficients or intensity histogram. In particular, we used very simple statistical feature-based signatures. We first generate a gray-scale image from each frame image, and normalize by mean and standard deviation of intensities. We then decompose each image into blocks (4×4 blocks for instance), and compute first- and second-order moments, namely, mean and

standard deviation, of normalized intensities for each block. Finally we concatenate all moments to generate a feature vector of the frame image. If we decompose each image into 4×4 blocks, the resultant feature will be a 32-dimensional vector. We compared the performance between this feature and DCT, and found this feature normally achieved the higher matching accuracy. We then binarize features with mean for each component to generate signature. We tested different length of signature and found 32bits signature gives reasonable performance. There could be 2^{32} buckets and thus very sparse, however, typically most buckets are empty.

Let a video sequence be $V = (f_1, f_2, \dots, f_n)$, where f_n is a frame and n is number of frames in V . Then, a feature vector is extracted from the frame f_i . A feature vector is given as:

$$S = (sig_{1,k}, sig_{2,k}, \dots, sig_{i,k}, \dots, sig_{n,k}) \quad (1)$$

where $sig_{i,k}$ is the k -th component of signature vector of frame f_i , $k = (1 \dots M)$ and M is the dimension of signature (our experiment shows that $M = 32$ gives reasonable performance). Then, feature vectors are encoded into signatures by using the following equations:

$$l_k = \begin{cases} 1 & \text{if } sig_{i,k} < m_k, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

$$d_i = \left(\sum_{k=1..M} (l_k C_k) \right) + 1 \quad (3)$$

where m_k is the mean of the k -th component, d_i is a signature and $C_k = 2^{k-1}$. The set of signatures is given as:-

$$D = (d_1, d_2, \dots, d_i, \dots, d_n). \quad (4)$$

Then we put all frames into buckets based on the value d_i . Frames with the same value d_i will be stored in the same bucket. TOP is then generated for each bucket. TOP is a binary sequence with the length n (the number of all frames in video archive) representing existence of frames in the bucket. Assume that a bucket contains p frames. Corresponding TOP is the binary sequence having p ones at temporal locations of frames.

C. Noise Reduction

From the previous step, ideally buckets corresponding to a commercial film occurring l times are assigned TOPs with l "peaks" (one peak corresponds to a consecutive ones), while TOPs of buckets of ordinary programs are expected to have only one peak. However, noise of video and/or signature hinders this ideal situation. In the next step, we compensate for these noisy situations, and then remove buckets corresponding to ordinary programs.

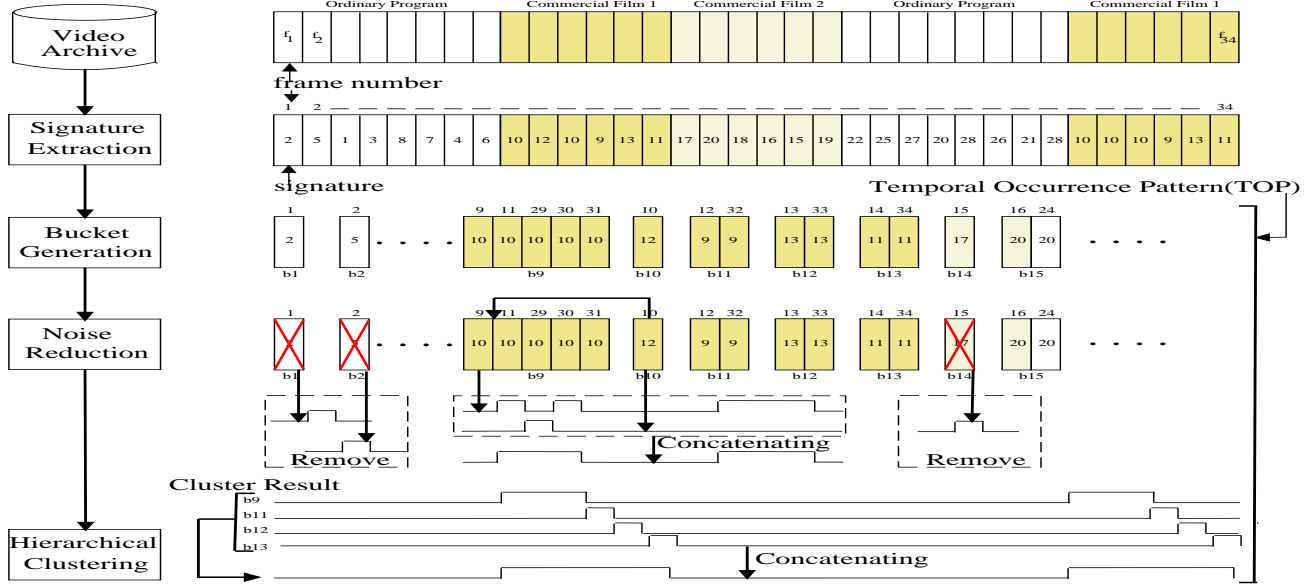


Figure 2. Proposed framework.

First, we detect the gap in TOPs (short consecutive zeros surrounded by ones) shorter than or equal to 10 frames and fill by ones. Second, we remove buckets with only one peak. Note that we also remove buckets corresponding to commercial films if they appear only once.

D. TOP Similarity Generation and Clustering

In order to classify the buckets with similar TOP into the same group, the clustering method is needed. In this work, we use standard hierarchical clustering algorithm. We use the following similarity between buckets based on TOP for clustering as follows:-

$$s(a, b) = \frac{2 \max_{\tau} \left(\sum_t u(P_a(t)) \cdot u(P_b(t + \tau)) \right)}{\sum_t u(P_a(t)) + \sum_t u(P_b(t))} \quad (5)$$

$$u(P(t)) = \begin{cases} 1 & \text{if } (P(t-1)=0) \wedge (P(t)=1), \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

where $s(a, b)$ is the similarity between two buckets a and b , P_a and P_b are TOPs of bucket a and b respectively, u is the function to select the first one of consecutive ones from TOP. τ is temporal offset maximizing the similarity between buckets.

In our experiment, τ will be maximized between -450 and 450 , which corresponds to the length of typical commercial film (15 seconds = 450 frames). The example is shown in Fig. 3.

Finally, we use a hierarchical clustering algorithm with ward's method to find clusters. Since the hierarchical clustering does not automatically determine the number of clusters, we empirically determine the threshold to stop merging clusters as follows:-

$$Th_{partition} = \beta \times d_{max} \quad (7)$$

where $Th_{partition}$ is the threshold for similarity, β is a ratio and d_{max} is the maximum distance value. In our experiment, β was varied between 0.1 to 1.0 for the optimal $Th_{partition}$, and we found $\beta = 0.7$ gave the best result.

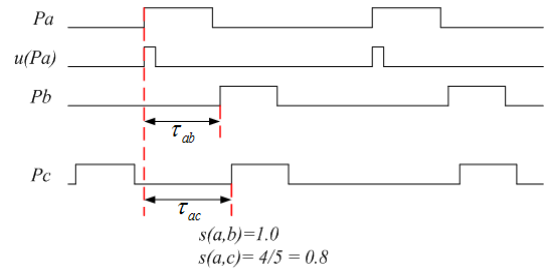


Figure 3. An example of TOP similarity determination.

III. EXPERIMENTAL RESULTS

In this section, we present the results of our newly proposed method with the actual broadcast videos in Japan. The video is in MPEG-1 format with the frame rate of 29.97 fps and the frame size of 352×240 pixels. We prepared a dataset of one hour video. We manually detected and identified commercial films in the dataset.

Table I
THE COMPARISON BETWEEN GROUND TRUTH AND GENERATED CLUSTERS

	C1	C2	C3	C4	C5	C6	C7	C8	Removed
CF1	0.00	0.00	0.00	0.00	0.00	0.00	92.33	2.89	4.78
CF2	0.00	0.00	90.89	0.00	0.00	0.00	0.00	5.78	3.33
CF3	0.00	0.00	0.00	93.82	0.00	0.00	0.00	1.38	4.80
CF4	0.00	3.11	0.00	0.00	76.44	0.00	0.00	0.74	19.70
CF5	0.00	99.44	0.00	0.00	0.00	0.00	0.00	0.00	0.56
CF6	3.11	0.00	0.00	0.00	0.00	77.22	0.00	15.78	3.89
CF7	87.00	0.00	0.00	0.00	0.00	0.00	0.00	9.78	3.22
Ordinary	0.11	0.43	0.05	0.19	0.05	0.00	0.10	57.50	41.57

We test our proposed method with one hour video which contains seven individual commercial films appeared multiple times. 10,803 non-empty buckets were generated and reduced to 1,242 buckets by noise reduction. The dendrogram of the clustering results is shown in Fig. 4. The results are summarized in Tab. I, where C1 to C8 are generated clusters and CF1 to CF7 are individual commercial films. The column C1 to C8 show the percentage of frames from commercial film CF1 to CF7 and ordinary programs that each cluster contains. The rightmost column shows the percentage of frames removed by noise reduction. From Tab. I, clusters C2, C3, C4 and C7 were successfully identified with more than 90% of frames of four individual commercial films, namely, commercial films CF5, CF2, CF3 and CF1, respectively. C5 contains only 76.44% of frames of CF4 because some buckets of CF4 contain only one peak and thus removed by noise reduction. C6 contains 77.22% of frames of CF6, because 15.78% of CF6 are contained in C8.

The largest C8 contains 57.50% of frames of ordinary programs, because they contain very similar video segments (logos, jingles, etc.). Therefore, TOPs of the corresponding buckets have couple of peaks, and only less than half of frames were removed by our method.

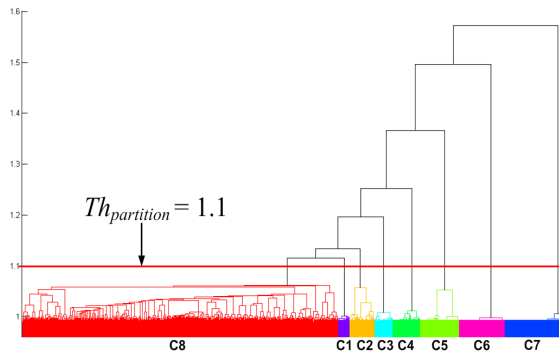


Figure 4. A dendrogram of one hour video.

IV. CONCLUSION AND FUTURE WORK

We propose a new method to detect and identify the commercial films in a large video archive, based on Temporal Occurrence Pattern (TOP). The hierarchical clustering algorithm is used to classify the TOPs into groups. This new method was found to work well with commercial films that have exactly repeated contents. We believe that our method scales well because the more videos we process, the more frequently we observe repeated commercial films, and thus the more stable TOP we can obtain. We will confirm this fact in the near future. We also will address adaptive thresholding for clustering and focus on the commercial films which have different duration.

ACKNOWLEDGMENT

The authors acknowledge financial support from Thailand Research Fund (TRF) and Commission on Higher Education (CHE).

REFERENCES

- [1] P. Duygulu, M. Y. Chen and A. Hauptmann, "Comparison and Combination of Two Novel Commercial Detection Methods", Proceedings of the 2004 International Conference on Multimedia and Expro (ICME'04), Taipei, Taiwan, June 2004, pp. 1267-1270.
- [2] Y. Li, J. S. Jin and X. Zhou, "Matching Commercial Clips from TV Streams Using a Unique, Robust and Compact Signature", Proceedings of the Digital Image Computing: Techniques and Applications, 2005(DICTA'05), Dec. 2005, pp. 266-272.
- [3] Y. Li, J. S. Jin and X. Zhou, "Video Matching Using Binary Signature", Proceedings of 2005 International Symposium on Intelligent Signal Processing and Communication Systems(ISPACS 2005), Hong Kong, Dec. 2005, pp. 317-320.
- [4] J. M. Gauch and A. Shivadas, "Finding and identifying unknown commercials using repeated video sequence detection", Computer Vision and Image Understanding, vol. 103, July 2006, pp. 80-88.
- [5] Y. Zheng, L. Duan, Q. Tian and J. Jin, "TV Commercial Classification by using Multi-Modal Textual Information", Proceedings of the 2006 IEEE International Conference on Multimedia and Expo (ICME'06), Toronto, ON, Canada, July 2006, pp. 497-500, doi:10.1109/ICME.2006.262434.
- [6] J. Wang, L. Duan, Q. Liu, H. Lu and J. Jin, "Robust Commercial Retrieval in Video Streams", In Proc. IEEE Conf. Multimedia and Expo (ICME'07), Beijing, China, July 2007, pp. 260-263.