

Enhanced Metadata Creation and Utilization for Personalized IPTV Service

Hyojin Park¹, Kireem Han², Jinhong Yang^{3(✉)}, and Jun Kyun Choi²

¹ Department of Information and Communications Engineering, KAIST, Daejeon, South Korea
gaiaphj@kaist.ac.kr

² School of Electrical Engineering, KAIST, Daejeon, South Korea
hanrob@kaist.ac.kr, jkchoi59@kaist.edu

³ HECAS, Seoul, Republic of Korea
jinhong.yang@hecas.co.kr

Abstract. Metadata is the clue to connecting videos and individual users. For a successful personalization, collecting more detailed and concrete metadata is important. This paper proposes a method to create and utilize metadata more precisely as a result of research on how to do the personalization better. For that, we propose a way to consume and utilizing a video in a series of segmented images so that can get the user's taste on specific sections or points of the video. Proposed method is implemented on the web site and the effectiveness of proposed method is revealed by comparing the site stay time and number of video usage per session between new visitors and returning visitors.

1 Introduction

The role of IPTV service providers can be defined as connecting videos to users, and the success of connection can be judged as the number of video consumption. To promote the IPTV users' video consumption, personalization is appealing solution but the problem is how to do. Because providing videos tailored to the user's tastes can surely increase the video consumption, but improper personalization will make not just decreasing the video consumption but make users to leave the service.

To successfully display the videos custom to individual users, it is very important to catch the user's preference and taste on video contents both in general and for the moment correctly. Metadata, which contains additional, explaining, and helpful information around the videos, service, user, and the user's community activity, gives all the information required to analyze the user's appetite and generate the personalized program guide.

In this matter, this paper proposes an enhance metadata creation and utilization method, i.e., serving a video into an image carousel form. Extracting multiple key frame images with full script of the video and serving in image carousel form increases the potential and efficiency of the video linkage and usage especially in social network services. As shown in Fig. 1, by allowing users' to consume segmented image based video content and tracing their activity in image unit while saving them as specific

metadata, IPTV service provider can collect and utilize more detailed, concrete, and precise metadata.

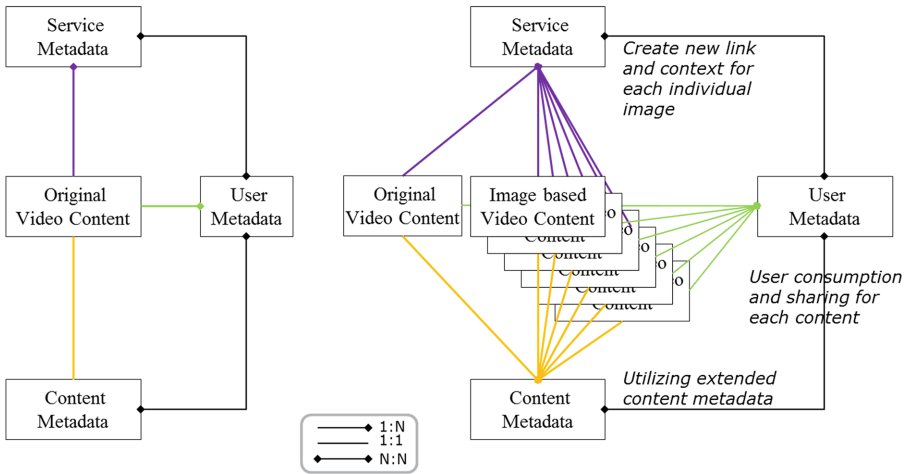


Fig. 1. Concept of the proposed method

Rest of this paper is composed as follows. Section 2 explains the proposed image carousel based video consuming method and the generation process. In Sect. 3, the list of newly created and extracted metadata elements by applying proposed method and their usage is introduced. Section 4 shows the implementation and service test results of the proposed method and concludes in Sect. 5.

2 Proposed Method for Generating Image Carousel

In this paper, we have studied how to divide video into several segments and package into a convenient form in order to understand user’s taste on video. In addition to consuming video in each video unit, we have invented a way to play, share, and utilize specific segments of video to create and reflect the new and more specific metadata while keeping the original video’s content.

Proposed method extracts keyframes representing each segment and to combine them into an image carousel. To ensure minimum loss in the original video content while saving playtime and data traffic, proposed method keeps the entire script automatically generated by audio mining which are actively researched owing to the development of the recent advances on audio recognition [1–3, 10, 12]. By doing so, users can grasp the content of the video without playing but reading and can share or play from any keyframe image on the image carousel. In addition, proposed method reduces the long play time and the demand for high data traffic, which are the disadvantage of existing video services.

Technically, serving video in image carousel can be regarded as one of the method in the video abstraction or video summary [4, 5, 11]. However, in the sense that proposed image carousel generation does not shorten the voice of the video but keeps the entire

story written in script, it is different from the existing approaches [6–8]. Also, generating image carousel can be regarded as one of the methods in the keyframe extraction [9]. However, proposed scheme extracts keyframes from the meaning unit which is distinguished from the script, not from the entire video or by image processing.

The process to generate the proposed image-carousel from a video is done by the following steps as shown in Fig. 2.

- Script generation: Generating script of the video by audio mining
- Script parsing: Time based script parsing and keyword extraction
- Video segmentation: Segmenting the video based on the maximum length of one script unit
- Keyframe extraction: Select a keyframe of each segment using key-word, expression, and extracting information
- Packaging: Overlay the script on the keyframe and packaging as image carousel format

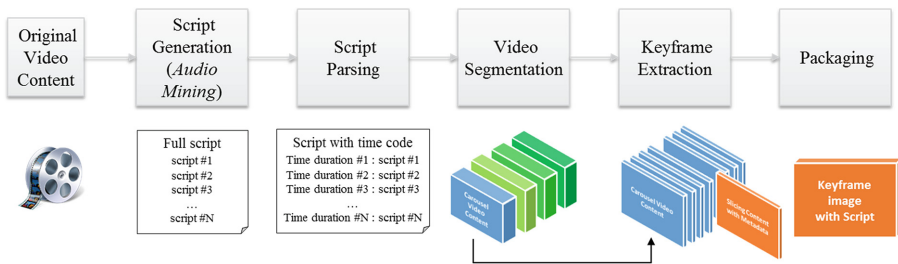


Fig. 2. Proposed image carousel generation process

3 Enhanced Metadata Creation and Utilization

When videos are served and consumed by proposed method, then new types of metadata elements are created from the video segmentation and extracted while the users are enjoying the image based video service. The list of new metadata elements are as follows:

- Newly created metadata elements from video segmentation
 - Script on video
 - Each script on time period
 - Keywords
 - Keywords on entire video
 - Number of times the keyword appeared in each segmented script
- Newly extracted metadata element from proposed service
 - Impressions by image
 - Inter-image exposure time
 - Image playback conversion frequency per image (conversion rate)
 - Number of shares per image

- Number of plays after conversion (conversion rate)
- Ad impressions by image
- Image-specific ad impression keywords
- Ad clicks by image (conversion rate)

Table 1 shows a comparison between the metadata provided by legacy service, i.e., YouTube and the metadata generated by the method of this paper. Also, the table is listing which new services are possible using the new metadata, not limited to but as typical examples including enhanced IPTV service personalization.

Table 1. Usage of newly created metadata

YouTube (Legacy service)	New metadata from proposed method	Advances and new service opportunities
Watch time (Estimated total audience watch time for videos)	<ul style="list-style-type: none"> ◆ Inter-image exposure time ◆ Keywords on entire video ◆ Number of times the keyword appeared in each segmented script 	<ul style="list-style-type: none"> ◆ Individual Ad service by enhanced advertisement matching ◆ Effective Ad insertion in the video
Views	<ul style="list-style-type: none"> ◆ Impressions by image ◆ Number of shares per image 	<ul style="list-style-type: none"> ◆ Effective Ad placement in service page
Average watch duration	<ul style="list-style-type: none"> ◆ Impressions by image ◆ Inter-image exposure time ◆ Each script on time period ◆ Number of times the keyword appeared in each segmented script 	<ul style="list-style-type: none"> ◆ Upgraded content recommendation ◆ Sophisticated service personalization
Average view rate	<ul style="list-style-type: none"> ◆ Impressions by image ◆ Image playback conversion frequency per image (conversion rate) ◆ Number of shares per image ◆ Number of plays after conversion (conversion rate) 	

When a video is played, the reason that the user likes the video can be different for actors, writers, directors, backgrounds, and so on. This difference in video consumption is hard to understand by tracking the consumption history on video units, but it can be better distinguished from the generation and consumption histories of the proposed specified metadata elements. The precisely seized user’s preference can be used to recommend videos and create a personalized IPTV service program guide that is more tailored to each user.

4 Implementation and Test Results

Figure 3 is the screenshot of the implemented web site to test the effectiveness of the proposed method. As shown in the figure, on each video has two overlaid buttons. When

a user chooses the left-hand side button, then the video is played, and when the right-hand side button is selected, then the video is served as the proposed image carousel. By running implemented service for one month, we could collect site stay time and content consumption information via google analytics. The drop-off rate of moving between contents is 21.55% based on the total users, 6.9 contents used during the visit, and the stay time is 5:57 min. When we divide this by the number of contents usage compared to the visit time, it can be understood that the use time for each piece of contents is about 51 s.

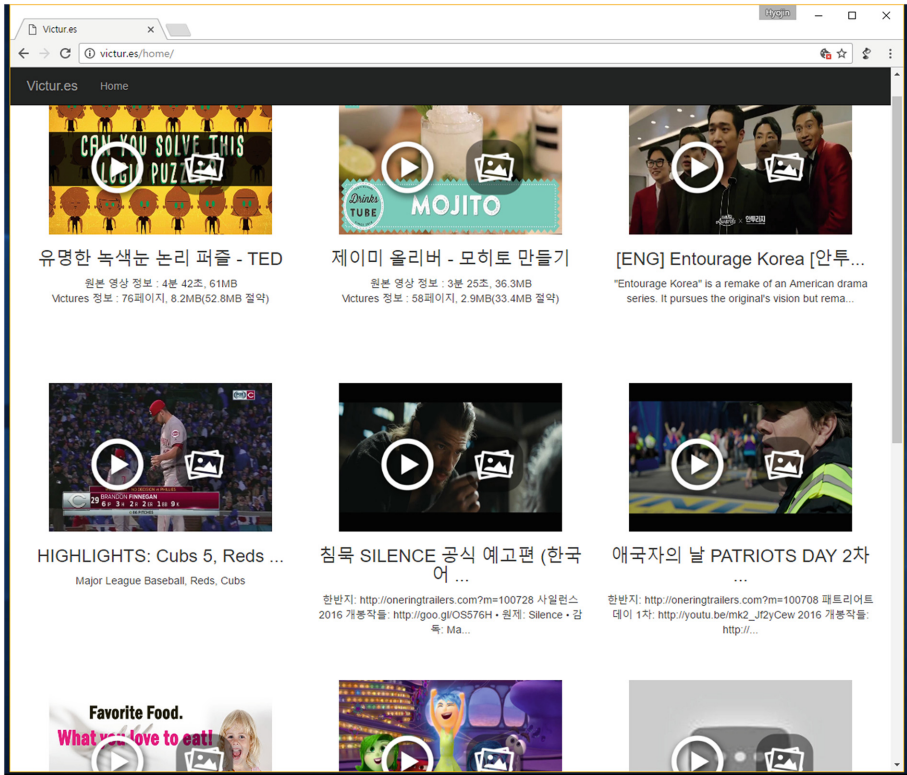


Fig. 3. Screenshot of the web site for running proposed method

The effectiveness of proposed method for making users to consume the more videos is revealed by comparing the site stay time and number of contents usage between new visitors and returning visitors. As shown in Table 2, from the aspect of drop-off rate, it can be seen that the drop-off rate of returning visitors is lower than that of new visitors. This implies that more content is consumed on the site by returning visitors, and that is confirmed by the fact that the number of content usage per session of returning users is more than 4 more than that of new visitors. Also, it was confirmed that the site stay time was very high as 9:01 min for returning visitors compared to the 3:45 min for new visitors. This data can directly or indirectly explain the satisfaction of return visitors to

the service and that shows the use of newly generated concrete metadata facilitates video consumption by video service visitors.

Table 2. Statistics on the implemented service with proposed method

	Session	Drop-off rate	Number of video usage per session	Average session time
All visitors	100%	21.55%	6.91	5:57 min
New visitors	58.16%	26.26%	5.2	3:45 min
Returning visitors	41.84%	15.00%	9.28	9:01 min

5 Conclusion

To link proper contents with different users, grasping the characteristics of contents and the interests of each user is very important. Proposed method for enhanced metadata creation and utilization for IPTV service personalization offers a new way of consuming video on IPTV. By providing multiple key frame images in carousel form with full script of the video, proposed method enables users can ‘read’ the content rather than ‘watch’. Segmenting the video and re-generate it in image carousel form allowed IPTV service providers to get the metadata on user’s preference and activity in specific sections or points and lead to the longer site stay time and the more video consumption as shown in the field test results.

Acknowledgement. This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No: R-20160906-004163, Developing Bigdata Autotagging and Tag-based DaaS System).

References

1. Yi, H., Rajan, D., Chia, L.-T.: Semantic video indexing and summarization using subtitles. In: Aizawa, K., Nakamura, Y., Satoh, S. (eds.) PCM 2004. LNCS, vol. 3331, pp. 634–641. Springer, Heidelberg (2004). doi:[10.1007/978-3-540-30541-5_78](https://doi.org/10.1007/978-3-540-30541-5_78)
2. Cour, T., Jordan, C., Miltsakaki, E., Taskar, B.: Movie/script: alignment and parsing of video and text transcription. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008. LNCS, vol. 5305, pp. 158–171. Springer, Heidelberg (2008). doi:[10.1007/978-3-540-88693-8_12](https://doi.org/10.1007/978-3-540-88693-8_12)
3. Košút, M., Šimko, M.: Improving keyword extraction from movie subtitles by utilizing temporal properties. In: Freivalds, R.M., Engels, G., Catania, B. (eds.) SOFSEM 2016. LNCS, vol. 9587, pp. 544–555. Springer, Heidelberg (2016). doi:[10.1007/978-3-662-49192-8_44](https://doi.org/10.1007/978-3-662-49192-8_44)
4. Truong, B.T., Venkatesh, S.: Video abstraction: a systematic review and classification. ACM Trans. Multimed. Comput. Commun. Appl. **3**(1), 1–37 (2007)
5. Hu, W., Xie, N., Li, L., Zeng, X., Maybank, S.: A survey on visual content-based video indexing and retrieval. IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.) **41**(6), 797–819 (2011)
6. Kim, W., Kim, C.: A new approach for overlay text detection and extraction from complex video scene. IEEE Trans. Image Process. **18**(2), 401–411 (2009)

7. Wactlar, H.D., Kanade, T., Smith, M.A., Stevens, S.M.: Intelligent access to digital video: informedia project. *Computer* **29**(5), 46–52 (1996)
8. Jung, K., Kim, K.I., Jain, A.K.: Text information extraction in images and video: a survey. *Pattern Recogn.* **37**, 977–997 (2006)
9. Gianluigi, C., Raimondo, S.: An innovative algorithm for key frame extraction in video summarization. *J. Real-Time Image Process.* **1**(1), 69–88 (2006)
10. Liu, Y.-J., Ma, C.-X., Fu, Q., Fu, X., Qin, S.-F., Xie, L.: A sketch-based approach for interactive organization of video clips. *ACM Trans. Multimed. Comput. Commun. Appl.* **11**(1), 1–21 (2014)
11. Mei, T., Tang, L.-X., Tang, J., Hua, X.-S.: Near-lossless semantic video summarization and its applications to video analysis. *ACM Trans. Multimed. Comput. Commun. Appl.* **9**(3), 16–39 (2013)
12. Money, A.G., Agius, H.: ELVIS: entertainment-led video summaries. *ACM Trans. Multimed. Comput. Commun. Appl.* **6**(3), 17–47 (2010)