July 2021

Identifying Suitable Slots for In-stream Video Advertisements

Victor Carbune

Matthew Sharifi

Identifying Suitable Slots for In-stream Video Advertisements

ABSTRACT

Video streaming services insert in-stream video ads at various points when providing users video content. However, such ad insertion can sometimes affect the viewing experience of the user if the content is paused at inappropriate moments or by unnecessarily prolonging the overall viewing session. This disclosure describes techniques to analyze video content using machine learning models or other suitable techniques to identify suitable portions that can be replaced by inserted ads or can be displayed together with an ad. By replacing sections of the video with little or no content, the techniques provide a more integrated, resource-efficient, and generally improved ad experience for users. For example, the user experience for video content such as cooking videos, exercise videos, or other videos with natural pauses can be improved by such analysis and ad insertion.

KEYWORDS

- Video advertisement
- In-stream advertising
- Ad insertion
- Advertisement serving
- Ad slot
- Podcast
- Video understanding
- Content understanding
- Silent period

BACKGROUND

On-demand video services sometimes insert advertisements within a video stream delivered to a user. In-stream advertising formats include skippable and non-skippable ads. In-stream video ads may be inserted at any point during a piece of content. Thus, an advertisement can sometimes be disruptive to the viewing experience of the user (e.g., when video content is paused/interrupted at inappropriate moments) or otherwise affect the quality of viewing experience, e.g., by prolonging the overall viewing session.

DESCRIPTION

This disclosure describes techniques to insert ads into a video stream by replacing a suitable portion of the underlying content. Online video content includes a wide variety of videos, many of which exhibit a natural downtime. For example, an exercise video may include a pause of 10 seconds, 30 seconds, etc. for the viewer to rest.

Sections of a video with little or no content are automatically identified and replaced with ads, thereby enabling a more integrated, resource-efficient, and generally improved ad experience for users. By replacing relatively content-free sections of the video with advertisements in an integrated and natural manner, user engagement can be maintained without needlessly prolonging the video. The techniques are generally applicable and particularly useful in videos that include natural downtime where there is little or no content for the user to consume, e.g., exercise videos, cooking/recipe videos, how-to videos, etc.

The described techniques are implemented with specific permission from the content provider to analyze their video content and replace portions with advertising. Certain videos can be excluded from such analysis and instead, ads can be placed at regular intervals, at user-

selected time points, or at the start/end of the video. Further, certain content may be displayed without advertising.

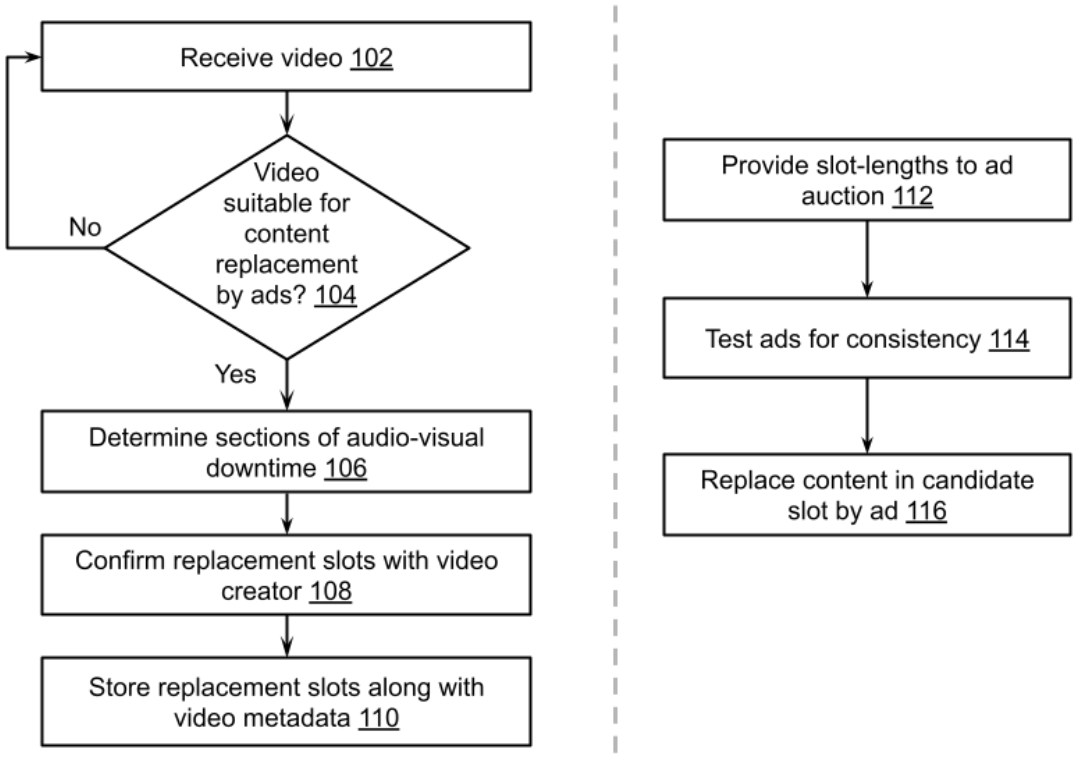

Fig. 1: (a) Determining sections of a video suitable for replacement by ads; (b) Serving ads during candidate slots

Fig. 1 illustrates automatic replacement of portions of video content by ads, per the techniques of this disclosure. Fig. 1(a) illustrates determination of appropriate sections of a video for replacement by ads, a procedure that can be done when the video is received/uploaded to the video service. A video is received by a content-hosting platform from content creators (102). The uploaded video is analyzed to determine its suitability (eligibility) for content replacement by ads (104). Some examples of videos eligible for content replacement include the following: monetization-enabled videos; videos that belong to specific content categories or verticals, e.g.,

workout videos, how-to videos, etc.; videos where the content creator has specified a preference towards content replacement by ads; etc.

The video content is analyzed using machine learning models (and/or other suitable techniques) to detect periods of audio-visual downtime (106). Some examples of machine learning (ML) models suitable for this purpose include deep neural networks (e.g., convolutional neural networks) trained on video and/or audio frames (either individually or in a sequence); residual neural networks; etc. The machine learning models can be trained, for example, in a supervised manner on segments of labeled video content. A frame or segment can be assigned a content richness score. For example, a segment with engaging speech or rich visual content can be assigned a high score. On the other hand, a blank frame (e.g., with a countdown timer) with no background audio can be assigned a low richness score. A sequence of frames, e.g., a static or relatively static sequence, may be identified and assigned a low richness score. The output from various ML models can be aggregated to detect runs of video downtime that are of a suitable length for ad insertion. Smoothing can be applied to allow for short gaps or noise in individual model outputs. A run that is above a length threshold can be selected as a replacement slot.

Possibly as a deferred step, the identified candidate replacement slots can be shown to the content creator to confirm or reject as replacement segments (108). Alternatively, they can be initially auto-confirmed, with the option given to the creator to reject later. The identified replacement slots are stored with the video metadata (110).

Fig. 1(b) illustrates the serving of ads during candidate slots. Each candidate slot is considered for replacement by an in-stream ad at the time of serving the video to a viewing user. Candidate replacement slots and their lengths are provided to ad auctions to identify ads of a suitable length, e.g., those that span the majority of the content-free segment (112).

Advantageously, the set of ads used for the video can be chosen to be consistent with each other (114), e.g., the same ad is not repeated, nor are inconsistent ads shown within the same piece of content. The selected ads can be inserted into the content, replacing the original video content in the playback stream (116). Replacement of video content by ad can be done by providing instructions to the video player, or for some clients by performing content substitution on the content itself. The user now consumes a composite video comprising original video interspersed with ads, where sections of the original video with low audio-visual activity are replaced by ads.

Alternatively, a hybrid ad can be created where the original video content (without audio) is shown alongside the ad. For example, this can be shown in a picture-in-picture format, e.g., as a small frame in the corner of the video, thus ensuring that the viewer does not miss anything important. The hybrid mode can support both skippable and non-skippable ads. A skippable ad enables the user to exit the ad and view the original content in full frame mode.

The described techniques can be utilized by video hosting and sharing services, audio services, podcasts, etc. For music or podcasts, segments with low/uninteresting audio content can be determined using an ML model and can be used as ad-insertion (or replacement) points. Similarly, developments in the storyline can be determined using an ML model and in turn used to determine ad insertion (or replacement) points.

Although the described techniques detect regions well-suited for replacement in an automatic manner, content hosting and serving platforms can also provide mechanisms for content creators to themselves provide this information, e.g., via metadata in the upload flow; via explicit markers in the video content; etc.

In this manner, the techniques of this disclosure optimize the overall length of video and audio content while still serving advertisements. Serving content in this manner can save time,

bandwidth, and other computing resources. The techniques enable ads to be inserted in an acceptable (nonintrusive) manner, e.g., during pauses in videos in specific verticals such as exercise videos, how-to videos, etc.

CONCLUSION

This disclosure describes techniques to analyze video content using machine learning models or other suitable techniques to identify suitable portions that can be replaced by inserted ads or can be displayed together with an ad. By replacing sections of the video with little or no content, the techniques provide a more integrated, resource-efficient, and generally improved ad experience for users. For example, the user experience for video content such as cooking videos, exercise videos, or other videos with natural pauses can be improved by such analysis and ad insertion.

REFERENCES

1. Dev, Soumyabrata, Hossein Javidnia, Murhaf Hossari, Matthew Nicholson, Killian McCabe, Atul Nautiyal, Clare Conran, Jian Tang, Wei Xu, and François Pitié. "Identifying Candidate Spaces for Advert Implantation." In *2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT)*, pp. 503-507. IEEE, 2019.
2. “Yospace Delivers Big Audiences For Dynamic Advertisement Replacement” available online at https://www.yospace.com/2015/07/27/yospace-delivers-big-audiences-for-dynamic-advertisement-replacement/, accessed Jun 12, 2021.