December 2020

Seamlessly Embedding Relevant Related Video Content in a Video Playback Stream

Alexandru Damian

Victor Carbune

Seamlessly Embedding Relevant Related Video Content in a Video Playback Stream

ABSTRACT

Online video content often includes links to additional video content related to the main video. The links to the related videos can be included in one or more places within the user interface , such as inline within the main video, within the description of the main video, etc. Users must interact explicitly with the links in order to view the linked related videos. This disclosure describes techniques to insert relevant related video content with permission inline within a main video to generate a video stream that seamlessly shows the combined content.

KEYWORDS

- Instructional videos
- How-to videos
- Video stream
- Inline video
- Video understanding
- Image understanding
- Video summarization
- Content summarization
- Video segmentation
- Video cropping

BACKGROUND

People commonly view instructional video content such as tutorials, course lectures, product reviews and deep-dives, etc. that is hosted by online video services, using devices such as computers, mobile devices, televisions, etc. Many video services as well as content creators

often provide links to additional video content related to the main video, such as videos that include in-depth explanations related to content within specific sections of the main video. Such links to the related videos can be included in one or more places within the user interface (UI), such as inline within the main video, within a description section of the main video, as a sidebar, etc.

The manner in which users interact with the linked videos can vary widely. User interactions with the linked video content can include practices such as: watching every linked video, watching only the few videos that are related to specific parts within the main video that are of interest to the user, watching only the main video without consuming any of the linked related video content, etc.

DESCRIPTION

This disclosure describes techniques to automatically embed relevant related video content (or portions/summaries of such content) inline within a main video. With user permission, a trained machine learning model is applied to infer the user's preferred video content style for video playback. If the user is determined to prefer inline added content in the main video, the related video content is played inline within the main video by temporarily pausing the playback of the main video and inserting the related video (or a version thereof) within the video stream. If needed, the inserted related video content is trimmed or summarized as appropriate to match the context of the main video content and ensure a seamless user experience (UX).

First, the user's preferred mode of consuming related video content is determined. The user can specify the preference directly by choosing an appropriate option among the available choices, such as "embed the full related video inline," "embed a summary of the related video

inline," "ignore related videos," etc. Alternatively, if the user permits, a suitably trained machine learning model can be used to infer the user's preference for the video content the user is watching. The model can be provided with input information regarding the video being watched, e.g., the type of video content (e.g., product assembly tutorial, product review of a smartphone), metadata of the video content (e.g., duration, content source), whether users of the video choose to watch related content, etc. Further, with user permission, the model can be provided relevant contextual and user information, such as the user's preference settings, video viewing history, search history, content purchases, schedule, chat message, etc. For instance, calendar and/or messaging information, obtained with user permission, can be indicative of the amount of time the user has available to watch the video content. For example, video and user information can be processed to obtain respective embeddings that are used together as input.

If it is determined that the user prefers a combination of the main video and related content, the content of the main video is combined with the related video content for playback. For example, the combined content can be generated via a suitable trained machine learning model that employs smart cropping and content summarization to output a video stream containing content from the main video and one or more related videos. Video cropping is used to remove content that is duplicated across videos, such as introduction, ending credits, repeated information, etc. Cropping can additionally remove content that is irrelevant to the topic of the main video, such as recommendations. If appropriate, the cropped content can be further shortened by trimming the content or generating a content summary. The cropped (or otherwise edited) related video content is inserted inline within the main video at an appropriate place to provide a seamless viewing experience for the video stream consisting of a combination of the main video with one or more related videos.

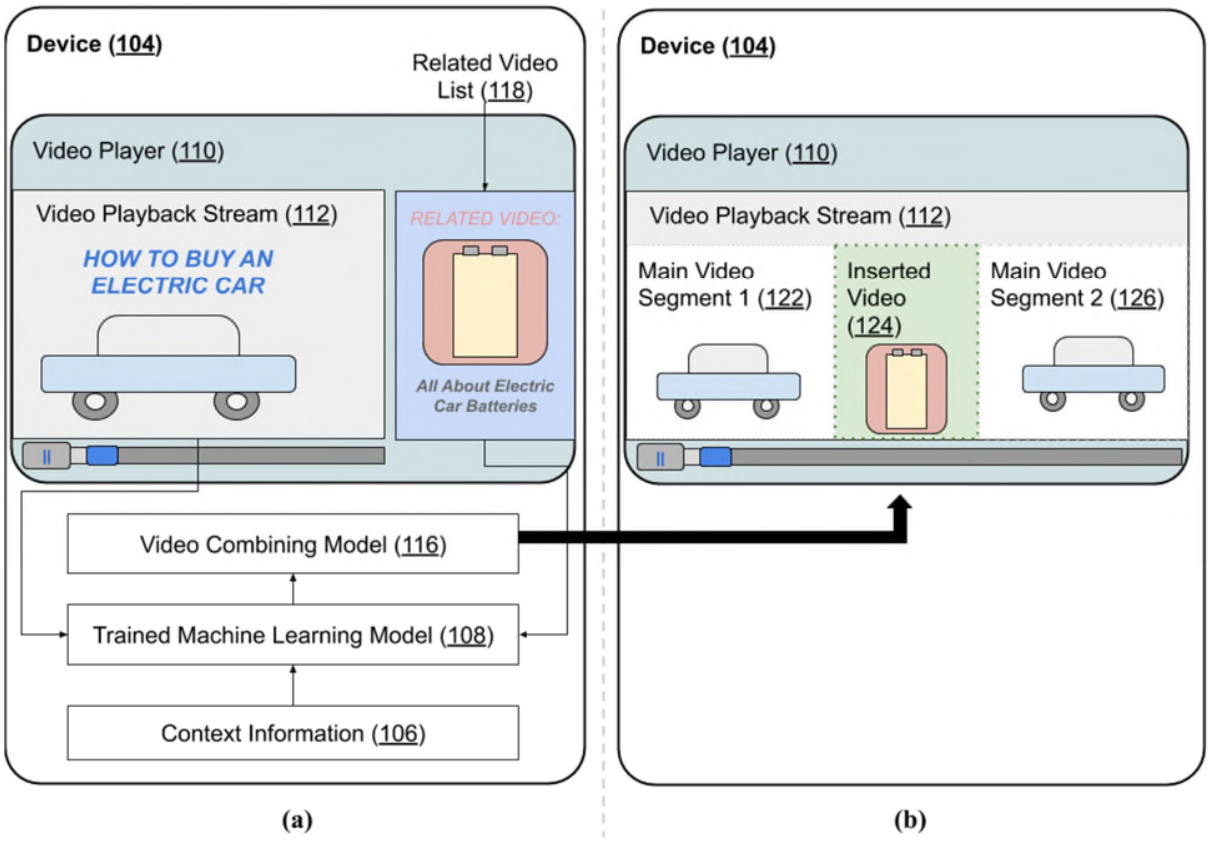

Fig. 1: Inserting a related video inline within the video content being viewed

Fig. 1 shows an example of operational implementation of the techniques described in this disclosure. A video player (110) on a user device (104) is used for playing a video stream (112). As Fig. 1(a) shows, a list of videos (118) related to the currently playing video is shown alongside the video playback. In this case, the user is watching a video on buying an electric car, and the related video pertains to batteries for the car.

Based on the video being played and the related video combined with relevant user-permitted context information (106), a trained machine learning model (108) is used to determine whether the user is likely to prefer that the related video content be embedded within the video playback stream. If such user preference is detected, a video combining model (116) which can be a separate trained machine learning model is used to insert the related video content at an

appropriate place within the playback stream. As shown in Fig. 1(b), the resulting playback stream consists of video content from the related video inserted (124) by splitting the main video into two segments, the first shown prior to the inserted video (122) and the second shown following the inserted video (126). While Fig. 1 shows insertion of a single content item, any number of related video content items can be inserted inline at appropriate positions within the main content.

The position at which the related content is inserted can be based on a semantic understanding of the main video as well as the related video content. For example, if the main content is a quick overview of the features of a new smartphone that has sections that cover the processor, memory, display, operating system, camera, battery, etc. of the smartphone, related videos that offer in-depth review of these components can be inserted inline at the appropriate section boundary. In another example, if the main content is a comparison of two different smartphones, in-depth reviews of individual phones can be inserted when each device is discussed during the comparison

The techniques described in this disclosure can support a broad variety of video viewing scenarios in which a user may desire watching additional video content related to the currently playing video. For instance, a video about buying an electric car can be associated with various videos pertaining to related aspects such as older models of a particular make, in-dash information and controls, battery information, comparison among electric vehicle types and manufacturers, etc. A user who has already decided on the car model to purchase may be interested in the battery information video; another who is undecided on the model may be more interested in the video comparing various models; and another user who is simply curious to know about electric cars in general may not want to watch any further videos on the topic. As

another example, some viewers of a video tutorial on sharpening chef knives may be interested in related video content on sharpening other types of knives while other viewers may be interested in related videos that provide additional information about sharpening chef knives, such as multiple types of sharpening angles and blades. With user permission, the techniques described in this disclosure can select and embed only those related videos that match the inferred or specified interests of each viewer.

Detecting content that is duplicated between two videos can be achieved via exact image based matching and/or approximate matching using machine learning algorithms to predict similarity given two video frame sequences. The threshold values for inferences used in the operation of the various machine learning models used in the implementation of the described techniques can be set by the developers and/or specified by the users and/or determined dynamically at runtime.

The described techniques can be implemented to support any type of video content and can be incorporated with any app, service, or platform that serves video content and related recommended videos. The techniques can be especially beneficial for consuming instructional video content such that relevant additional explanatory videos are automatically selected and embedded within the stream with user permission, thus creating a personalized learning experience. With user permission, the techniques enable a streamlined experience for consumption of video content, thus reducing interaction burden for the user and enhancing the overall user experience (UX) of watching video content.

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable collection of user information (e.g., information about a user's video viewing history

and/or preferences, a user's calendar, purchases, browsing and/or search history, a user's preferences, or a user's current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

Online video content often includes links to additional video content related to the main video. The links to the related videos can be included in one or more places within the user interface , such as inline within the main video, within the description of the main video, etc. Users must interact explicitly with the links in order to view the linked related videos. This disclosure describes techniques to insert relevant related video content with permission inline within a main video to generate a video stream that seamlessly shows the combined content. Trained machine learning models are used to determine user's preferences regarding inlining related content and to generate a combined stream based on the user's preferences. The related content can be cropped, summarized, or otherwise edited prior to insertion into the main content. The techniques provide a streamlined video consumption experience.