Technical Disclosure Commons

October 11, 2018

Preference-Based Acceleration of Video Material

Dimitri Kanevsky

Sara Basson

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Recommended Citation

Kanevsky, Dimitri and Basson, Sara, "Preference-Based Acceleration of Video Material", Technical Disclosure Commons, (October 11, 2018)
https://www.tdcommons.org/dpubs_series/1585

Preference-Based Acceleration of Video Material

Dimitri Kanevsky
Sara Basson

Abstract:

Video material, such as movies and TV shows, often includes visual and auditory information for users that have sensory disabilities. For example, a blind movie watcher will listen to audio descriptions of visual scenes of the movie and a deaf movie watcher will read descriptions (e.g., closed captioning) of audio information of the movie. Users may prefer to skip or accelerate certain portions of the video material while other portions of the video material are presented in full. These users may include viewers with various disabilities, such as blindness, deafness, autism and so forth, that prefer to only watch portions of the video material that contain specific details while skipping other portions of the video material. Alternately or additionally, the users may include a viewer that simply prefers to only watch portions of the video material that is of interest to them, such as particular types of scenes, certain actors, or particular details, while skipping the other portions that may not be of interest. It would be beneficial if playback of video material could be customized to a particular user, such that portions of the video are accelerated or summarized, and only preferred portions of the video are presented in full. Techniques for a preference-based acceleration of video material are described.

Keywords: Video accelerated playback, audio description, visual description, captioning, sensory disabilities, blind, visual impairment, deaf, hearing impairment, preference-based video playback

Background:

Today, video material is presented to users with sensory disabilities by audio description (e.g., narration track) or text description (e.g., closed captioning).   When video material is made accessible for users that are blind or have a visual impairment, the video material can include long

tracks of audio description that describes what is taking place visually in scenes of the video material. For example, the audio description can include extensive dialogs or conversations, musical interludes, and narrations that describe scenery, landscapes, and other visual cues to provide context of the visuals taking place in the scenes of the video. When video material is made accessible for users that are deaf or have a hearing impairment, the video material can include long text descriptions that describes sounds that are taking place audibly in the video material. For example, the text descriptions can include long descriptions that depicts dialogue, background noises, and other audio cues to provide context of the audio taking place in the scenes of the video.

Description:

Video material, such as movies and TV shows, often includes visual and auditory information for users that have sensory disabilities. Users may prefer to skip or accelerate certain portions of the video material while other portions of the video material are presented in full. These users may include viewers with various disabilities, such as blindness, deafness, autism and so forth, that prefer to only watch portions of the video material that contain specific details while skipping other portions of the video material. Alternately or additionally, the user may be any viewer that simply prefers to only watch portions of the video material that is of interest to them, such as certain types of scenes, particular actors, or selected genre, while skipping the other portions of video material that may not be of interest or of less importance. Preferences of the users consuming the video material can be utilized to present the video material in an accelerated manner.

As depicted in Fig. 1 below, a deaf user is watching a video, such as movie or TV show, where sounds and dialogue of scenes in the video are presented to the deaf user via text description on screen. The sounds and dialogue of the video are presented in text form, which can include

conversations, descriptions of background noises, ringing phones, and other audio cues that are described for the deaf user to have context of the sounds taking place in the scene of the video.
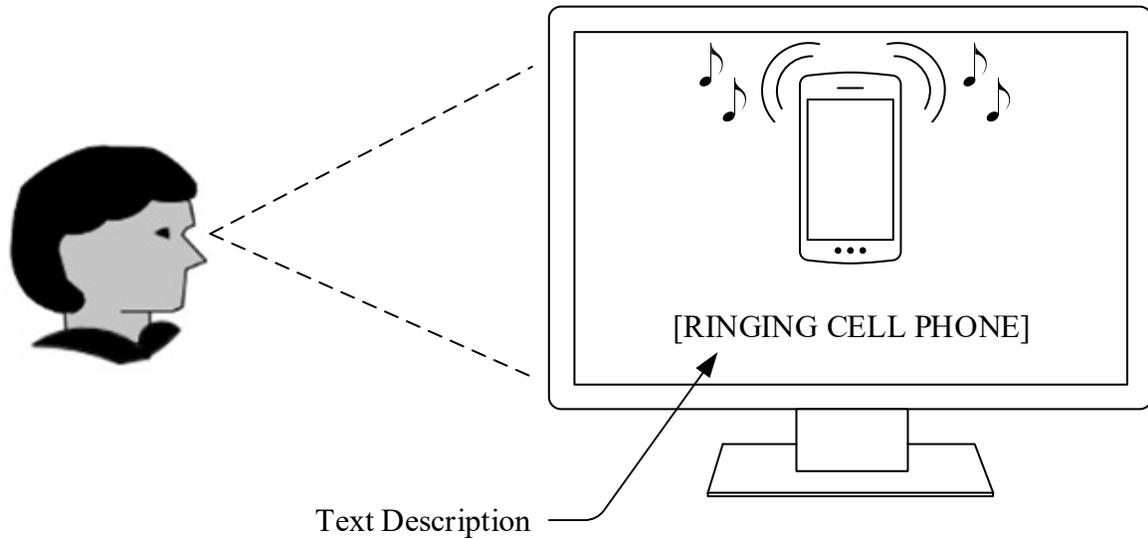

Fig. 1

As depicted in Fig 2 below, a blind user is watching a video such that the visual information of the video is presented to the blind user via audio description. The audio description consists of a narrator verbally describing what is taking place visually in a scene of the video during natural pauses in dialogue or audio of the video.
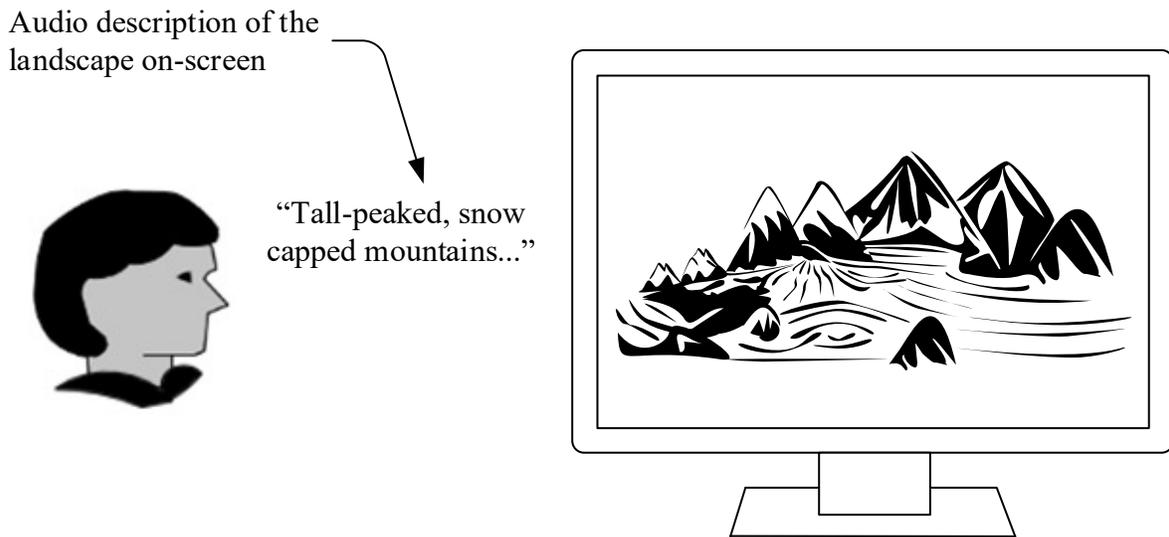
Fig. 2

Improvements for users watching video material can be achieved using a variety of techniques. In particular, techniques relying on machine learning can be applied to identify significant portions of the video material that can be shown or heard, and identify less-significant portions of the video material that can be skipped and summarized. This can be accomplished by collecting feedback from users on what portions of the video material they consider most significant or judging users' attention via eye tracking and other sensors.

For example, a user can indicate through an application that they want all audio-heavy portions of the video material to be presented in an abbreviated, summarized fashion while the video information is presented in full form. The audio-heavy portions of the video material can be scenes that include long periods of musical interludes. The user may prefer to not watch these long periods of musical interludes, and prefers to see a summarized description of these periods instead.

Another user can indicate that visual-heavy information is presented in an abbreviated, summarized fashion while the audio information is presented in full form. The visual-heavy information can include scenes having long tracks of audio description where a narrator verbally describes what is taking place visually in the scenes. The user may prefer to not hear these long audio descriptions, and prefers to listen to a summarized version of the long audio descriptions. The summarized version can be created using existing language summarization software, and presented to the user through text-to-speech synthesis to enable the user to skip through these portions of the video that have long audio descriptions.

Portions of the video material can be identified where audio and/or video information is presented in full form or in an abbreviated, summarized manner. Fig. 3 illustrates steps of an example method used to identify and present video-heavy information of the video material based on a time-based user preference. As illustrated, the techniques determine whether a scene of a video has an audio description that is greater than 5 minutes. It should be noted that the determination of the audio description being greater than 5 minutes is based on a user preference that is inputted by the user or learned by an application that determines that the user prefers to see audio descriptions longer than 5 minutes in an abbreviated, summarized manner. The threshold of 5 minutes in the example method shown below can be other values that the user prefers, such as 30 seconds, 2 minutes, 15 minutes, and so forth, *e.g.*, by selection or machine learning.

If the audio description of the scene is greater than 5 minutes (the "yes" path), then the audio description is skipped. A summarized description of the audio description is provided to the user and playback of the video continues. For example, the scene may contain a long narration that describes the scenery, landscape, and other visual cues of the scene. Instead of listening to the long narration in real-time, an abbreviated summary of the narration is provided to the user. If the audio description of the scene is not greater than 10 minutes (the "no" path), then the audio description is provided in full form to the user.
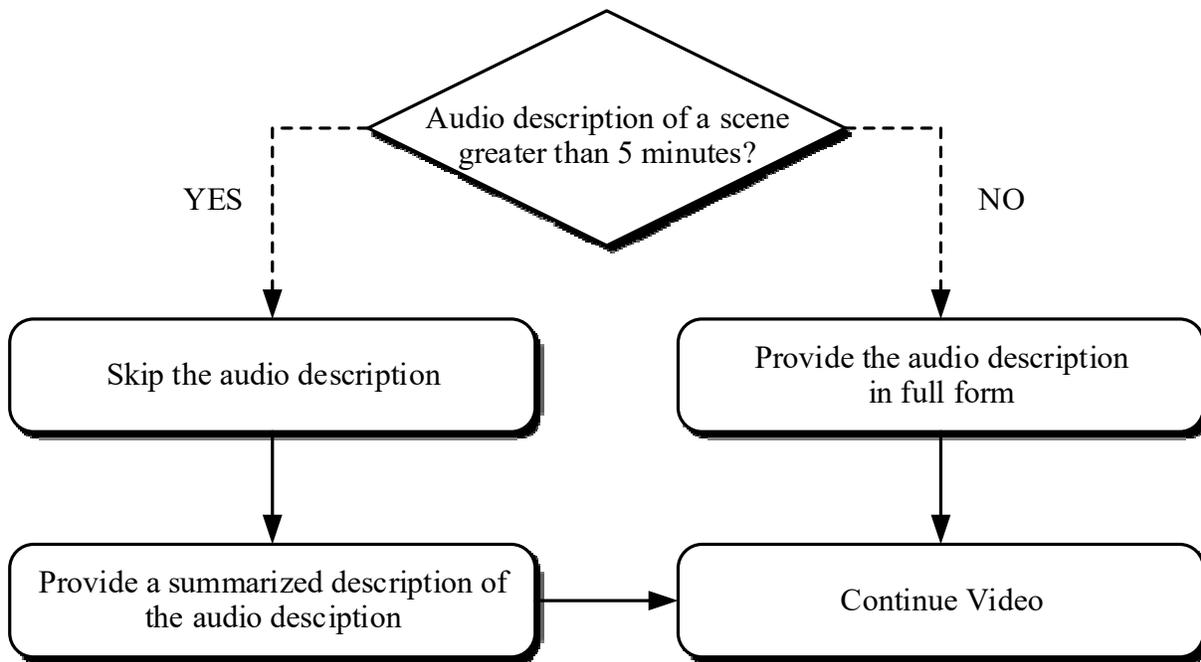
Fig. 3

Fig. 4 illustrates steps of an example method used to identify and present audio-heavy information of the video material based on a time-based user preference. As illustrated, a determination is made as to whether a scene of a video has a soundtrack that is greater than 5 minutes. It should be noted that the determination of the soundtrack of the scene being greater than 5 minutes is based on a user preference that is inputted by the user or learned by an application that determines that the user prefers to see soundtracks longer than 5 minutes in an abbreviated, summarized manner. It should also be appreciated that the threshold of 5 minutes in the example method shown below can be other values that the user prefers, such as 30 seconds, 2 minutes, 15 minutes, and so forth, *e.g.*, by selection or machine learning.

If the soundtrack of the scene is greater than 5 minutes (the "yes" path), then the soundtrack for the scene is skipped. A summarized description of the soundtrack is then provided to the user and the video continues. For example, the scene may contain long periods of music, and instead

of playing the music in real-time, a summary of the music in the scene is provided to the user, such as "musical composition playing <song> by <artist> for 6 minutes." If the soundtrack of the scene is not greater than 5 minutes (the "no" path), then the soundtrack is provided in full form to the user.
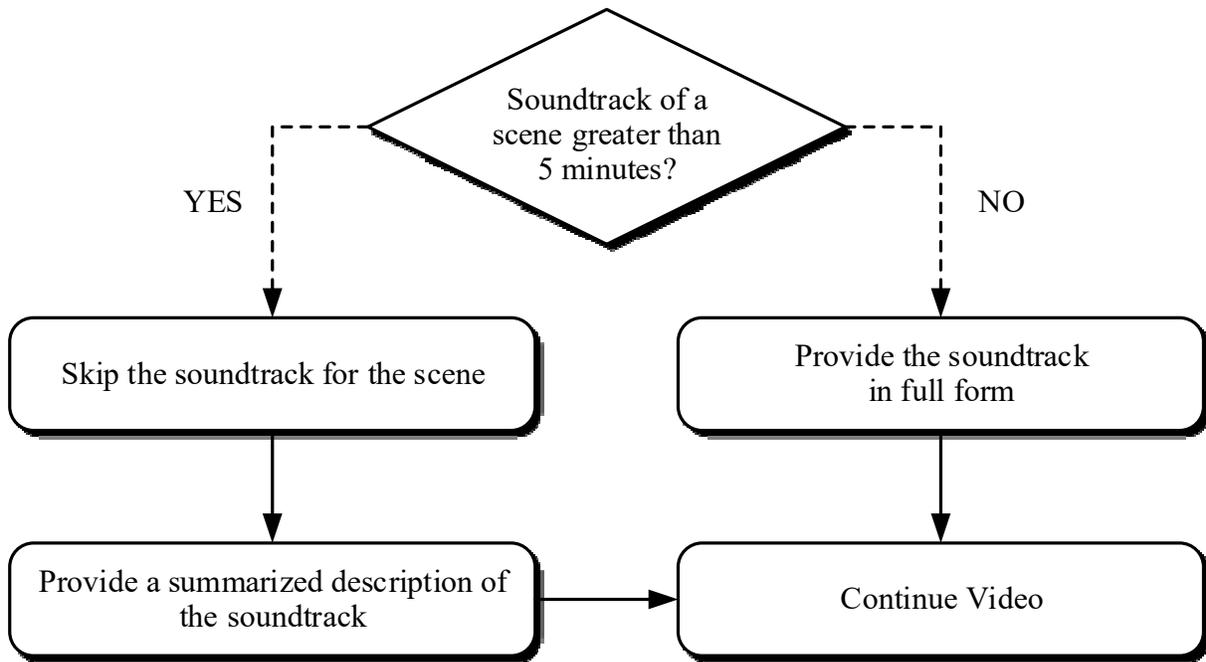

Fig. 4

Preferences other than the time-based user preference described above can be utilized to accelerate playback of the video material. For example, portions of the video material can be marked by the user that they would prefer to watch in an accelerated and abbreviated manner. If the user's preferences are consistent, the application can learn what portions of video material the user prefers to see in full and other portions the user prefers to watch in an accelerated or abbreviated manner.

By way of example, consider three user preferences, and how the techniques use these preferences. Assume a user selects, or has a prior habit of watching, only scenes that include a

particular actor (or skips through an actor or selects not to watch a particular actor). Scenes that do not include this actor can be skipped and replaced with summarized descriptions inserted between scenes that include the actor (or vice versa).

Second, assume a user prefers viewing video material without certain types of scenes of the video material. For example, the user may not want to watch scary or sexual-themed scenes of a video, or the user may not want to watch violent or gory scenes. Scenes that are identified as any of these by the techniques, such as through video material analysis, user selection, or other users' indications, can be skipped and replaced with summarized descriptions.

Third, assume a user indicates a certain amount of time the user is available for watching the video material. For example, a user traveling on a plane can indicate a traveling time of one hour that reflects the amount of time the user is on the plane or allowed to watch the video material during flight. If the video is two hours in length, the techniques adjust playback of the video to remove portions in order to complete the video material during the one hour traveling time. The removed material may instead be summarized into a stream of the video as described above.