

Technical Disclosure Commons

Defensive Publications Series

April 2020

Video Identification Based On Video Snippet Fingerprinting

Roman Vorushin

Filip Pavetić

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

Recommended Citation

Vorushin, Roman and Pavetić, Filip, "Video Identification Based On Video Snippet Fingerprinting", Technical Disclosure Commons, (April 03, 2020)
https://www.tdcommons.org/dpubs_series/3096



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.

Video Identification Based On Video Snippet Fingerprinting

ABSTRACT

This disclosure describes techniques that enable a user to identify a video or a movie by capturing a video snippet of a screen on which the video or movie is being displayed with a camera, e.g., a smartphone or other camera. The captured video snippet is cropped to the content of the screen. An embedding (fingerprint) is generated for the video and is matched to stored embeddings of video content to identify the video or the movie. Information regarding the identified movie, e.g., options to purchase or stream the movie, read reviews, etc. are provided to the user.

KEYWORDS

- Video identification
- Song identification
- Video embedding
- Video fingerprint
- Audio embedding
- Audio fingerprint
- Virtual assistant

BACKGROUND

Software applications are available that enable a user to identify a song being played. For example, a user can issue a query to a virtual assistant, e.g., “what song is playing now?” in order to get a response that identifies the song. In a similar manner, a user sometimes catches a glimpse of a movie or a TV show and wants to know the name of the movie or TV show, for example, to watch it later.

DESCRIPTION

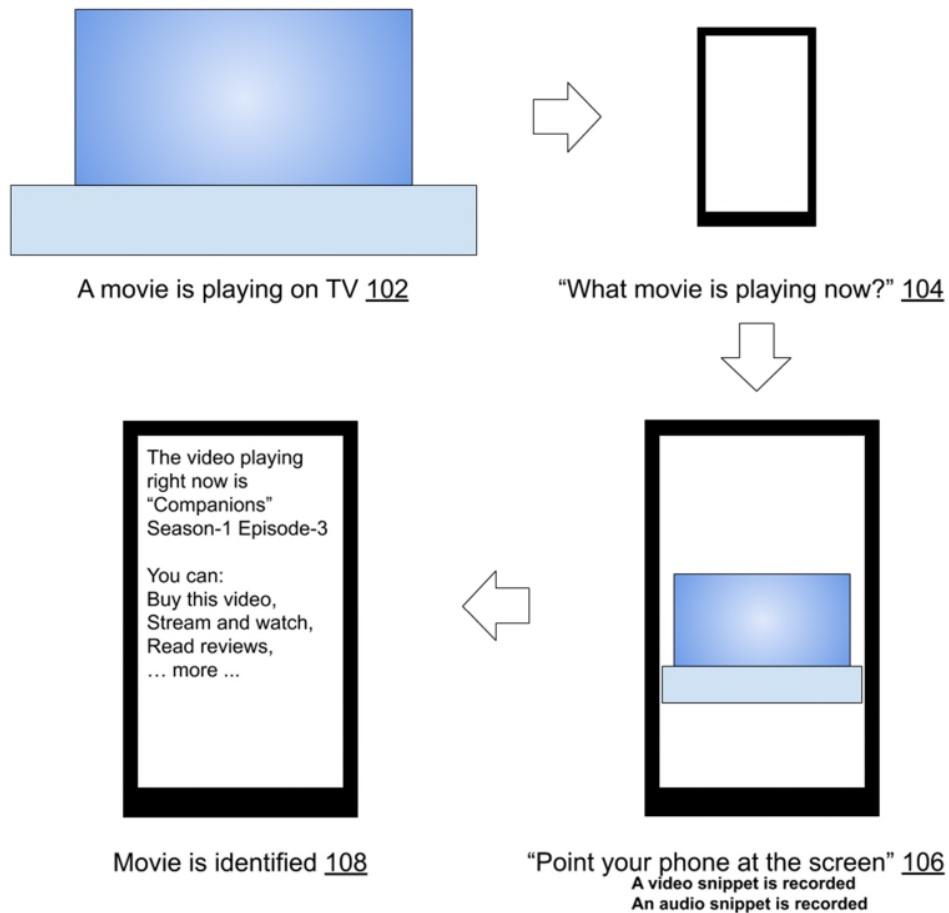


Fig. 1: Identifying a video or a movie using a virtual assistant or mobile device

Fig. 1 illustrates identification of a video or a movie using a virtual assistant or mobile device, per the techniques of this disclosure. A video or a movie plays on a screen, e.g., a television (102) or other screen. The user requests the virtual assistant or mobile device to identify the movie (104), e.g., by asking “what movie is playing right now?” The virtual assistant responds (106) by requesting the user to point the device camera at the screen. Alternatively, the user can provide a video clip that was previously captured and request identification of the video content.

With user permission, a video snippet of the video that is displayed on the screen is recorded. The movie is identified (108) using techniques described with reference to Fig. 2 below. Further options or links can be presented to the user, e.g., to buy a copy of the video, to stream the video, to read reviews, to share the video with other users, etc.

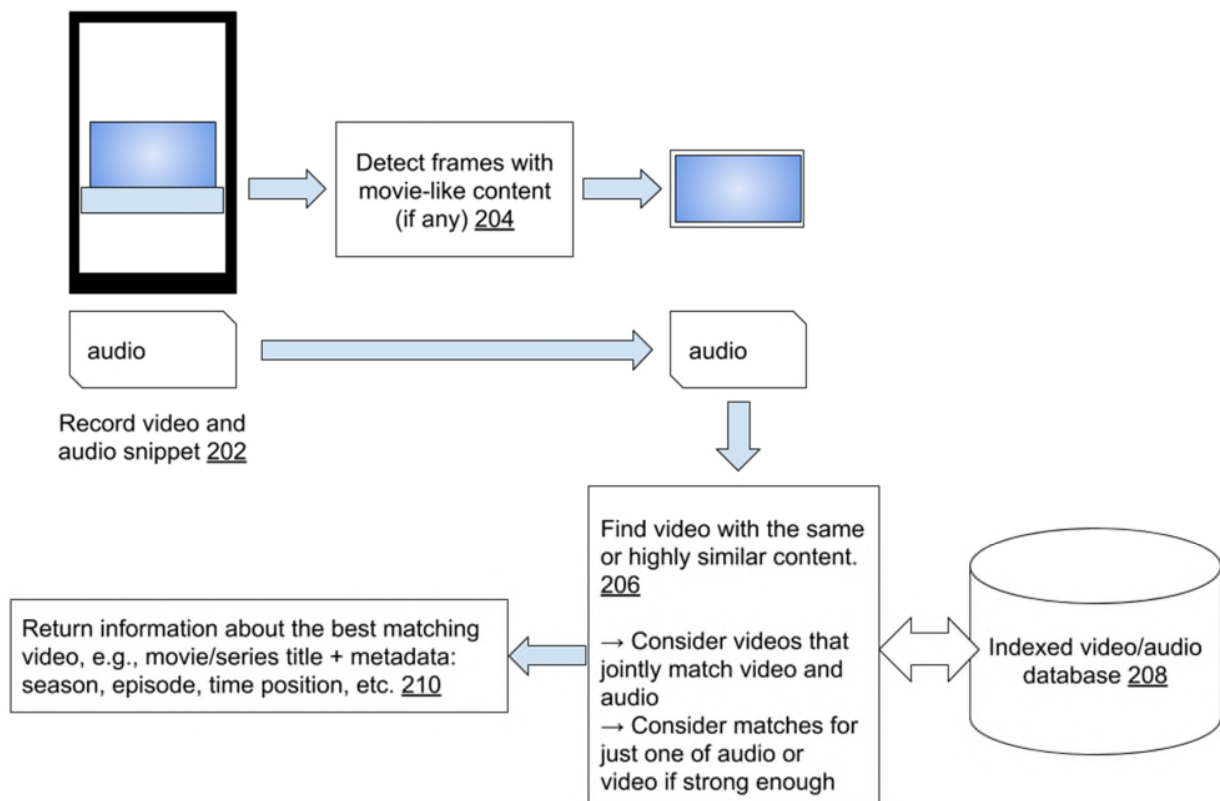


Fig. 2: Video identification by matching of video and/or audio snippet

Fig. 2 illustrates video identification by matching a video snippet (and optionally, audio from the snippet) per the techniques of this disclosure. With user permission, a snippet of a video being displayed on a screen is captured, e.g., on a mobile device (202), along with audio if available. The snippet can be a short snippet, e.g., 15-30 seconds long. The captured video may include portions of the room or location that the screen is placed in, e.g., other objects in the room. The video snippet is analyzed to remove the portion that includes such objects and retain

only video content that is displayed on a screen, e.g., frames with movie-like content are detected and extracted (204).

The captured video snippet is transformed into a video embedding (fingerprint), e.g., by providing it as input to a machine learning model (e.g., a deep-learning neural network) that generates embeddings from the video. In a similar manner, an audio fingerprint is obtained of the captured audio.

The embeddings of the captured video snippet are compared with a database (208) of embeddings of known shows and movies. The database is indexed by audio/video embedding. By matching the embeddings, videos from the database with content similar to the snippet (206) are identified. A joint audio-video match is sought; however, only audio or only video matches can also be considered, provided such matches are strong enough.

Information regarding the best matching movie or series, e.g., title and metadata such as season, episode, time-position of the captured snippet, etc., is returned (210). Processing of the captured snippet can be done locally on the device, on a server, or a combination thereof. Upon identification of the video, additional information is retrieved to provide further options to the user. For example, such information can include information regarding online or offline merchants or services that offer the video for sale or rent, reviews of the video, etc.

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 social network, social actions or activities, profession, 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

This disclosure describes techniques that enable a user to identify a video or a movie by capturing a video snippet of a screen on which the video or movie is being displayed with a camera, e.g., a smartphone or other camera. The captured video snippet is cropped to the content of the screen. An embedding is generated for the video and is matched to stored embeddings of video content to identify the video or the movie. Information regarding the identified movie, e.g., options to purchase or stream the movie, read reviews, etc. are provided to the user.

REFERENCES

[1] "Are there any apps like Shazam for movies?" <https://www.quora.com/Are-there-any-apps-like-Shazam-for-movies>, accessed Mar. 24, 2020.