January 28, 2019

Identifying Abusive Videos Inserted In A Video

Filip Pavetic

Roman Vorushin

IDENTIFYING ABUSIVE VIDEOS INSERTED IN A VIDEO

ABSTRACT

Disclosed herein is an improved mechanism for identifying abusive videos that have been inserted in a video. The mechanism can identify a video to be analyzed. The mechanism can then identify, for each frame of the video, a group of candidate locations within the video at which a two-dimensional video is likely to be embedded. The system can cluster the identified candidate locations and identify a subset of the candidate locations based on the clusters. The system can then identify a two-dimensional video embedded at one of the locations included in subset of candidate locations.

BACKGROUND

Video content providers can receive uploaded video content, store the uploaded video content, and then provide the uploaded video content to many users, for example, by streaming the video content to multiple user devices. These video content providers may determine whether the uploaded video content matches video content in a reference database, for example, that includes reference files of copyrighted content. However, in an effort to bypass or circumvent these detection systems, some content uploaders may upload a first video, where a second video is embedded within the first video and where the second video may be copyrighted or otherwise protected. It can be difficult to identify abusive videos that are embedded within another video. Thus, there is a need for an improved mechanism for identifying abusive videos inserted in a video.

DESCRIPTION

The systems and techniques described in this disclosure relate to identifying abusive videos inserted in a video. The system can be implemented on a server, such as a server associated with a media content sharing service. FIG. 1 illustrates an example process for identifying abusive videos inserted in a video.

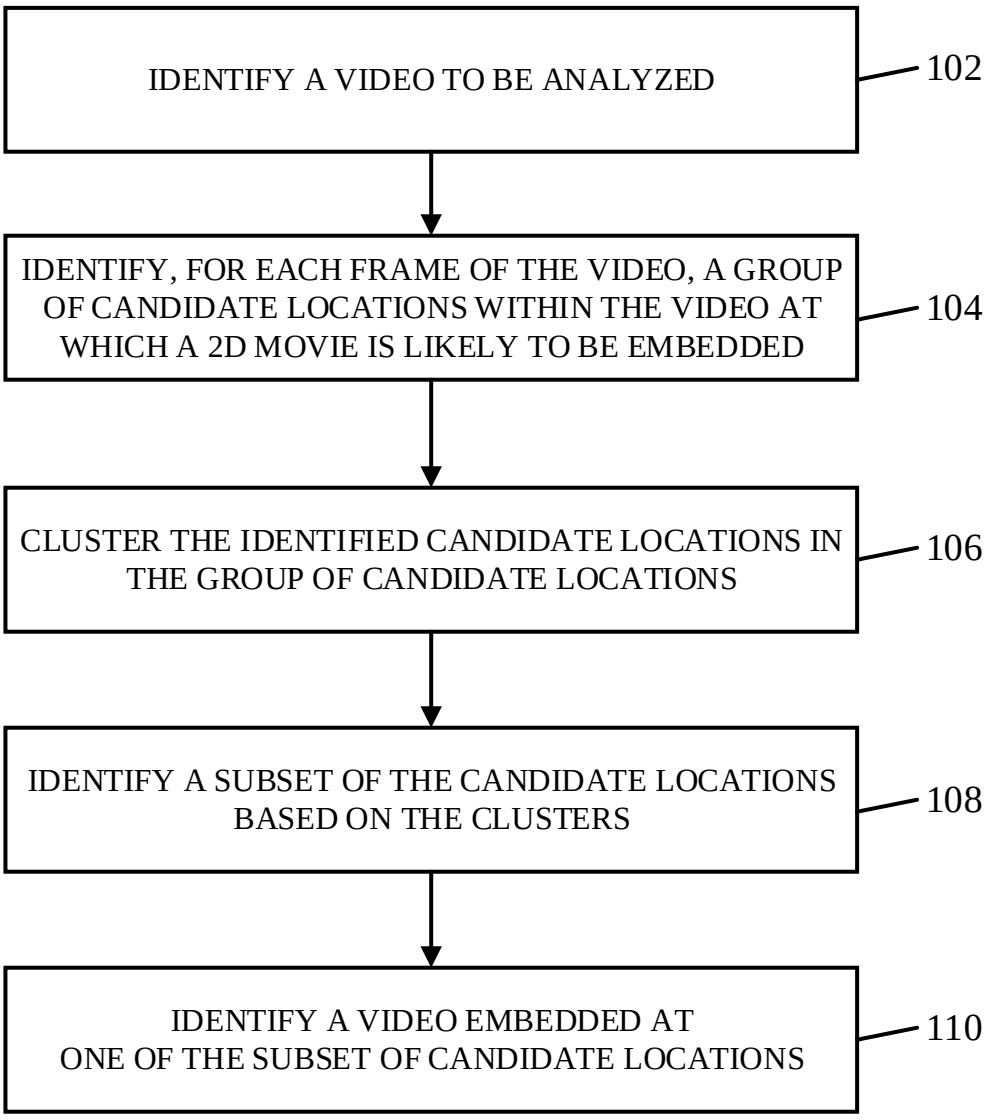

FIG. 1

At 102, the system can identify a first video to be analyzed. In some instances, the system can identify the video to be analyzed in any suitable manner. For example, the system can identify a video from a queue of videos that are to be analyzed (e.g., recently uploaded videos, and/or any other suitable queue of videos). In some instances, the first video to be analyzed can be any suitable type of video, such as a two-dimensional video, a three-dimensional video, a spherical video that can be manipulated by a viewer, and/or any other suitable type of video.

The system can identify, for each frame of the first video, a group of candidate locations within the video at which a two-dimensional movie is likely to be embedded at 104. In some instances, the system can identify each candidate location in the group of candidate locations using any suitable technique or combination of techniques. For example, the system can use a neural network that has been trained to identify spatial positions within a video that are likely to contain an embedded two-dimensional video. In some such instances, the neural network can be trained using any suitable training set, such as a training set of videos that have been manually annotated as having an embedded video at particular spatial positions. In some instances, an identified candidate location can be of any suitable size.

Note that, in some instances, each identified location in the group of candidate locations can be associated with a confidence level that indicates a confidence that the location includes an embedded video during the corresponding frame of the first video. For example, in some instances, an output of the neural network can indicate that an identified location at a particular frame of the first video has a 50% likelihood of including an embedded two-dimensional video.

Additionally, note that although the system is described herein as identifying a group of candidate locations for each frame of the first video, in some instances, the system can identify

the group of candidate locations for any suitable subset of the frames of the first video (e.g., for every other frame, for every tenth frame, for frames identified in any suitable manner as likely to include interesting content, and/or any other suitable subset of the frames).

At 106, the system can cluster the identified candidate locations in the group of candidate locations. The system can cluster the identified candidate locations based on any suitable criteria. For example, in some instances, the system can cluster the identified candidate locations such that candidate locations identified at different frames of the first video are assigned to different clusters. As a more particular example, in some instances, a first location (e.g., an upper left portion of the first video, and/or any other suitable location) identified for a first frame and a second location (e.g., an upper right portion of the first video, and/or any other suitable location) identified for a second frame can be assigned to two different clusters. As another example, in some instances, the system can cluster the identified candidate locations such that a candidate location identified for multiple frames of the first video are clustered together. As a more particular example, in an instance where a particular location (e.g., an upper left portion of the first video, and/or any other suitable location) is identified for multiple frames of the first video as being likely to include an embedded two-dimensional video, the multiple instances (that is, those instances of the particular location identified at each of the multiple frames of the first video) of the particular location can be clustered together. In some instances, the clustering can result in candidate locations that have been identified at multiple time points of the first video to be clustered together.

At 108, the system can identify a subset of the candidate locations from the group of candidate locations. In some instances, the system can identify the subset of the candidate locations in any suitable manner. For example, in some instances, the system can identify

candidate locations corresponding to clusters having more than a predetermined number (e.g., more than ten, more than twenty, and/or more than any other suitable number) of candidate locations or more than a predetermined number of instances of a candidate location. As another example, in some instances, the system can identify the *N* (e.g., one, two, five, and/or any other suitable number) largest clusters identified at 106 and can identify the candidate locations assigned to each of the *N* largest clusters.

At 110, the system can identify a two-dimensional video embedded at at least one of the candidate locations in the subset of candidate locations. The two-dimensional video can be identified in any suitable manner. For example, in some instances, the system can apply any suitable video fingerprinting technique(s) to each candidate location in the subset of candidate locations to identify a two-dimensional video embedded at the candidate location. To determine whether an abusive video has been embedded within an uploaded video at a candidate location, these fingerprints can be compared against a database of reference files that contain, for example, fingerprints of previously uploaded videos.

Accordingly, a mechanism for identifying abusive videos that have been inserted in a video is provided.