Technical Disclosure Commons

Defensive Publications Series

December 12, 2017

Using Fingerprints To Position Advertising Content

Johan Granstrom

Steven Dropsho

USING FINGERPRINTS TO POSITION ADVERTISING CONTENT

ABSTRACT

Disclosed herein is a mechanism for using content-based video and audio fingerprints to find suitable locations for placing midroll advertising content. The mechanism can include assigning penalty information to different locations within a video, where the penalty information can be a combination of an audio penalty score and a video penalty score. The mechanism can use the penalty information to determine a suitable location for placing or positioning content within the video, such as midroll advertising content.

BACKGROUND

Advertising content placed within a video, such as midroll advertising content, is increasingly being provided to users. Such advertising content is generally considered to be less disturbing to the user as the user is invested in the consumption of the video. Because of this, advertising content placed within a video is more likely to be viewed to completion by users.

Moreover, this type of advertising content is more likely to increase in value if it is provided at a suitable position within a video. For example, a midroll video advertisement may not be well-received or, in some cases, viewed to completion if it interrupts a particular portion of the video. In another example, a midroll video advertisement may not be well-received or consumed if it is positioned at a particular portion of the video that interrupts the audio portion of the video. Determining whether one or more suitable positions within a video is suitable for placing such advertising content proves to be a difficult task.

DESCRIPTION

Generally speaking, the mechanism relates to using content-based video and audio fingerprints to find suitable locations for placing midroll advertising content. The mechanism can include assigning penalty information to different locations within a video, where the penalty information can be used to determine a suitable location for positioning content within the video, such as midroll advertising content. For example, penalty information can be assigned to each location in a video, where lower-scored penalty information can indicate that the location is a preferable location to place advertising content within the video, such as midroll advertising content.

FIG. 1 is an illustrative example of a method for using content-based video and audio fingerprints to determine one or more suitable locations for placing advertising content, such as midroll advertising content, within a video.

Turning to FIG. 1, at step 110, a video server can begin by obtaining an audio fingerprint associated with a particular portion or segment of a video. For example, in some instances, the video server can retrieve the video, divide the video into a particular number of segments (e.g., segments containing 1,000 milliseconds of audio content), and generate an audio fingerprint for each segment of audio content.

It should be noted that an audio fingerprint is analogous to a human fingerprint where small variations that are insignificant to the features characterizing the fingerprint are tolerated or ignored. In some implementations, the audio fingerprint is a numerical representation (e.g., a vector) of the audio segment including one or more attributes, such as average zero crossing rate, estimated tempo, average spectrum, spectral flatness, prominent tones across a set of bands, and bandwidth. One or more of these attributes can be determined by performing a frequencydomain spectral analysis of the audio segment. Compared with the audio segment itself, the audio fingerprint can be generated to focus on the perceptual characteristics of the audio segment. For example, if two audio segments sound alike to the human ear, their corresponding audio fingerprints should match, even if their binary representations are different.

In a more particular example, the video server can generate an audio fingerprint (e.g., a 1,000-dimensional floating point vector) for each 1 second segment that corresponds to 10 seconds of audio content centered around a location and that encodes features of the audio content in the 10 second window.

At step 120, the video server can determine an audio penalty score from the audio fingerprint. In some instances, the audio penalty score for a given location can be calculated as the magnitude of the audio fingerprint vector for that location. It should be noted that the magnitude of the audio fingerprint vector can correspond to how much interesting sound was present within the window of audio content (e.g., a 10 second window). Accordingly, by determining an audio penalty score based on the complexity of the sound in an audio segment, the audio penalty can be used by the video server to inhibit the placement of advertising content, such as a midroll advertisement, at a location that interrupts a portion of the audio content stream that is likely to contain important or interesting audio content.

It should be noted that, in some instances, scaling can be applied to balance the audio penalty score with the video penalty score described hereinbelow. For example, a linear scaling function, a unary function, or any other suitable scaling function can be applied.

At step 130, similar to step 110 above, the video server can obtain a video fingerprint associated with a particular portion or segment of a video. For example, in some instances, the video server can retrieve the video, divide the video into a particular number of frames, and generate a video fingerprint for each frame of video content.

It should be noted that the video fingerprint for each segment of video content can be generated using any suitable approach. For example, a video fingerprint can be generated based on unique video features identified through a particular set of frames. In a more particular example, a video fingerprint can be generated based on a hash of a spectrogram window (e.g., a min-hash technique). In another example, a combination of computer vision techniques and/or data stream processing algorithms can be implemented to generate a video fingerprint. In yet another example, a video fingerprint can be generated based on one or more wavelet vectors.

In a more particular example, the video server can generate a video fingerprint (e.g., a 1,000-dimensional floating point vector) from a single frame of the video that encodes features of the video content in that frame.

At step 140, the video server can determine a video penalty score from the video fingerprint of a particular frame. In some instances, the video penalty score for a given location can be calculated as the magnitude of the video fingerprint vector for that location. It should be noted that the magnitude of the video fingerprint vector can correspond to the complexity of the image within the frame and the distance between two video fingerprints can correspond to how visually different they are from each other. It should be noted that any suitable distance between two video fingerprints can be calculated, such as a Euclidean distance measurement or a cosine similarity measurement. Accordingly, by determining a video penalty score based on the complexity of the image in a video frame and based on the distance between two adjacent video fingerprints, the video penalty can be used by the video server to inhibit the placement of advertising content, such as a midroll advertisement, at a location that interrupts a portion of the video that is likely to contain important or interesting video content.

For example, a suitable location within the video (e.g., a natural break in a video stream) can be found from one or more low-complexity fingerprints (e.g., low magnitude of the video fingerprint vector), where the video fingerprint before and/or after this region of one or more low-complexity fingerprints has a large distance (e.g., visually different video regions).

In a more particular example, this can be represented as follows:

E(i) = embedding at position *i*, i.e., at *i* * 1,000 milliseconds into the video

M(i, j) = mean of *len(E(t))* for *t* in the interval [*i - floor(j/2), i + ceil(j/2)*]

Cosine similarity of two vectors: *C(m, n)* = (*m* · *n*) / *len(m)* / *len(n)*

Similarity of two vectors (where cosine similarity is translated to interval [0, 1] and modified so that long vectors pointing in different directions are less similar than short vectors): *S(m, n)* = (1 + *C(m, n)*) / 2 / log10(*len(m)*) / log10(*len(n)*)

Accordingly, the video penalty score at position *i* looking at a window of size *j* can be represented as:

P(i, j) = M(i, j) / S(E(i - floor(j/2) - 1), E(i + ceil(j/2))), where

j > 0 and the window is *floor(j/2)* to the left and *ceil(j/2)* to the right.

In addition, each window size *j* is given a multiplicative factor *F(j),* e.g.,

F(1) = 1.0,

F(2) = 1.0,

F(3) = 1.0,

F(4) = 1.0 (1s),

F(5) = 0.95,

F(6) = 0.95,

$F(7) = 0.9,$

$F(8) = 0.9$ *(2s)*,

$F(9) = 0.85,$

$F(10) = 0.85,$

$F(11) = 0.9,$

$F(12) = 0.9$ *(3s)*,

$F(13) = 0.95,$

$F(14) = 0.95,$

$F(15) = 1.0,$

$F(16) = 1.0$ *(4s), etc.*

It should be noted that, in some instances, these numbers are used so that the preferred window of an advertisement break is between 2.25 seconds to 2.5 seconds.

The final penalty score for a given location i can be represented as:

$P(i) = min(P(i, j) * F(j))$ for j in [1, 16].

It should be noted that this mathematical definition of the penalty score of having an advertisement break at position i captures a location that is determined to be visually "quiet" and surrounded by different types of content, i.e., a natural pause in the video.

It should also be noted that the functions, constants, and/or parameters mentioned in the equations above can be modified in any suitable manner to yield different ways to penalize the placement of advertising content at an unsuitable location.

Referring back to FIG. 1, at 150, the video server can calculate, for each location of the video, a penalty score that combines the audio penalty score and the video penalty score. For example, the penalty score can be the summation of the audio penalty score and the video penalty score for a location of the video. In another example, the penalty score can be the product of the audio penalty score and the video penalty score for a location of the video. It should be noted that any suitable function can be used to combine the audio penalty score and the video penalty score, such as a min function, a max function, a sigmoid function, etc. It should also be noted that, prior to the combination of the audio penalty score and the video penalty score for a location of the video, any suitable scaling, such as linear scaling, can be applied to balance the audio penalty score with the video penalty score.

At 160, the video server can determine one or more locations within the video for positioning advertising content, such as midroll advertising content. Using the penalty score for each location (e.g., with a granularity of 1,000 milliseconds) in a video, the video server can determine one or more locations within the video for positioning or placing advertising content based on those locations having the lowest penalty scores.

Accordingly, the video server implementing the mechanism described herein can place advertising content within a video at one or more suitable locations. For example, the video server can determine whether to place a midroll advertisement within a video or whether the video is not suitable for the placement of any midroll advertisements (e.g., there are no locations having a penalty score less than a particular threshold score). In another example, the video server can rank order multiple locations within a video that are determined to be suitable locations for placing a midroll advertisement and provide advertisers with opportunities to bid on particular locations.

Accordingly, a mechanism for using content-based video and audio fingerprints to find suitable locations for placing midroll advertising content is provided.

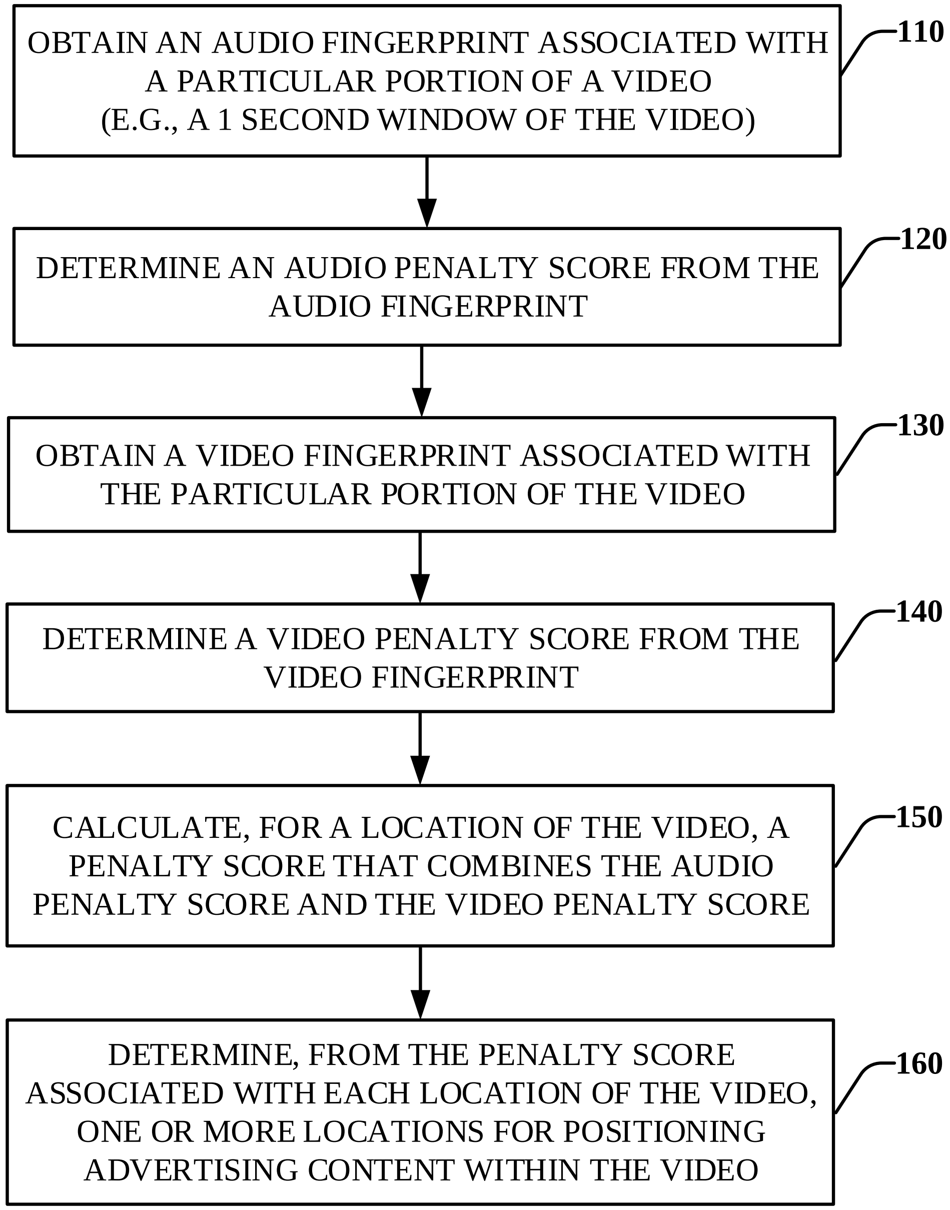
OBTAIN AN AUDIO FINGERPRINT ASSOCIATED WITH A PARTICULAR PORTION OF A VIDEO (E.G., A 1 SECOND WINDOW OF THE VIDEO)
110
DETERMINE AN AUDIO PENALTY SCORE FROM THE AUDIO FINGERPRINT
120
OBTAIN A VIDEO FINGERPRINT ASSOCIATED WITH THE PARTICULAR PORTION OF THE VIDEO
130
DETERMINE A VIDEO PENALTY SCORE FROM THE VIDEO FINGERPRINT
140
CALCULATE, FOR A LOCATION OF THE VIDEO, A PENALTY SCORE THAT COMBINES THE AUDIO PENALTY SCORE AND THE VIDEO PENALTY SCORE
150
DETERMINE, FROM THE PENALTY SCORE ASSOCIATED WITH EACH LOCATION OF THE VIDEO, ONE OR MORE LOCATIONS FOR POSITIONING ADVERTISING CONTENT WITHIN THE VIDEO
160

FIG. 1