Technical Disclosure Commons

Defensive Publications Series

January 13, 2017

INCREASING AD DETECTION RELIABILITY BY LEARNING PER-CHANNEL TEMPLATES AND TRANSFORMATION

Yaroslav Volovich

Ant Oztaskent

Simon Rowe

INCREASING AD DETECTION RELIABILITY BY LEARNING PER-CHANNEL TEMPLATES AND TRANSFORMATION

ABSTRACT

Content generators, such as video producers and advertisers, can provide their content to TV networks and other publishers for display. Automatically determining if the provided content appeared on a specific TV network can be difficult. Matching captured frames from the video streamed by the TV networks to frames from the provided content can result in a low match rate. The match rate can be low because the TV networks often apply various image transformations, such as scalings, croppings, color transformations, and overlays such as logos, tickers and other on-screen graphics that can obscure the content. When the transformations are applied, the frame images ultimately displayed by the publisher don't match the original frame images. The present paper discusses a system and method for determining, on a channel by channel basis, transformations and overlays that are applied by TV networks.

SYSTEM DESCRIPTION

After generating their content, content generators provide the content to TV networks and other publishers for display to the public. The content generator often desires to know how often and when the publisher displays the content. The content generator may also desire to know which publishers are displaying the content. One method content generators and other parties can use to determine if specific content was displayed is to capture screen frames displayed by the publisher. The captured screen frames can then be matched to screen frames from the content generator's content. If one of the displayed screen frames matches a frame of the content generator's content, then the publisher is determined to have displayed the content. However, the

publishers can often alter the content prior to display. Once altered, it can be difficult to match the captured screen frames back to the original content. For example, Figures 1–3 illustrate how the same video may be displayed by three different publishers.

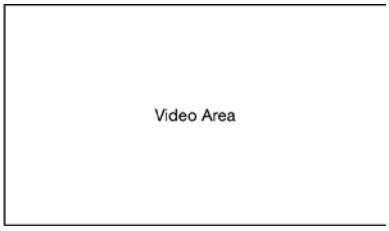

Figure 1

Figure 2

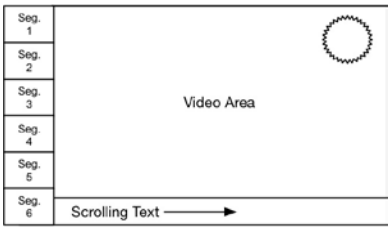

Figure 3

Figure 1 illustrates an example of a video that is displayed by a first publisher or provided via a first service (e.g. terrestrial, satellite, or internet broadcast, unicast, multicast, streaming, etc.). The first publisher does not alter the video and the video is displayed with no overlays, transformations, or other alterations. Frames from this video would have a high likelihood of matching back to the content generator's original content because the original content is not altered in a substantial way. Figure 2 illustrates video displayed by a second publisher. Here the video content was provided to the publisher in a widescreen aspect ratio. The publisher converted the video to a standard-width video by "letterboxing" the video, or adding mattes (e.g., black bars) above and below the video. Figure 3 illustrates a video displayed by a third publisher. Here, video several overlays are applied and the video is scaled (and/or cropped) to fit within the remaining area. The video illustrated in Figure 3, as with some news-based TV stations, includes a first overlay that can include scrolling text that is applied to the bottom of the screen. A side overlay was added that can include, for example, the upcoming video segments that will be displayed by the publisher. The original video is either scaled or cropped to fit into the remaining screen area. Figure 3 also illustrates that a badge was overlaid in the top right-hand corner of the

video. This overlay may include an onscreen graphic or station identification information (e.g., the publisher's station call letters, trademark, name, or other such information). Color or visual transformations can also be applied to the original video (e.g. contrast compression or expansion, hue adjustment, white balancing or rebalancing, etc.). Because transformations were applied to the video illustrated in Figures 2 and 3, when frames from the displayed video are compared to frames form the original video, the frames will have a low likelihood of matching.

The present system can identify transformed video and then generate transformation templates that indicate how video, for each channel, is transformed prior to its display. As an overview, during a learning phase, the system can collect reference video data (e.g., streamed TV video from TV networks or web-based sources) and compare the collected data to the original video to determine how each of the publishers transform displayed content. The system can determine, for each of the publishers (or TV channels) which transforms (e.g., overlays, scaling transformations, cropping transformations, color transformations, and other alterations) are likely to be applied to the content when displayed by the respective publisher. Once the transforms are determined for each of the publishers, they can be saved as templates. During a matching phase, the templates can be used to adjust the parameters and regions of a video/image based matching algorithm that matches subsequently captured video content to original video content.

As one example, during a learning phase, the system can capture TV channel streams from publishers to find occurrences of specific advertisements. The specific advertisements can then be compared to the original version of the advertisement from the content generator to generate a conversion template (e.g. scale by 80%, shift up by 60 pixels and right by 50 pixels, increase brightness by 10%, rotate hue by 5 degrees, etc.). The TV channel streams can be traditional broadcast TV channels from a cable or satellite provider or web-based video streams.

The system can extract the frames from the streams that correspond to the specific advertisements. Having been streamed by the publisher the extracted frames of the specific advertisements can include the TV channel's transformations. The system can extract a portion of the broadcast that is likely to be least altered, such as the central portion (e.g., the central 1/3 of the extracted frames), to remove scrolling text and other overlays that may be added by the TV channel. The system can compare each of the extracted frames to the corresponding frames of the original version of the advertisement (called the reference frames). First, the system can match scaling or cropping of the extract frames by applying horizontal and vertical scaling and offsets (e.g., a gradient descent on average, median brightness, or color difference) to match the extracted portion to a portion of the corresponding reference frames. The type and amount of cropping and scaling to match the extracted frames to the reference frames can be saved. Once the central portion of the extracted frames are scaled to match the central portion of the reference frames, the system can also take the difference between the central portion of the extracted frames and the central portion of the reference frames to generate a binary map. The binary map for a frame can indicate differences between the extracted frame and the corresponding reference frame. The system can identify areas of the binary map that are non-zero as corresponding to overlays in the extracted frame. In some cases, areas of the binary map above a predetermined (non-zero) threshold are identified as corresponding to overlays in the extracted frame. In some cases, the binary maps can be simplified by producing masking rectangles which cover the regions where significant differences are found. The binary mask, in its simple or full form, provides a mask of "ignore" regions that are not used to match newly captured video to reference data after the completion of the learning phase. The binary mask and the type and amount of cropping and scaling to match the extracted frames to the reference frames, as well as any other

detected transformations, are saved as a template into a database in association with an identification of the publisher.

After the learning phase, frames of live (or otherwise new) video are extracted by the system. The template for the corresponding publisher is applied to each of the live frames. The live frames, with the template applied, can then be compared to reference frames from predetermined content to determine if the content was displayed by the publisher. In some implementations, the system can generate multiple templates per publisher. For example, different templates can be generated for specific days of the week or times within each day.

Thus, the use of transformation guides or templates may allow for faster, more efficient computer vision recognition or “screen-scraping” of video streams to recognize third party content, by applying a common template to content on a channel or stream associated with a provider