Technical Disclosure Commons

Defensive Publications Series

November 2019

Change detection for optimized semantic video analysis

Elly Nattinger

Austin McCasland

Change detection for optimized semantic video analysis

ABSTRACT

Semantic analysis or annotation of videos is most useful when done frequently enough to capture the significant moments of a video, but not so frequently that annotations become busy and repetitive. With current techniques, semantic analysis is done too often, overloading the semantic analyzer and overwhelming the viewer with frequent, repetitive, or similar annotations of insubstantially differing frames. This disclosure presents techniques that detect substantial changes in the video for the purposes of semantic analysis. Timely and relevant annotations are presented to viewers without overwhelming them and without overloading the semantic analyzer.

KEYWORDS

- Video annotation
- Video summarization
- Semantic analysis
- Computer vision
- Image understanding
- Frame differencing

BACKGROUND

Semantic analysis or annotation of videos is most useful when done frequently enough to capture the most significant moments of a video, but not so frequently that annotations become busy and repetitive. With current techniques, semantic analysis is done too often, overloading the semantic analyzer and overwhelming the viewer with frequent annotations. Insubstantially differing frames are semantically analyzed, resulting in repetitive or similar annotations of the same objects in different frames and reduced user perceptions of quality of annotation.

A related issue is when the video changes too fast relative to the speed of the semantic analyzer. In such cases, the annotation lags the video, such that annotations no longer relevant to current content are shown on the screen, causing confusion to viewers. This is commonly a problem in livestream videos or just after a scene cut in any video.

DESCRIPTION

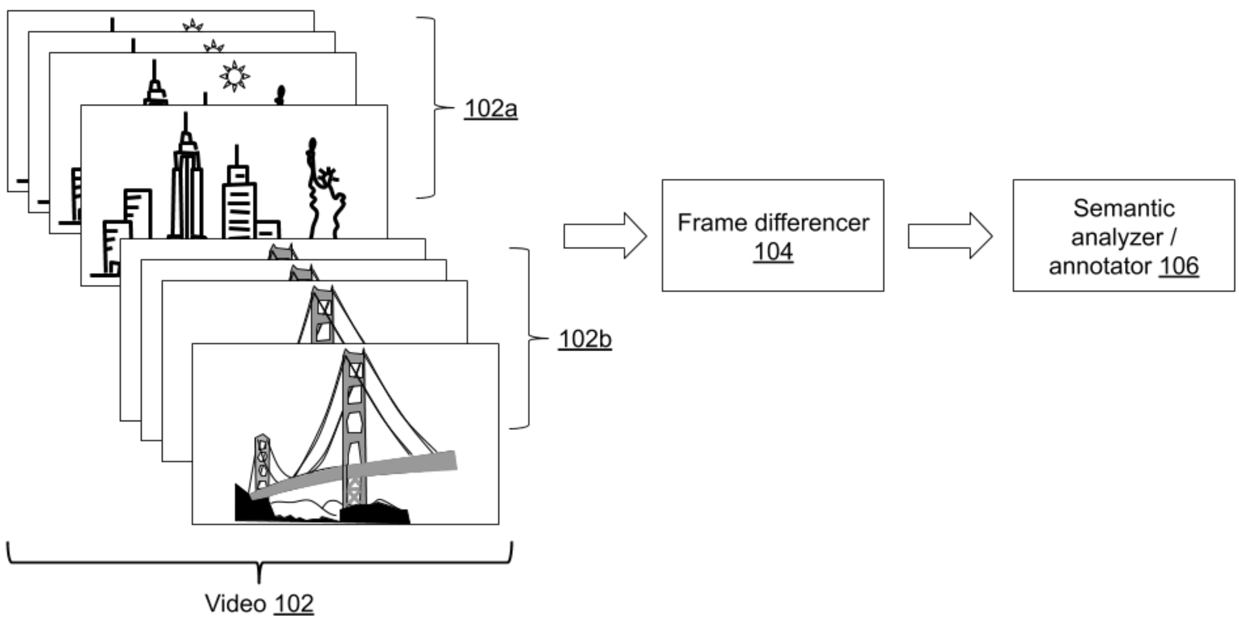

Fig. 1: Optimizing frame capture for semantic analysis

Fig. 1 illustrates optimizing frame capture for the purposes of semantic analysis, per techniques of this disclosure. A video feed (102) is fed into a frame differencer (104), which picks out pairs of frames with substantial differences between them. Substantially differing frames are fed to a semantic analyzer or video annotator (106). For example, for the purposes of semantic analysis, the frame-set 102a may be represented by a single frame, while the frame-set 102b may be represented by another, distinct frame. In this manner, frames are sent for semantic analysis when the change in the scene depicted in the video is sufficient (e.g., meets a change threshold) to trigger a round of semantic analysis. Excessive annotation is thereby prevented from reaching the viewer. The semantic analyzer is invoked less and is thereby prevented from being overloaded.

Differences between frames can be found in a number of ways. For example, the pixels of a video frame can be blurred and grey-scaled. Consecutive blurred and grey-scaled frames can be differenced, and the difference frame thresholded. If the number of pixels that have changed more than the threshold is greater than a certain amount, semantic analysis is triggered. The threshold can be chosen based on the type of video, e.g., videos of people may have different thresholds from videos of landscapes, or from videos of products for sale. Similarly, videos with fast-moving objects may have a different threshold than the threshold for videos with slow-moving objects. Thresholds can be established that detect scene cuts, camera panning, that the movement of people across the screen, etc.

In this manner, if there hasn't been a substantial change since the last semantic analysis, the video is determined to be relatively static and video analysis is not performed. Triggering semantic analysis over substantially different frames reduces viewer distraction by reducing the number of multiple, similar annotations for the same objects.

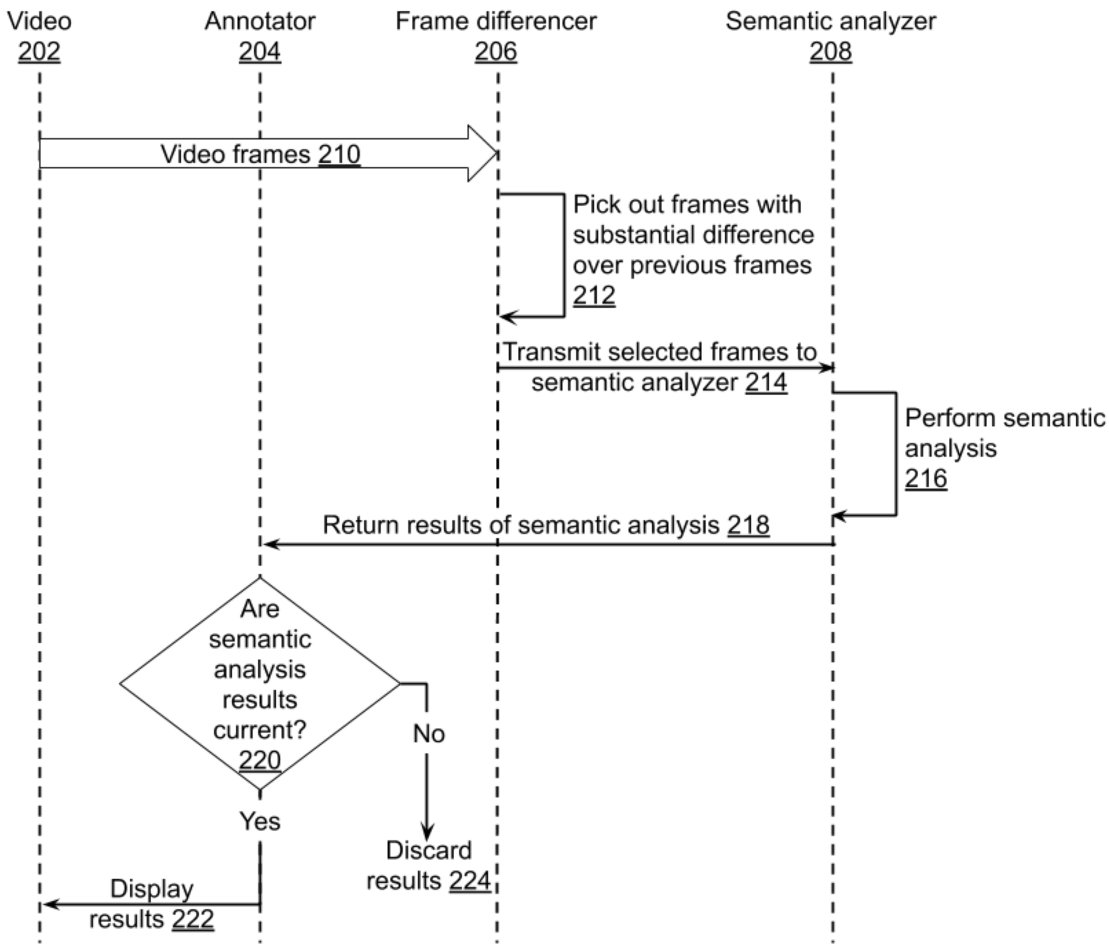

Fig. 2: Displaying annotations that are current to a frame

Sometimes semantic analysis can take enough time that the video has changed substantially by the time the annotations are ready to display. This can happen, for example, during livestream videos or during scene cuts in any video. In such cases, an object positioned or identified by the semantic analyzer is not relevant to the current frame of the video.

Fig. 2 illustrates displaying annotations that are current to a video frame, per techniques of this disclosure. A video (202) is received and played continuously. Video frames are sent (210) to a frame differencer (206). The frame differencer picks out frames with substantial difference over previous frames (212) and sends the selected frames (214) to a semantic analyzer. The semantic analyzer analyzes the frames (216) and returns the results of the semantic analysis (218) to a video annotator (204). The annotator verifies whether the results (220) are current with respect to the video frame being displayed, e.g., by determining the amount of pixel movement between the current frame and the frame that triggered the most recent semantic analysis. If the semantic results are current, e.g., there is not much movement between the current frame and the frame that triggered semantic analysis, the annotations derived from semantic analysis are displayed (222). If the semantic results are not current then the semantic analysis results are discarded (224).

Advantages

- Display of timely and relevant annotations to a video, without overwhelming the viewer with excessive annotations.
- Optimal loading of the semantic analyzer.
- Optimal use of bandwidth to cloud-based machine vision services, e.g., transmitting fewer and more relevant frames of a livestream video to cloud-based machine vision services.
- Reduced compute requirements for the device that captures or plays the video.
- Auto-generation of new content (in relation to a video) that provides better user experience than content that is generated at random or equally-timed display intervals.

CONCLUSION

This disclosure presents techniques that detect substantial changes in a video for the purposes of semantic analysis. Timely and relevant annotations are presented to viewers without overwhelming them and without overloading the semantic analyzer.