August 2020

Audio/video Quality Measurement

Anonymous

Audio/video Quality Measurement

ABSTRACT

Referential measurement tools that are used to measure video quality are typically very sensitive to synchronization and scaling issues that occur when video captured at a recipient device is out of synchronization with a source video and/or is scaled, e.g., to a different aspect ratio. When the captured video has such differences from the source video, conventional referential measurement techniques produce erroneous results. This disclosure describes techniques to process captured video to account for scaling effects and to synchronize captured video with source video. Referential measurement can then be performed with greater accuracy. A synchronization measurement technique is also described.

KEYWORDS

- Broadcast video
- Streaming video
- Video quality
- Referential measurement
- Video synchronization
- Video scaling

BACKGROUND

Referential measurement tools that are used to measure video quality are typically very sensitive to synchronization and scaling issues. Such tools typically compare frames from the source and captured material sequentially e.g. frame 1 from both source and captured material, then frame 2 for both, and so on. If the captured material is desynchronized even slightly, e.g. frame 1 in the source is actually frame 11 in the captured material, such measurement tools end up performing a comparison of the wrong frames and produce bad results. Further, if the captured material has a different number of frames-per-second, such measurement tools desynchronize quickly.

Some displays at recipient endpoints have a different ratio than the original material. Referential measurement tools are also sensitive to scaling and aspect ratio/resolution conversion. A reliable mechanism that is robust to desynchronization and/or scaling and that provides accurate measurement of the quality of impression of audio/video is valuable in delivery of such content, e.g., via broadcast.

DESCRIPTION

This disclosure describes techniques to reliability measure video quality and audio/video synchronization for broadcast material. The techniques provide referenced (also known as referential) video quality measurement by determining quality based on a comparison of the source material (original broadcast) with what is actually displayed at a receiver device (captured material).

The measurement of audio/video quality using the techniques described herein includes four main parts, which are also illustrated in Fig. 1 below.

1. A ***preprocessing (synchronization) step*** in which the source material and the captured material are synchronized. This is a complex problem, since referential measurement is sensitive to synchronization issues. The described techniques also address scaling problems.
2. A ***referential video quality measurement*** mechanism.
3. A ***referential audio quality measurement*** mechanism.
4. A mechanism to ***measure audio/video synchronization***.

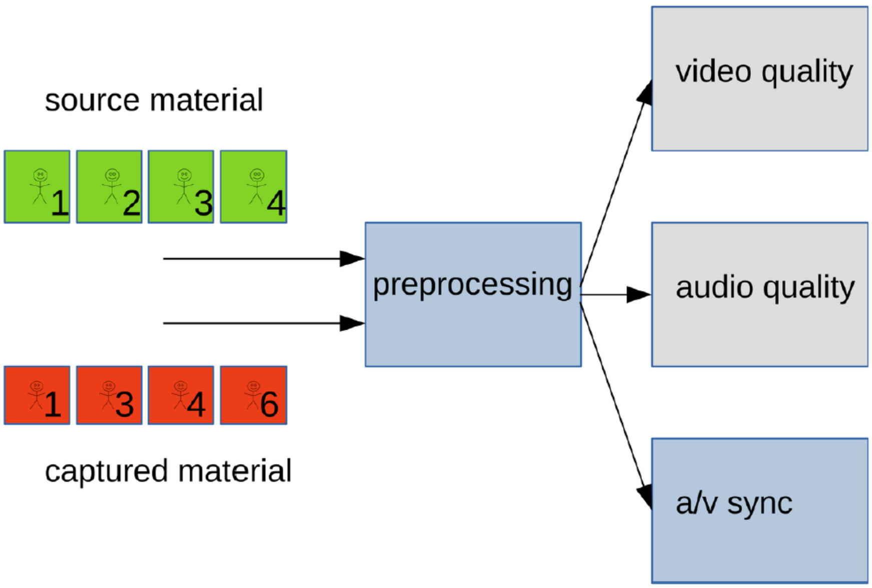

Fig. 1: Solution architecture

Fig. 1 shows the architecture of audio/video quality measurement, per techniques of this disclosure. As shown in Fig. 1, video synchronization and/or descaling is performed prior to the actual video quality measurement to ensure that the resultant measurement is accurate.

Preprocessing: Video descaling

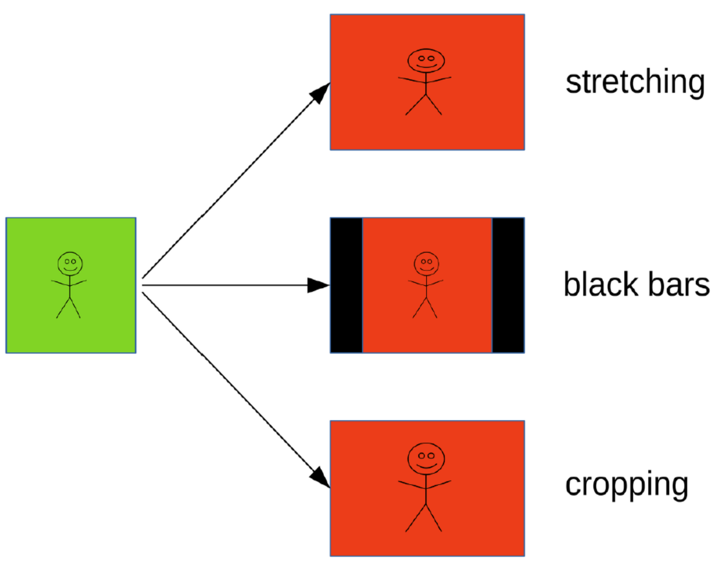

Fig. 2: Ratio adaptation mechanisms (descaling)

Fig. 2 shows ratio adaptation or descaling. A scaling function that the captured material has been subjected to is estimated. For example, some displays have a different ratio than the original material (e.g. 16:10 vs. 16:9). Aspect ratio adaptation is usually performed using 3 main approaches - cropping, adding black bars, and stretching/scaling. Measurement tools are sensitive to each of these modifications.

To estimate the aspect ratio adaptation mechanism, a comparison of the first frame of the source and captured material is performed. A check for stretching is performed by converting the first frame of the captured material to the same resolution as that of the original material. The presence of black bars, if any, is detected and such bars are removed. Further, it is determined if there is any cropping and the video is scaled accordingly. After these checks, a scaling function is established for the entire material, with the assumption that the nature of scaling between the source and captured material does not change during the course of the broadcast, thus providing high performance, since subsequent frames need not be processed for descaling.

Preprocessing: Video synchronization

Video synchronization is performed with the assumption that the reference input includes all the frames. In other words, the source material acts as a source of truth. The problem is then that of matching the frames from the captured material to those in the source material. Differences between the source and captured material can include: (a) multiple copies of the same frame (duplicates); and/or (b) missing or lost frames. The synchronization mechanism works as follows:

First, a frame in the whole captured material is identified that, after application of the scaling function, presents a good match to the first frame in the source material. To calculate the match, a cheap image comparison mechanism is utilized, e.g., PSNR, instead of the more expensive VMAF. For performance, a small initial portion of the captured material, e.g., first 5 seconds, is checked. Assume that frame s0 (first frame of the source material) is matched to a frame cI (frame I of the captured material).

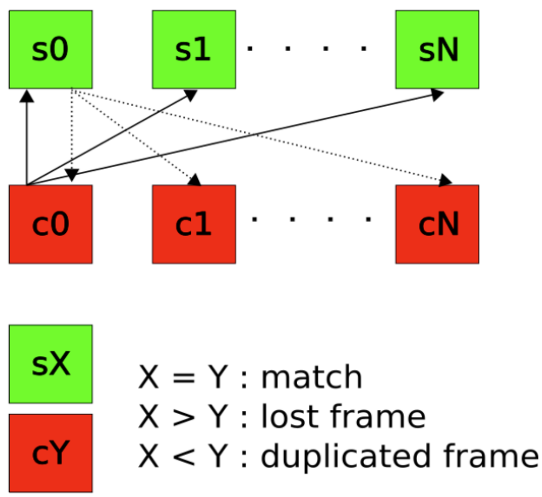

Fig. 3: Synchronization mechanism

Fig. 3 is a visual illustration of the synchronization mechanism, described below. For each frame after ci in the captured material (i in [I, N]), where N is the total number of frames, the following process is performed:

- Apply the scaling function to frame ci.
- For every frame sj in [s0,....]:
  - Calculate PSNR (scaled(s0), scaled(ci))
  - Calculate PSNR(scaled(co), scaled(si))
  - Find best match for c0 and s0 maximizing PSNR:sX and cY
    - If X=Y, it is a match
    - If X>Y, the captured stream has a duplicated frame
    - If X<Y, the captured stream has a missing frame
  - Repeat the loop for s{j+1}

This technique requires reasonably good frame integrity - the captured video under evaluation needs to use single frame compression, and a short group of pictures (GOP) with high bitrate/quality or raw YUV values. Further, the technique may be unsuitable for certain contexts, e.g., videos with sophisticated 3:2 pulldown conversions, where non-integer frames-per-second conversions are solved by mixing frames from the original source.

Measurement of audio and video quality

After synchronization and scaling, video quality measurement is performed using standard techniques, e.g., VMAF as described in [1], and audio quality measurement is performed using standard techniques, e.g., PEAQ as described in [2] or a variation. Both of these techniques require that the captured stream and the reference be perfectly synchronized.

Measurement of audio/video synchronization

To measure audio and video timing, a source video with known properties is used. The known properties include a known audio signal that is played at known intervals and a known video signal that enables identification and counting of individual frames. For example, the known audio signal can be in the form of a short frequency sweep or chirp that is played at the start of every second in the source clip. To determine the position in the captured video where the chirps are being played, a correlation measurement is made that compares the chirp with a downsampled version of the captured audio. The generated video includes a visual clock that can be read automatically and is usable for machine interpretation as well as human checking. Using such a known source video, the number of frames by which the captured video is leading or lagging can be determined compared to the audio for every second.

In actual measurement, video is captured from a suitable source, e.g., HDMI at 60 fps and at a suitable resolution. For example, 4K can be used for quality and a lower resolution be used for audio/video synchronization measurements. For the synchronization measurements, frames can be compressed and put in a container format such as mp4, mkv, etc. and a short group of pictures (GOP) is used.

CONCLUSION

Referential measurement tools that are used to measure video quality are typically very sensitive to synchronization and scaling issues that occur when video captured at a recipient device is out of synchronization with a source video and/or is scaled, e.g., to a different aspect ratio. When the captured video has such differences from the source video, conventional referential measurement techniques produce erroneous results. This disclosure describes techniques to process captured video to account for scaling effects and to synchronize captured video with source video. Referential measurement can then be performed with greater accuracy. A synchronization measurement technique is also described.

REFERENCES

[1] Netflix/vmaf: Perceptual video quality assessment based on multi-method fusion

[2] BS.1387 : Method for objective measurements of perceived audio quality