White Paper

Video Dropouts and the Challenges they Pose to Video Quality Assessment

Video Quality R & D Team,
Digital Media Group, Interra Systems, Inc.

1. ABSTRACT

The media industry is rapidly adopting file-based workflows in all stages of the content lifecycle including transcoding, repurposing, delivery, etc. Additional complexities could be introduced during media transformations, which if not handled properly, could lead to issues in video perceived by the end consumer. While adoption of file-based workflows provided more flexibility with the basic paradigm of file processing, it has also added complexities during media transformations. Improper handling of these complexities can lead to perceived video quality issues for the end consumer. The issues are due to errors caused by media capturing devices, encoding/transcoding devices, editing operations, pre- or post-processing operations, etc. A significant majority of video issues nowadays are due to the loss or alteration in coded or uncoded video information, resulting in the distortion of the spatial and/or temporal characteristics of the video. These distortions in turn manifest themselves as video artefacts, termed hereafter as video dropouts. Detection of such video quality (VQ) issues in the form of dropouts are gaining importance in the workflow quality checking and monitoring space, where the goal is to ensure content integrity, conformance to encoding standards, meta-data fields and most importantly, the perceived quality of the video that is ultimately delivered. This end video quality can certainly be measured and verified using manual checking processes, as was traditionally the case. However, such manual monitoring can be tedious, inconsistent, subjective, and difficult to scale in a media farm.

Automated video quality detection methods are gaining traction over manual inspection as these are more accurate, offer greater consistency, have the ability to handle large amount of video data without loss of accuracy and moreover, can be upgraded easily with changing parameters and standardizations. However, automatic detection of video dropouts is complex and a subject of ongoing research. The source where the artefacts are introduced has a bearing on the way the artefact manifests itself. Automatic detection of the variety of manifestations of video dropouts requires complex algorithmic techniques and is at the heart of a "good QC tool". This paper discusses various kinds of video dropouts, the source of these errors, and the challenges encountered in detection of these errors.

2. INTRODUCTION

The media industry has had to deal with the continuously increasing demand and volume of media data. Furthermore, broadcasters and production houses need to ensure interoperability and robustness for seamless playout and file delivery. With the increased demand and efficient delivery targets in such a complex and dynamic media processing environment, media houses have begun to feel the need to use automated processing tools to detect inconsistencies and thereby ensure quality delivery to the end consumer. An important function of the automated QC tool is to check for any errors introduced within the media files during the various stages of processing of the media.

Typically, video data is characterized by the smooth and steady variation of color and luma values in temporal and spatial domain. Information loss can occur within the encoded video data or during media capture. Such losses lead to observable artefacts called *Video dropouts.* For efficient and robust workflow operation, it is critical to understand the type of errors as well as the stage at which these have occurred. Detection of these errors cannot be achieved by just looking at the encoding parameters or variation in some meta-data information. Also, *No-Reference (NR)* based video quality analysis is the only possibility as the original content may not be available. Therefore, such video dropout detection algorithms should have the ability to automatically detect artefacts, without expecting too many inputs from the user, just as any human observer would be able to do without any *prior* knowledge of the content being viewed. However, it is also practical to accept that the designed automatic detection algorithm will not be perfect and may report false-detection (*false-positives*) and may miss errors (*false-negatives*) along with the correct detections (*true-positives*). The detection performance of a designed algorithm is ranked as high if its false detections and misses are very low, even as its true detections are high. False detections needs to be confirmed before taking any further actions and could involve manual inspection of the media. A high degree of false positives will increase the frequency of manual inspections and will not be cost effective. On the other hand, the misses will be even more critical false then detections, as video content with error is allowed to be delivered without intervention. Ideally, the rate of misses by an automatic detection system should be less than the rate of false positives. Another aspect of automatic detection methods is the throughput or the machine cycles used while analyzing the content. Increasing the accuracy of detection typically goes hand in hand with higher algorithmic complexity. This in turn leads to an added burden on the finite compute resources within the given VQ monitoring system. So in practice, the dropout detection algorithms should implement methods which can give high detection rates while simultaneously optimizing the number of machine cycles being used.

Therefore, a good QC system at the outset must satisfy the following criteria:

1. Low rate of false detection (false positives)
2. Low rate of misses (false negatives)
3. Consistency and accuracy of the detected issues
4. Good performance – This aspect becomes important as the nature of video dropouts is complex and involves multiple algorithms and techniques. For any QC system to be effective, the performance of such checks must be fast. Often, there is a trade-off between accuracy and the speed of detection of the artefacts.

In the following sections, we will look at various types of video dropout issues, their source(s) and manifestation; then we will summarize with a section on the requirements and challenges to detect the dropout issues.

3. VIDEO DROPOUT ISSUES: SOURCES & CLASSIFICATION

A scene captured with high precision and a good quality video camera, produces video (typically) depicting smoothly varying spatial and temporal data. Natural scenes have the property to exhibit steady and smooth changes in the spatial domain. Usually abrupt or irregular changes in the temporal attributes are not present – legitimate exceptions of these could happen at the time of scene/shot changes, introduction of commercial advertisements, etc. So barring these legitimate exceptions, an abrupt change in the spatio-temporal characteristics is easily perceived by the human observer, since human perception/intelligence is well trained by its continuous viewing of natural phenomenon. If the observed video is perceived to not have a steady flow of spatial and/or temporal characteristics, the intelligent human perception system can detect it easily. It may well be that the originally captured video may not have had issues to begin with. Rather, various issues resulting in dropouts could potentially be introduced within the video data as it is modified during various stages of the media workflow. Video dropouts can be in the form of a line or a set of lines, (vertical or horizontal), a small block being misplaced or deformed, or patterns, such as a zigzag or a host of other abrupt artefacts. Machine detection of this plethora of changes is a complex subject requiring complex algorithms. Also, the characteristics of video dropout issues will vary depending on the stage at which these are produced e.g., content capture, compression, delivery, storage or reproduction. The following sections describe the high level classification of video dropouts, their properties, and possible sources of errors.

3.1 ANALOGUE DROPOUTS

Prior to digitized video, there was an era when the video and audio information was stored on *magnetic tapes* in the form of analog video signals.

Troves of legacy yet valuable media assets have been traditionally archived in this manner. With the recent adoption of file-based storage, there was an urgent need to convert these important media assets to digital files. This required an analog to digital conversion device, capable of interpreting stored analog formatted data and converting it to its digitised form. Basically, the stored analog signals can be interpreted as pattern of variations in the voltage levels guiding the formation of video frames, fields, and pixels. Due to mishandling, ageing, and improper maintenance of these tapes, the stored analog data could be adversely affected, leading to inconsistencies in the received voltage levels by the conversion device. The resulting video artefacts in this way are collectively termed here as Analog dropouts. Some of these analog dropouts are blotches, scratches, miss-tracking, head clog, skew error, horizontal/vertical sync pulse loss, etc. The following sections will discuss some of the most frequently observed types of analog video dropouts.

a) Horizontal / Vertical Sync Pulse Loss

The video frame consists of multiple horizontal scan lines spread across the vertical resolution. A specific voltage level exists at the end of each scan line indicating its end and start of the next scan line. Variation in the voltage level (due to noise) will result into shift in content lines towards left/right perceptually viewed as horizontal lines. This is shown in the snapshot below (Figure 1 (a)).

Vertical sync pulse is another such voltage level controlling the start/end of a new video frame. A deviation in this voltage level will disturb the start of formation of next frame. Vertical sync pulse loss will result into merging of two adjoining frames at the frame boundary (Figure 1 (b)).

Figure 1 (a) Horizontal Sync Pulse Loss (b) Vertical Sync Pulse Loss

b) Skew Error

Physical changes in the magnetic tape could be introduced due to continuous expansion or shrinkage of the tape surface over time. The ***recorded track alignment with respect to the playback head is in turn*** affected by these changes. During playback/recording, this loss of alignment could result in shift of a band of scan lines at the top/bottom of the picture. This horizontal shifted portion of the video frame at top or bottom part is termed as Skew Error (Figure 2).

Figure 2: Skew Error

c) Line Repetition Error

The digital to analog conversion device receives the video data in the form of scan lines. Buffers storing each scan line data are updated regularly after each sample and hold duration. Due to issues in controlling signals, the current scan line is not captured and is replaced by the previously fetched scan line. This error in the control signal continues for a while and the same is manifested as a repeated set of horizontal content lines. This type of artefact is shown in Figure 3.

Figure 3: Line Repetition Error

d) Blotches

Blotches are due to the presence of dirt/sparkle on the surface of magnetic tape. Dirt/blotches disrupt the reception of signals during video data capture. The area for which the data is not received appears as white or black spots. A snapshot of the video frame with blotches is shown in Figure 4.

Figure 4: Blotches

3.2 DIGITAL DROPOUTS

As a consequence of the digital revolution, media (audio and video) in its digital form started its storage in magnetic tapes instead of its analog counterpart. Digital video has adopted many evolving digital formats for storing the video data in uncompressed or compressed forms. Some of the uncompressed were Betacam, D-1 up to D-5, etc. For compressed versions, DV standard was widely used and its more robust forms were DVCAM and DVCPRO. DV is based on *lossy* video compression using only an intra-frame compression scheme. The spatial compression method was based upon Discrete Cosine Transform (DCT).

The digitized video data is compressed using video coding standards, such as MPEG2/4, H.264/5, VC-1, VP6/9, etc., JPEG2000 and some proprietary formats, such as Apple ProRes. Block-based coding methods typically use DCT coefficients for representing variation in image data, while the temporal variations are represented using motion vectors. The quantized DCT coefficients and resulting motion vectors are entropy coded, based on different kinds of algorithms defined by the underlying coding standards.

Video compression removes information redundancy within a given piece of content so that in its compressed form it can be efficiently stored, transmitted, and received in bandwidth limited environments. A loss or alteration in a single bit or a group of received bits will make the underlying video de-compressor (decoder) lose its synchronization with respect to the coding standard being followed. It is difficult for the decoder to figure out the error event and in the event of an error it continues to parse and decode the data. Each decoding step will then output irrelevant and inconsistent decoded output leading to video issues, and some of these manifest themselves in the form of digital dropouts. Unlike analog video errors, digital errors typically follow some defined geometries in the frame. For block-based coding schemes, the errors will be in the form of rectangular blocks or horizontal slices. In the case of a more complicated error, it may happen that the block boundaries are slightly deformed also. The following sections outline some of the commonly observed digital video dropouts.

a) Block Errors

Blocks of video data in block-based encoding methods are typically encoded using DCT coefficients. In the case of intra coded pictures, the content itself is coded using DCT coefficients. Specifically, in the case of DV video stored on digital tapes, the blocks are coded using intra coding. When the DV data from tapes are being converted to files, there is a possibility of bit error(s) introduced in the received data by the conversion device. Such bit errors (tape hits) can affect decoding of the DC and/or AC coefficients. If errors are introduced within the AC coefficients, then the decoded blocks will be with high-frequency patterns similar to DCT basis functions. This type of defect is termed here as a high-frequency block error. On the other hand, errors within a DC coefficient can result in the formation of blocks different in intensity from the neighboring blocks. Both of these types of errors are depicted in Figure 5.

Figure 5: High Frequency block Errors

b) Stuck/Misplaced Blocks

The temporal information in a video is represented by motion vectors along with block residuals encoded using DCT coefficients. The decoder will output either 0 (no-motion) or incorrect valued motion vectors if there exists an error in encoded data representing motion vectors. If the motion vector is zero, the blocks actually in motion will not be displaced. If the motion vectors are not correct, the blocks will not be placed at their actual positions in the video frame. These types of errors are depicted in Figure 6.

Figure 6: Stuck / No-motion Blocks

c) *Non-decoded Data / Concealment Errors*

The underlying decoding engine can sense and will try to adopt various concealment methods to rectify the error. One of the simplest concealment strategies is copying the correctly decoded data in previous frame(s) at the co-located position. This method works quite well for video with no motion or slow motion but the same fails in the case of video with a considerable amount of object motion. The concealed data will be in the form of square blocks or a horizontal strip. Figure 7 depicts the concealment error for an error in slice.

If decoder is not able to identify the error, decoder will not be able to decode and rectify the same and due to this the video data is not decoded. The non-decoded area in the frame will be filled with default/undefined values leading to rectangular blocks being filled with a green/gray color. If the data is not concealed, the blocks, or slices or part of the frame will be filled with undefined values.

Figure 7: Concealment Error

4. VIDEO DROPOUT DETECTION: REQUIREMENTS AND CHALLENGES

Previous sections have talked about various kinds of video dropouts. Detection of these dropouts is important so as to identify and rectify any kind of issue during the various stages of media processing workflows. Based on the various types of video dropouts observed and for efficient processing within the workflow, following are the high-level requirements for detection methods to report various video dropout issues,

1. The dropout detection method(s) need to follow the *No Reference* principles, as original data might not be available in most of the scenarios at the Quality Checking site.
2. The detection method should be fully automatic, i.e. having a minimum set of external inputs/parameters, (- preferably none).
3. The detection accuracy should be high enough, i.e. fewer number of false positives. At the same time, it should avoid missing actual errors (false negatives).
4. The analysis time by these methods should not be high, so as to optimize the throughput of the overall media workflow.
5. Finally, the accuracy of these methods should be content independent. For example, it should be able to handle natural video, cartoon/animated content, text/graphics within content and so on, efficiently.

The detection methods/algorithms use image processing, computer vision, pattern recognition and machine learning techniques. Looking at the above set of requirements, it is important to apply focused and appropriately selected techniques within the algorithms so as to minimize false detections and misses while at the same time minimizing adverse performance impaction.

Some of the widely known and used video processing concepts are Edge detection, Spatial Frequency analysis, temporal consistency analysis using motion estimation/compensation, etc. The goal behind all these techniques is to replicate aspects of the human vision system (HVS); i.e. to have perceivable change in response to the spatial and temporal irregularities present in the video. If we talk about the spatial irregularities, then spatial details of video content are measured using spatial analysis techniques, such as edge detection, frequency analysis, morphological operations, etc. Similarly, temporal characteristics of the observed video are analysed using Motion estimation / compensation techniques, object tracking, scene change detection, and so on. Appropriate selection and combination of these techniques works quite well for most of the obvious errors, but may not detect all errors, because of the various assumptions followed during implementation.

The first fallback of these spatial techniques is the selection of a threshold to differentiate an abnormal response from a normal one. For example, an edge threshold strategy to select an edge of an erroneous block will fail if the block boundaries are not so visible (mostly in the case of old videos). HVS inherently adjusts these thresholds depending upon the content type and quality. However, it is very difficult to consistently adjust these thresholds in an automated system. Another major problem is the various types of noise present in video. HVS is very good at differentiating noise from information content. We can perceive minor glitches even in the presence of high noise, but replication of the same in an automated system is very complex and difficult, given the high level requirements.

For temporal techniques, a major issue is the accurate modeling of very fast motion or affine (rotational, shear) motion. HVS can track an object over time irrespective of the changes in shape and visibility of that object, but it is very difficult for an automated system to adjust to so many transformation possibilities. Similar to spatial techniques, it is very difficult to say how many irregularities in motion tracking can be attributed to an error. This discrepancy in adjustment of threshold for detection of irregularity will either lead to false detections or missing actual errors. For example, unlike natural videos, cartoon content exhibits non-linear and complex temporal and spatial characteristics that become similar to video dropout due to temporal discontinuities.

As another example, a piece of animated content may consist of highly visible rectangular blocks changing their shape or color. These changes will be detected as false positives. It is difficult to model the temporal behavior of an animated object rapidly changing its position across frame boundaries. Artistic effects that are used to enhance the content may also introduce discontinuities being detected as false dropouts. A typical editing example of this type is the formation of an artistic object created by merging smaller subcomponents. Another scenario is introduction of burnt-in subtitles forming an extra layer over the continuous spatial information. For example, a colored rectangular block over the letter 'í' will be identified falsely as a non-decoded block at the frame location where the text has just started. A classic example of false detection of block error dropouts is in the case of '*Confetti*' (figure 8).

In such cases, human intelligence helps in deciding whether the spatial and temporal irregularities are desired or not. The sudden appearance or disappearance of objects can be either magical or a case of occlusion (desired) or due to some error (not desired). Similarly, irregular changes in the shape of an object can be attributed to its natural characteristics (e.g., fog, rain drops, etc.) or an error. Although it is possible to detect such possibilities using specific techniques, one has to accept that it is not possible to put various heuristics or take steps to handle each and every complex case of false detection or a miss. Also, advanced techniques or algorithms, using high machine time, cannot be employed here.

Figure 8: 'Confetti' Sequence

Based on the available set of techniques, the requirements for detection methods, and possible cases of false detection and misses, we can say that while there have been notable successes in alleviating dropouts, there exist several challenges in the technology realm for the detection of dropouts.

5. INTERRA'S BATON CONTENT VERIFICATION SYSTEM FOR VIDEO DROPOUTS

Baton is a highly advanced Auto QC system with deep video and audio quality checks. Baton supports the detection of a large variety of video dropout, both analog and digital issues. The detection algorithms deploy appropriately selected and *patented* advanced Image processing/computer vision techniques. Supported video dropout checks are designed and targeted for various stages of media workflow. For digital workflow, Baton supports the detection of luma/chroma block errors, data corruption, and concealment errors. For the analog workflow, it supports detection of line repetition errors, horizontal/vertical sync loss, skew error, artefacts at frame boundaries, and line errors. Apart from these, detection of some generic errors, such as unwanted frames, field/frame corruption, and chroma dropouts, are also supported.

The algorithms have been fine-tuned and performance optimized to detect the issues in real-time for HD content. Additionally, the techniques have been fine-tuned to minimize false positives and false negatives, while retaining accuracy of results.

Baton detects the largest number of video dropouts in a variety of content compared to any other QC tool in the market, but it does not stop there. The Baton VQ team within Interra is doing research and development for a broader coverage of issues, and improvement in accurate detection and performance of the existing ones. As the industry moves into the 4K era along with newer codecs, such as HEVC, there are new types of issues cropping up. Interra's Baton solution is continuously being enhanced to incorporate these changes.

6. CONCLUSION

Complex video quality issues in the form of video dropouts are being observed while creating, transforming, and processing the media content. Accurate detection of the video dropout issues are certainly challenging, with an ever increasing reliance on automated QC solutions for no-reference assessment of video quality. Detecting these complex and varied artefacts needs to be handled by a good QC tool. Most QC tools today are at an early stage of detecting these artefacts accurately. QC solutions are focused on research and development of advanced techniques that result in accurate and robust detection of varying kinds of video dropouts. In addition, these solutions are continuously upgrading and enhancing their techniques to handle newer and more challenging types of video dropouts.

Interra Systems, Inc.
1601 S. De Anza Boulevard, Suite 212
Cupertino, CA 95014
USA

