**3GPP TSG-SA WG4 #126 S4-231659 r3**

**Chicago, USA, 13th-17th November 2023**

**Source: Dolby Laboratories Inc. (Rapporteur)**

**Title: [FS\_FGS] Some Updates on Film Grain Synthesis Testing**

**Version: 0.3**

**Agenda Item: 3.7**

**Document for: Discussion**

# 2 Proposed Scenario and Evaluation Framework

In order to evaluate the performance and justify potential benefits a qualitative and quantitative analysis on what would be the costs and benefits would be needed.

|  |  |  |
| --- | --- | --- |
| # | Parameter | Definition |
| 1 | Scenario name | Film Grain Synthesis Scenario |
| 2 | Motivation for the scenario | Study, test and analyze the effect of film grain synthesis applied to varied test scenarios encoded with H.265/HEVC, in order to ascertain whether there is benefit in (1) preservation of artistic intent and/or (2) masking artifacts / saving bitrate, while maintaining visual acuity, etc. |
| 3 | Description of the scenario | Film grain can be unavoidably lost by a conventional video compression workflow as shown in Figure 1. High frequency grain in the source content () can be either completely or partially lost due to quantization in adaptive streaming or broadcast applications, in which case the resulting decoded output could be void of the artistic effect of grain intended in the source. Even if grain is partially preserved at high bitrates, overall coding efficiency is poor as the grain is not spatially or temporally correlated, which can negatively impact the coding characteristics of the scene.  In the case of the Film Grain Preservation Workflow, as per Figure 2, the source video () is input to a denoising process that outputs a video sequence from which noise or film grain is attenuated or removed. A film grain parameterization process then compares the source and denoised videos to determine film grain model parameter values, which relate to the variance, spatial frequency characteristics, colour correlation, and other statistical characteristics of the film grain. The process of denoising followed by the film grain model parameter estimation is commonly referred to as the film grain analysis process. After these processes are performed, the denoised video is then encoded and the film grain model parameter values are either signalled in the coded bitstream () or provided to the decoder by some external means.  On the decoder side, the film grain model parameter values are parsed and input to a film grain synthesis process that generates simulated film grain and blends the grain with the decoded video () to output decoded video with simulated film grain (). |
| 4 | Supporting companies and 3GPP members |  |
| 5 | Source format properties   1. Spatial resolutions 2. Chroma Format 3. Chroma Subsampling 4. Aspect ratios 5. Frame rates 6. Colour space formats 7. Transfer Characteristics 8. Bit depth 9. Other signal properties | 1920x1080, 3840x2160  YUV  420P  16: 9  24, 29.97, 30 ,59.94 ,60  ITU-R BT.709 or BT.2020  BT.709, PQ  8, 10 |
| 6 | Encoding and decoding constraints and settings: Typical encoding constraints and settings such as   1. Relevant Codec and Codec Profile/Levels according to TS26.116 and TS26.511. 2. Random access frequency 3. Error resiliency requirements 4. Bitrates and quality requirements 5. Bitrate parameters (CBR, VBR, CAE, HRD parameters) 6. ABR encoding requirements (switching frequency, etc.) 7. Latency requirements and specific encoding settings 8. Encoding context: real-time encoding, on device encoding, cloud-based encoding, offline encoding, etc. 9. Required decoding capabilities | HEVC/H.265  Main Profile & Main10 Profile, Level 5.3  HM version used:  HM 18.0  <https://vcgit.hhi.fraunhofer.de/jvet/HM/-/tags/HM-18.0>  Detailed config parameters embedded in cfg file below |
| 7 | Performance Metrics and Requirements   1. A clear definition on how the performance needs to be evaluated including metrics, etc addressing the main KPIs of the scenario. 2. Objective measures such as PSNR, VMAF, etc, may be used. 3. Subjective evaluation is not excluded and may be done, but needs commitment | Subjective Metrics   1. MOS (pairwise method with hidden reference) 2. Relative Grading (A v/s B method) 3. JND (Just Noticeable Difference) method |
| 8 | Interoperability Considerations for the application   1. Streaming with DASH/HLS/CMAF 2. RTP based delivery | The following media formats have been tested and validated for FGS interoperability.   1. MP4 2. MPEG-DASH 3. HLS |
| 9 | Test Sequences   1. A set of selected test sequences that are provided by the proponents in order to do the evaluation. They should cover a set of source format properties | Preserving Artistic Intent  Scenario FHD   1. BQTerrace 2. OldTownCross 3. InToTree   Scenario 4K-TV   1. Crowdrun 2. Meridian (from existing 3GPP dataset)   It has been reported in S4-231787 that there is the new “ground truth-based” film grain material available, that was generated using the “Tears of Steel” open movie and new content captured by Apple. Such material can be used for investigating methods on film grain analysis, synthesis, and quality evaluation among others.  (Proposal is to put this text to TR26.955 Sec.6.7.3) |
| 10 | Detailed test conditions:   1. Provides a proposal for detailed test conditions, for example based on a reference software together with the sequences and configuration parameters. | After denoising the content, HEVC software encoders and decoders with film grain analysis and synthesis support, respectively, are used to assess the performance of film grain modelling:   * **Denoiser**: The ffmpeg ‘nlmeans’ filter is used to denoise the source content before encoding.   Sample Command  ./ffmpeg -f rawvideo -pix\_fmt yuv420p -s:v 1920x1080 -I ../InToTree\_1080p50.yuv -vf nlmeans=3.5:5:3:9:7 InToTree\_nlmeans\_3.5\_5\_3\_9\_7.yuv;   * **Film grain analysis algorithm**: The frequency filtering analysis algorithm in the HM-18.0 reference encoder is used to model the grain parameters to be inserted in the FGC SEI message. * **Film grain synthesis algorithm:** The SMPTE RDD5 based Frequency filtering grain synthesis algorithm in the HM-18.0 reference decoder is used for blending the synthesized film grain with the decoded video * **Browser player with FGS support**: For Subjective evaluation, Ittiam HEVC FGS web browser player is used as it supports SMPTE RDD5 based frequency film grain synthesis using the FGC SEI messages embedded in HEVC bitstream.   [https://dem](C:\\Users\\alexismt\\Library\\Containers\\com.apple.mail\\Data\\Library\\Mail Downloads\\B1B576F5-673D-4315-A962-5C2C77A35F62\\https:\\dem)o.ittiam.com/demo/i265\_fgs/ |
| 11 | External Performance data |  |
| 12 | Additional Information |  |

An independent scenario for FGS may be defined.

# 4 Communication with JVET

It is agreed to send an LS to JVET informing them of our intent to evaluate the impact of FGS on our 3GPP scenarios listed in this study.

This LS is intended to request information on their ongoing activity with AG5 on similar evaluation in order to see if they can meet our timeline.

This clause is a placeholder for information we want to share with JVET and our requests.

* Inform JVET on the study item on Film Grain Synthesis (list the objectives or point to/attach the SID)
* Inform JVET on our timeline justified by the intent to meet release18 for normative work.
* Explain to JVET the type of evaluations that we will conduct in SA4
* Ask JVET if they have similar activities planned/ongoing/done and see if they can share results (or by when).

An LS is expected to be sent before the next JVET meeting (11th October 2023).

Update: The LS has been sent (S4-231587) to MPEG/JVET (m65645) and a response statement is generated from the October MPEG/JVET meeting (S4-231648, 3GPP SA4 #126).

ANNEX: FGS Testing Methodology and Results

**FGS workflow for preserving artistic intent.**

Film grain can be unavoidably lost by conventional video compression workflows, as shown in Figure 1. The high frequency grain in the source content () can be either completely or partially lost due to quantization in adaptive streaming or broadcast applications, and the resulting decoded output may be void of the artistic effect of the grain intended in source. Even if grain is partially preserved at high bitrates, overall coding efficiency may be poor as film grain is not spatially or temporally correlated, impacting the coding characteristics of the scene and a significant number of bits may be needed to achieve an acceptable coding and at the same time preserve the film grain.

Figure 1 HEVC workflow without film grain

HEVC Encoder

Video Stream

**Film Grain Analysis Disabled**

HEVC Decoder

The following notations are used for the film grain preservation workflow depicted in Figure 2:

* is the source video (uncompressed)
* is the denoised source video.
* is the compressed bitstream of the denoised video.
* is the decoded denoised video.

is the grain synthesized video, using, for example, the FGC SEI message parameters.

Figure HEVC film grain modeling and synthesis workflow

De-Noising

HEVC Encoder

Film Grain Analysis

VideoStream

**Film Grain Analysis Enabled**

Difference Noise

Film Grain SEI

HEVC Decoder

Conventional codec

Film grain framework modules

*Encoder side*

*Decoder side*

Film Grain Synthesis

**Quality Evaluation Methods for FGS**

To assess the benefits of grain characterization and synthesis for preserving artistic intent two independent subjective quality verification tests were recently conducted. The first test used the MOS scoring method and used 14 test subjects/viewers, while the second test used the A-B Preference method and used 11 test subjects. It should be noted that it is commonly recommended that for a subjective test to be reliable a higher number of test subjects, i.e. >=25. In addition, encodings were performed without the consideration of subjective optimization tools or quantization matrices, which can have a significant impact especially in the presence of film grain in the content. Therefore, one should be cautious in making any conclusions based on these tests given their limitations.

**Subjective Performance Evaluation Methods**

* MOS Scoring Method
  + ITU-T\_Rec. P.190 based subjective performance evaluation method (pair-wise testing with hidden reference). This method is a category judgement where the test sequences are presented one at a time and are rated independently on a category scale (1 to 10). The present test procedure must include a reference version of each test sequence shown as any other test stimulus.

| **Test Site** | **On-site** |
| --- | --- |
| **Display, size, connection  (resolution setting)** | Samsung 65” S95B, HDMI (3840×2160) |
| **Viewing distance** | 2 viewers sitting at 1.5H, 1 viewer standing at 1.6H |
| **Viewing angle** | ±75°, 90° (at screen center) |
| **Total number of viewers** | 14 (5 female, 9 male) |

* A-B Preference Method
  + The test sequences were evaluated using a 5-score scale, the 5 scores (A>>B, A>B, A==B, A<B, A<<B) were mapped to the values (3, 1, 0, -1, -3) and the random A-B assignment was reverted. In the comparison of the compressed sequences with and without FGS, positive numbers indicate that the FGS version was scored higher.

| **Test Site** | **On-site** |
| --- | --- |
| **Display, size, connection  (resolution setting)** | Samsung 65” S95B, HDMI (3840×2160) |
| **Viewing distance** | 2 viewers sitting at 1.5H, 1 views standing at 1.6H |
| **Viewing angle** | ±75°, 90° (at screen center) |
| **Total number of viewers** | 11 (3 female, 8 male) |

**Scenario FHD Test Results (Subjective)**

**Scenario 4K-TV Test Results (Subjective)**

**FGS workflow for masking coding artifacts**

Film grain technology can also be used as a mechanism for masking artifacts that may have been introduced during the coding process. This process could even be used on content that did not originally contain film grain. In particular, film grain could potentially help in improving the subjective quality by subtly masking artifacts such as blockiness, banding, and blurred textures at low bitrates.

Figure 3 HEVC workflow for film grain estimation for masking coding artifacts

HEVC Encoder

Film Grain Estimation

VideoStream

**Film Grain Estimation Workflow**

HEVC Decoder

*y*

Conventional codec

Film grain framework modules

*Encoder side*

*Decoder side*

Film Grain Synthesis

FG parameter Insertion

The following notations are used for the film grain estimation for masking coding artifacts workflow that is depicted in Figure 3:

* is the source video (uncompressed)
* is the compressed bitstream that also includes some Film Grain parameters, e.g. based on the Film Grain Characteristics SEI message.
* is the decoded video.
* is the grain synthesized video based on the film grain parameters included in the bitstream.

As in the previous test, no subjective optimization was used when encoding the test content.

**Subjective Test Results A|B Method with DMOS**

**Just Noticeable Difference (JND) Subjective Test Method**

In experimental psychology, the term Just Noticeable Difference (JND) means the amount of something that must be changed in order for a difference to be noticeable, such that it may be detected at least 50% of the time. This is applied in various sensation and perception studies. In the context of the film grain, this JND method may be used for either to test for artistic intent or for masking visual artifacts.

The JND objective test measures these two things depending on the use case via side by side comparison:

1. Preservation of artistic intent: measure the difference between the clean reference and the reference video with added FGS,
2. Applying film grain to mask coding artifacts: find the highest compression level, adjusted by QP, at which point the coding artifacts become noticeable with and without added FGS.

|  |  |  |
| --- | --- | --- |
|  | Without film grain synthesis | With film grain synthesis |
| Reference | true lossless quantization | true lossless quantization with film grain synthesis |
| Test | compressed at various QPs in [18, 42] | compressed at various QPs in [18, 42] with film grain synthesis |

* 1. If the video file size is not limited, just use the original reference video. If the video size needs to be small, use QP=4 for the reference video.
  2. Keep the QP the same between the video with / without FGS applied.

The steps of this JND test method is as follows:

1. Initial stimulus: set at lowest quality. Initial ΔQP set at 8.
2. When a subject provides a correct response (i.e., there is difference between the two videos or the reference video looks better than the test video), the QP value is decreased by ΔQP.
3. When a subject makes an incorrect response (i.e., no difference between the two videos or the test video looks better than the reference video), the QP value is increased by ΔQP.
4. How we determine the value of ΔQP:
   1. The initial ΔQP is set to 8.
   2. Keep the previous ΔQP when the same response follows an earlier response. That is, an incorrect response after earlier incorrect response, or a correct response after a previous correct response occurs.
   3. ΔQP is reduced to half of the previous ΔQP, when an incorrect response occurs after a correct response or vice versa.
   4. The chart below illustrates an example of adaptive QP adjustment according to a subject’s response.

A graph with numbers and lines

Description automatically generated

* 1. This process stops when it reaches either of the stopping conditions below:
     1. 4 times of upward and downward pattern then arrive at QP=1 step size,
     2. When the number of responses reach the maximum trial number of responses. (This maximum number is TBD)
  2. The JND threshold value is determined as the last correct response. For example in the chart above the JND threashold value corresponds to the last point in the chart.
  3. Prior to the above test, a training session is needed; the best set of examples showing significant difference and no difference using all test sequences will be presented to the test subject.
     1. 2 different types of comparison
        1. Both the reference and the test are FG-free. The other set is for the reference and the text material both to have FGS added.
        2. There is no direct side by side comparison between the clean (no FG) and the video with FG.
        3. The order of the test videos is randomized. The playback time for each sample video is 10 seconds.

1. Viewing test conditions:

| 1. **Test Site** | **On-site  (for 4K resolution)** |
| --- | --- |
| **Display, size, connection  (resolution setting)** | Samsung 65” S95B, HDMI (3840×2160) |
| **Viewing distance** | 1 viewer sitting at 1.5H |
| **Viewing angle** | ±75°, 90° (at screen center) |
| **Total number of viewers** | 12-15 |

| **Test Site** | **On-site  (for 1080p resolution)** |
| --- | --- |
| **Display, size, connection  (resolution setting)** | DELL 29” TV/ Monitor, HDMI (1920x1080) |
| **Viewing distance** | 1 viewer sitting at 1.5H |
| **Viewing angle** | ±75°, 90° (at screen center) |
| **Total number of viewers** | 12-15 |

(Rapporteur’s note: the JND test results will be added after SA4 #126 and shared at a forthcoming Video SWG AhG meeting. The intent is to use some of the FG content that Apple has kindly made available.)