Work Plan: Evaluation of short 360° sequences

VQEG Meeting, Berlin, 06/03/2019

A bit of context…

- Decided in audio in July: Consider use case of **360-degree video for joint work**.

- Started working on a **test plan**:
  https://drive.google.com/drive/folders/1UWpGRqfo4ILF48hjmmdtl01Pqdoirx06?usp=sharing

- Objective: Recommend methodology for subjective assessment of quality for 360-degree videos that valid to evaluate "typical" degradations, e.g., coding artifacts (homogeneous and hetereogeneus), stitching artifacts, etc.

- We started defining it @Google/Youtube meeting with SoA presentations and discussions of different aspects covered in the test plan:
  - **Monitoring user behavior,**
  - **Methodologies** for subjective quality evaluation of **short and long** 360-degree videos,
  - Approaches for assessing the **simulator sickness**
  - Evaluation of **immersion and presence** in VR.

Quality assessment for 360˚ sequences

- **Decisions** on last meeting: Test the effects of...
  - Length of sequences
  - ACR vs DCR (double stimulus)
  - Influence of HMD or display HW (e.g., desktop, normal screens, random viewports, etc.)
  - Simulator sickness questionnaire or other relevant questionnaires

- Two-fold work plan considering:
  - Short sequences.
  - Long sequences.

Quality assessment for short 360° videos

- **SRCs --** Review datasets and get more content: At least 10 SRCs
  - **Resolution**: at least 4K
  - **Length** of content: Consider at least 30s
  - Looking for **uncompressed/high quality content**… and with **CC licensing**… Difficult to consider videos on Youtube
  - Possible set:
    - 3 videos form Nokia
    - 3 videos from TU Ilmenau
    - From datasets with "professional" content and CC licensing (awaiting confirmation by email):
      - V-Sense dataset: 8 videos, equi-rectangular.
        - https://v-sense.scss.tcd.ie/research/3dof/directors-cut-research/
      - ImmersiaTV dataset: Several scenes, unstitched content.
        - http://www.immersiatv.eu/project-outcomes/datasets/

Quality assessment for short 360˚ videos

- **SRC1** – Nokia. 4m20s. Audio. Stitching artifacts. Raw/high quality version available.

Quality assessment for short 360˚ videos

- **SRC2 –** Nokia. 5m40s. **No** audio. Artifacts due to rain. Raw/high quality version available.

Quality assessment for short 360˚ videos

- **SRC3 –** Nokia. 3m11s. Audio. Slight stitching artifacts. Raw/high quality version available.

Quality assessment for short 360˚ videos

- **SRC4 –** TUI (CheerLeading). 30s. Audio. Raw/high quality version available.

Quality assessment for short 360˚ videos

- **SRC5 –** TUI (Brazil). 30s. Audio. Raw/high quality version available.

Quality assessment for short 360˚ videos

- **SRC6 –** TUI (Thieves at Lake). 30s. Audio. Raw/high quality version available.

Quality assessment for short 360° videos

- **SRC7 –** Vsense (Vaude). 4096x2048, 29,97fps. 2m25s. Audio **in German**. High quality version available (H264, 50Mbps). Documentary, indoor and outdoor **short shots**.
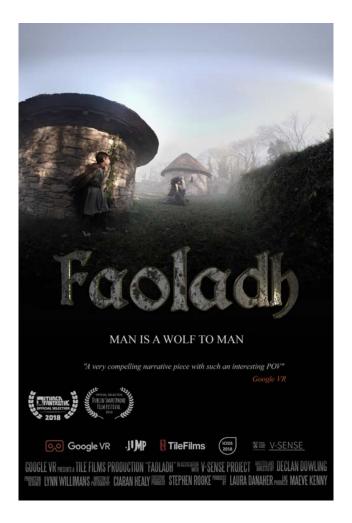
Quality assessment for short 360˚ videos

- **SRC8 –** Vsense (Luther). 4096x2048, 29,97fps. 4m25s. Audio **in English**. High quality version available (H264, 50Mbps) . Documentary (animation character), various **short shots** indoor and outdoor.

Quality assessment for short 360˚ videos

- **SRC9 –**

(awaiting confirmation by email for more videos)

Quality assessment for short 360° videos

- **SRC10 –** ImmersiaTV (Scene 5). ImmersiaTV (Scene 8). Recorded: GoPro H3Pro6 Rigs. Audio?. Raw/high quality version available. **Unstitched: To Check!**

Quality assessment for short 360° videos

- **SRC11 –** ImmersiaTV (Scene 8). Recorded: GoPro H3Pro6 Rigs. Audio?. Raw/high quality version available. **Unstitched: To Check!**

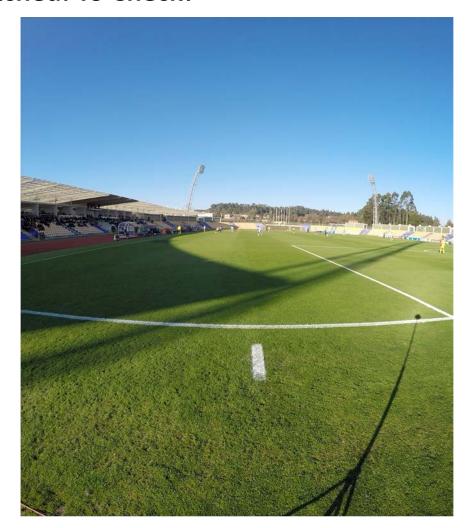# Quality assessment for short 360˚ videos

- **SRC12 –** ImmersiaTV (Scene 9). Recorded: GoPro H3Pro6 Rigs. Audio?. Raw/high quality version available. **Unstitched: To Check!**

Quality assessment for short 360˚ videos

- **SRC13 –** ImmersiaTV (Media 5). Recorded: GoPro H3Pro6 Rigs. Audio?. Raw/high quality version available. **Unstitched: To Check!**

Quality assessment for short 360˚ videos

VQEG

- **SRC14/15 –** ImmersiaTV (MEDIA_T05_1). Recorded: 7 camera GOPRO rig Hero4. Audio?. Raw/high quality version available. **Unstitched: To Check!**

Quality assessment for short 360˚ videos

- **SRC16/17 –** ImmersiaTV (MEDIA_T05_2). 9 minutes. Recorded: 7 camera GOPRO rig Hero4. Audio?. Raw/high quality version available. **Unstitched: To Check!**

Quality assessment for short 360˚ videos

- **SRC18 –** ImmersiaTV (MEDIA T07). 10 minutes. Recorded: Orah4i and AZilPixStudio. Audio. Raw/high quality version available. **Unstitched: To Check!**

Quality assessment for short 360˚ videos

Testing interfaces

- Same tool to run the tests in all labs:
  - Play test videos → ACR and DCR.
  - Show rating interface and collect ratings.
  - Record head pose (and eye gaze).
  - Compatible with various HMDs.
- Some existing alternatives to consider as starting points to develop:
  - AVTrack360 from TU Ilmenau: Python, working with HRC Vive and Oculus, no rating interfce.
  - Tool from Nokia: Samsung GearVR, Google Daydream
  - UdN: Unity3D, OpenVR (tested on HTC Vive), need to adapt rating interface (used in emotion test, SAM scale), record of eye-gaze.

Quality assessment for short 360° videos

VQEG

- **HRCs –**
  - **Coding:**
    - Uniform coding: Homogeneous degradations
      - Selection of QPs for each SRC according to pretests.
    - Non-uniform coding (e.g., tile-based) -> heterogeneous degradations
      - Number of tiles / size of tiles
      - Abrupt vs. smooth quality changes between tiles?
      - Using open source tools
        - Reference SWs
        - VP9 --> with support from Youtube for setting parameters
  - **Stitching:** Add one/two SRC containing stitching artifacts in the test set.
  - Different **projections**: Equirectangular and cubemap
    - Not compare for all HRCs.

  - Duration of a whole test session
    - 10 SRCs x (8 HRCs X 2) x 20 sec < 60 min approx.

	HRCS		
1	Reference		
2	Coding	Homogeneous	QP1
3	Coding	Homogeneous	QP2
4	Coding	Homogeneous	QP3
5	Coding	Hetereogeneous	N tiles - abrupt change
6	Coding	Hetereogeneous	M -tiles abrupt change
7	Coding	Hetereogeneous	N -tiles smooth change
8	Coding	Hetereogeneous	M tiles smooth changes

Quality assessment for short 360˚ videos

Other aspects to evaluate and compare: Not all HRCs!

- Effects of **duration**: Identify minimum duration required to evaluate the considered degradations
  - Consider test clips of 30s, 20s and 10s
  - The same observer does not see the same clip with different durations → How many observers?
  - Randomizing so each observer sees clips of different duration
- Effect of **methodology**: ACR vs DCR (double stimulus)
  - Subset of observers for ACR and a another subset for DCR
  - DCR limiting session duration (sequential visualization) x2
- Effect of different **display HW**: HMDs, desktop, screen
  - Different HMDs, desktop browser, normal screens...
  - Different devices in different labs?
- Effect of **audio:**
  - Compare ratings/exploration with and without audio for certain contents and HRCs.

Quality assessment for short 360° videos
Current Status

VQEG

Item	Status	Open Points
SRCs	9 (+9?) sequences	Other SRCs? More variety? Stitching tools? How many with audio? 180 content?
HRCs	Coding (homogeneous, heterogeneous), stitching, projections	QPs for each SRC? Tiling patterns?
Tools	TU Ilmenau, Nantes, Nokia Bell Labs	Refine and select. All HMDs and desktop?
Methodologies	ACR, DCR	Duration of test sessions? Question to ask in DCR?
Labs	UdN, Nokia Bell Labs, UPM, TU Ilmenau, CWI, Roma Tre, RISE, Gent.	Who else? Distribution of test conditions among labs: small common set, and whole test set split among labs.

- Coordination with ITU?
- Tests in May?