VQEG NORM Overview + Updates on Spatial Information/Temporal Information Indicators

VQEG NORM, May 2022

Werner Robitza <werner.robitza@gmail.com>
Cosmin Stejerean
Ioannis Katsavounidis
Lukas Krasula
… and many others

Overview — Project Introduction

VQEG No Reference Metrics (NORM)

Activities:

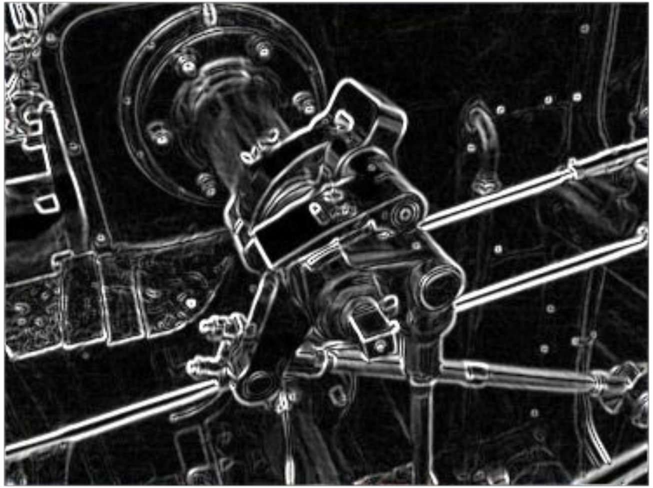
- NR Metric Development—What Is Our Design Goal?
  - Lead: Margaret Pinson
  - Evaluation of various NR metrics on different datasets
  - Lots of resources
- SI and TI Clarification
  - Lead: Werner Robitza
  - Update of spatiotemporal complexity indicators (← will give more info on this)
  - Meeting minutes
- Video Quality Metadata Standard
  - Lead: Ioannis Katsavounidis
  - Provide quality-related metadata in video sequences
  - Motivation document
  - Meeting minutes
  - Metadata proposal (T.35 metadata payload)

SI/TI Overview

- Defined in ITU-T Rec. P.910
- Classify spatiotemporal complexity of video sequences
- Definitions:
  - SI: Standard deviation of Sobel-filtered image
  - TI: A basic motion difference feature for adjacent frames

$$SI = max_{time}\{std_{space}[Sobel(Fn)]\}$$

$$M_n(i,j) = F_n(i,j) - F_{n-1}(i,j)$$

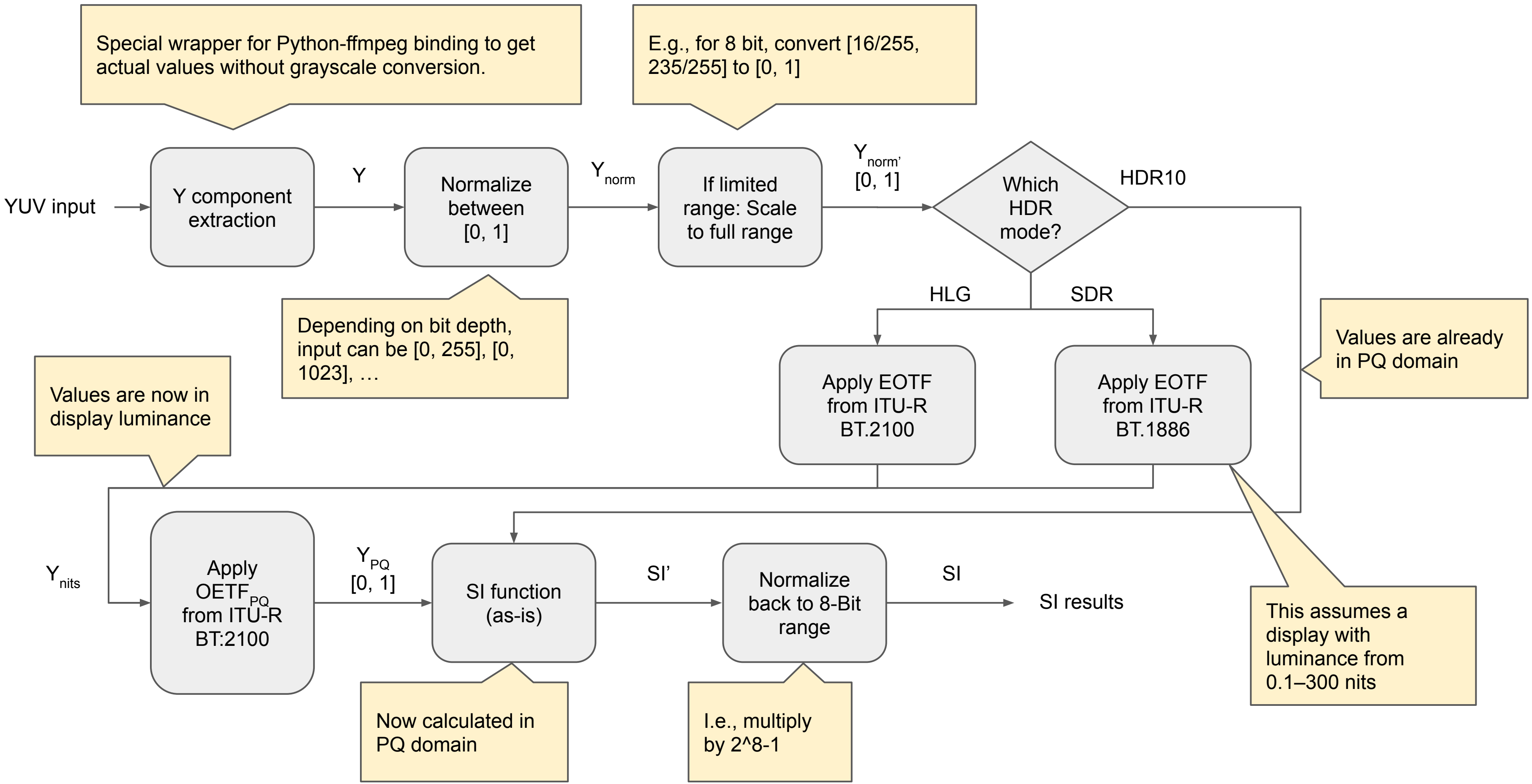


Overview: VQEG Updates to SI/TI Functions

- Make SI/TI future-proof:
  - Harmonize existing implementations wrt. handling of edges and full/limited range conversions
  - Handle content > 8 Bit per channel
  - Handle HDR content
  - Provide an update for ITU-T Rec. P.910
- Provide an even better encoding complexity metric:
  - Enhance SI/TI with basic motion compensation features
  - Ioannis/Cosmin from Meta will publish code to perform motion estimation analysis
  - Generic method to remove impact of content features that can be easily predicted with little information overhead → should give more accurate results

Activities so far

- Monthly meetings, open for everyone (link at VQEG website)
- Previous meeting notes:
  - https://docs.google.com/document/d/1pjAJet6YMznf1pPZ_5Xp0L3UiJvCh78x2LRIHvKApro/edit
- New software developed:
  - Main code written by Werner, additional input on HDR conversion functions from Lukas Krasula
  - Code is written in Python (slow, but ongoing activity to multi-thread)
  - Moved to VQEG organization now: https://github.com/VQEG/siti-tools
  - Tests included for SDR, HDR content
  - Legacy branch for old version of code
- Evaluations of old vs. new scores and others:
  - https://github.com/slhck/siti-evaluation
  - Various data, scripts and plots to reproduce analyses
- Writing of ITU-T Rec. P.910 recommendation update
  - https://docs.google.com/document/d/1pGqvifcoYk_nZ33Q-xnbOMTzQl4wpwgS/edit#heading=h.s8a3itid4k6q
  - Please add your suggestions!


Special wrapper for Python-ffmpeg binding to get actual values without grayscale conversion.
E.g., for 8 bit, convert [16/255, 235/255] to [0, 1]
YUV input
Y component extraction
Y
Normalize between [0, 1]
Y norm
If limited range: Scale to full range
Y norm' [0, 1]
Which HDR mode?
HDR10
HLG
SDR
Depending on bit depth, input can be [0, 255], [0, 1023], …
Values are already in PQ domain
Values are now in display luminance
Apply EOTF from ITU-R BT.2100
Apply EOTF from ITU-R BT.1886
Y nits
Apply OETF PQ from ITU-R BT:2100
Y PQ [0, 1]
SI function (as-is)
SI'
Normalize back to 8-Bit range
SI
SI results
This assumes a display with luminance from 0.1–300 nits
Now calculated in PQ domain
I.e., multiply by 2^8-1

Further Contents

- New vs. old SI/TI, brief comparison
- Analysis of resolution dependency
- Analysis of compression efficiency
- P.910 updates

New vs. old SI/TI

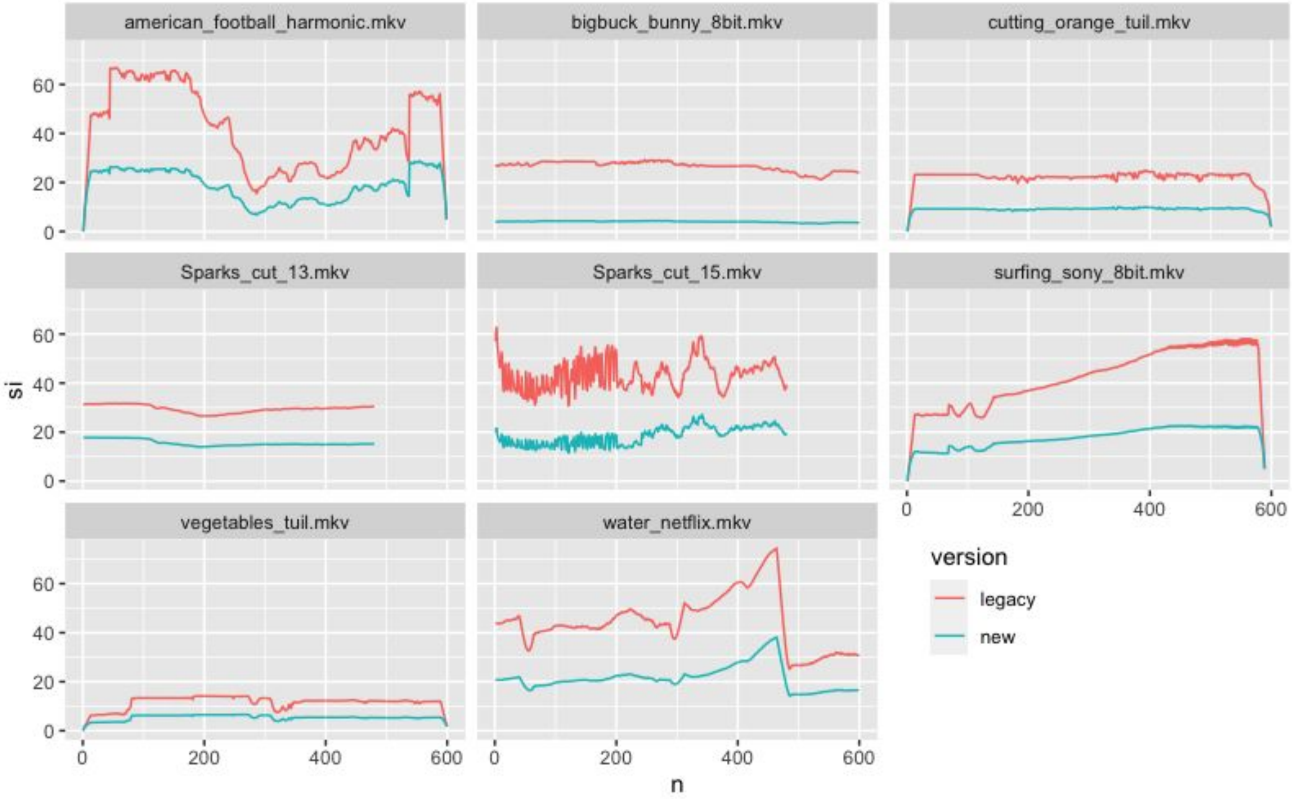
https://github.com/slhck/siti-evaluation/tree/master/analyze-siti-new-vs-legacy

Evaluation of new SI/TI

New vs. “legacy” SI/TI on AVT-VQDB-UHD1

10s, 2160p, 8-bit SDR

→ Values are lower due to shift to PQ domain

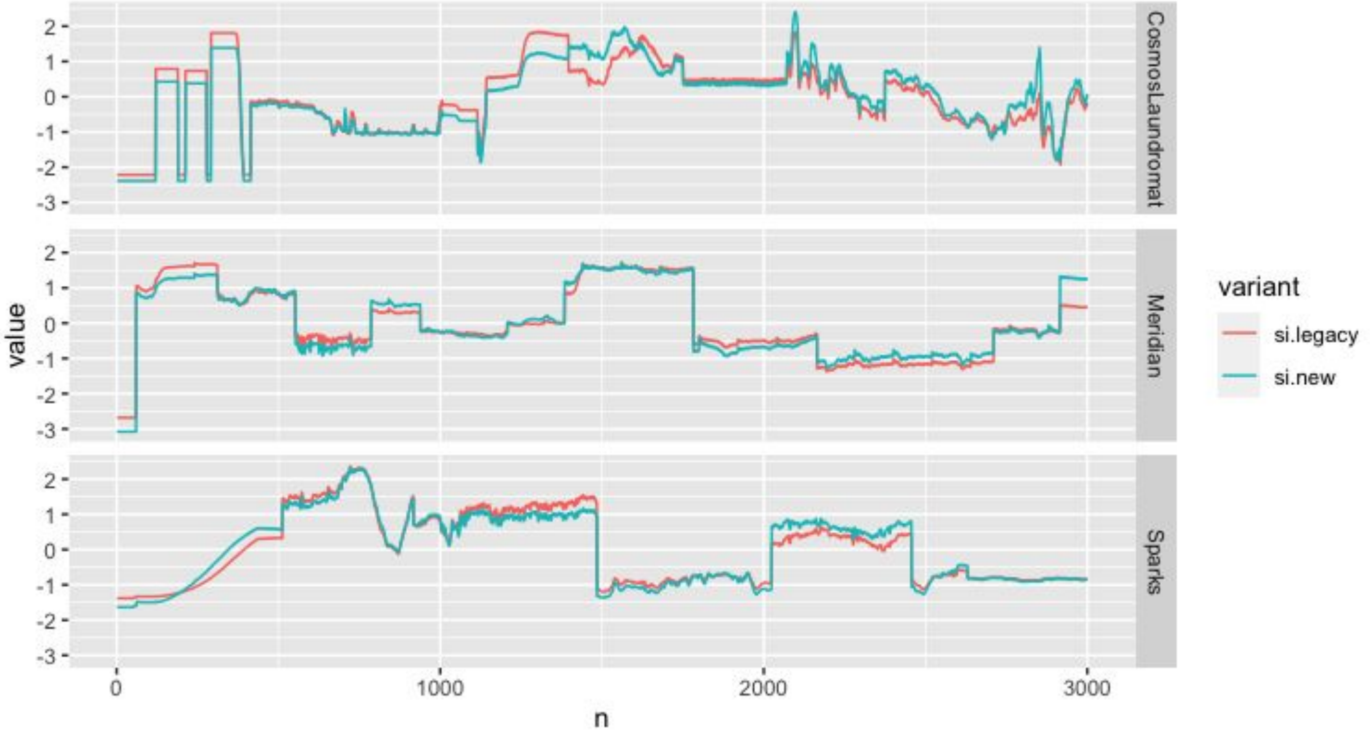


Evaluation of new SI/TI

New and old values for SDR for another set of sequences (Netflix Open Content)

Here, y-axis is normalized, so we see that the scores are basically the same.

Minor differences in scores may be related to “HDR-ness” of the content, especially for the Sparks sequence



Conclusion

- New SI operates the same way for SDR contents, just on a different scale
  - We did not want to change the underlying functionality here, just extend it for more use cases
  - So this makes sense and is a good result
- New SI gives about the same results for HDR10 compared to old
  - Goal achieved? For HDR10, values are already in PQ, so SI was expected to be the same

Resolution dependency

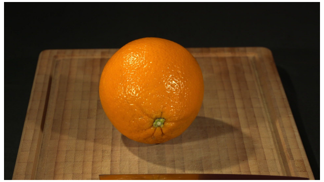
https://github.com/slhck/siti-evaluation/tree/master/analyze-siti-resolution

Caveat: SI Resolution Dependency

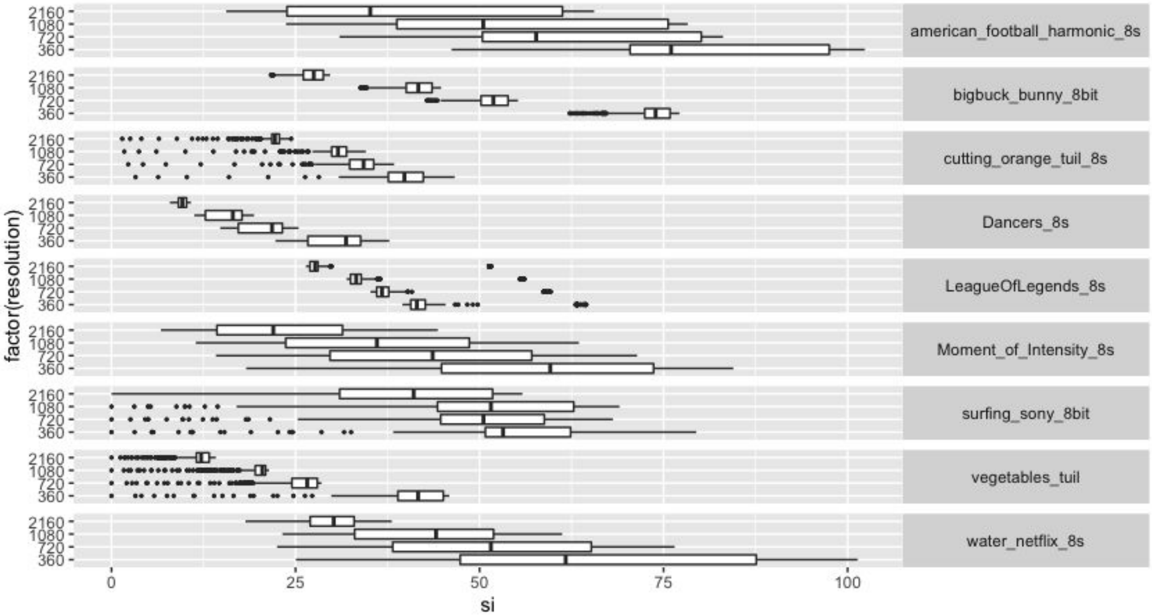
- The magnitude of SI intrinsically depends on the content resolution
- The gradients in the Sobel filter depend on the content itself
- With natural images, these gradients are not the same
- Compare e.g. Big Buck Bunny in 1080p vs 240p

Dancers

Vegetables



Cutting Orange

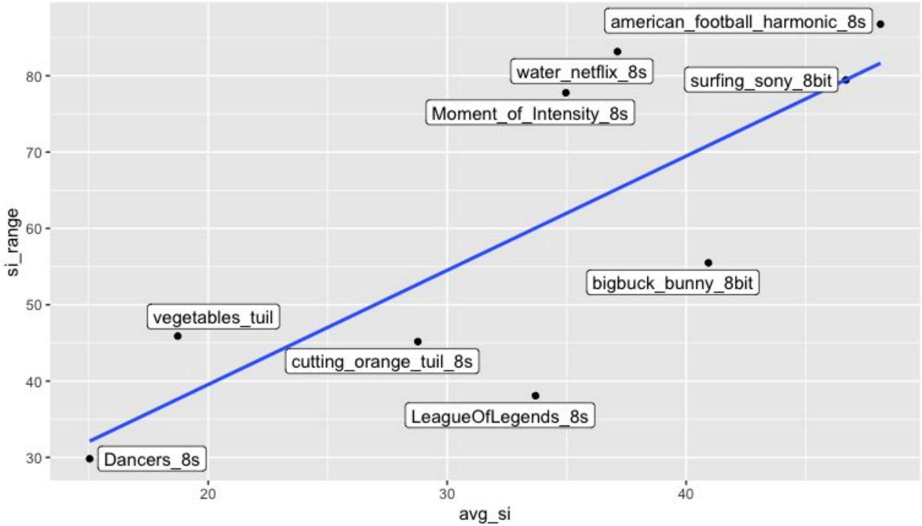


- Spread of values depends on content properties
- Similar range for “smooth” images without many gradients (*Dancers*, *Vegetables, Cutting Orange*)
- Wide range for sports clips (*Surfing Sony, American Football*)
- Largest differentiation in *Big Buck Bunny*, which also retains the largest SI for the lowest resolution

Water Netflix

Intrinsic Resolution Dependency

- If there is little information to begin with, changing the resolution will not have a big impact
- Absolute range of SI across all resolutions depends on the absolute value of SI (e.g. for the 1080p resolution)
- Seems to be the case for the dataset
- Reasonable linear fit between avg. SI for 1080p variant, and range of all SI values across entire resolution set
- Differentiator is still the content type (here, e.g. more natural vs. CGI/gaming)



Conclusion

- Range of SI depends on
  - Original resolution
  - Absolute SI value of content
  - Content properties (smoothness/naturalness etc.)
- A compensation function would be nice so that we could compare SI across contents with different resolutions, but seems challenging to develop without analyzing a broad set of contents
- Suggestion to stick with what we have for now, and warn users not to compare across resolutions

Analysis of compression efficiency

https://github.com/slhck/siti-evaluation/tree/master/analyze-siti-vs-other-metrics

Compression Efficiency

- We want to understand how well a certain clip can be compressed
    - i.e., retain quality under lossy compression
    - Quality is defined subjectively or via instrumental metrics
- How well do SI/TI and other metrics explain this compressibility?
- Analysis from QoMEX 2021 paper:
    - Robitza, W., Rao Ramachandra Rao, R., Göring, S., & Raake, A. (2021). Impact of spatial and temporal information on video quality and compressibility. 2021 13th International Conference on Quality of Multimedia Experience, QoMEX 2021, 65–68.
    - Also gave presentation at VQEG meeting in June, 2021.
- Code and data are now open-sourced

Goals — Working Hypothesis

Videos with higher SI/TI should be harder to compress

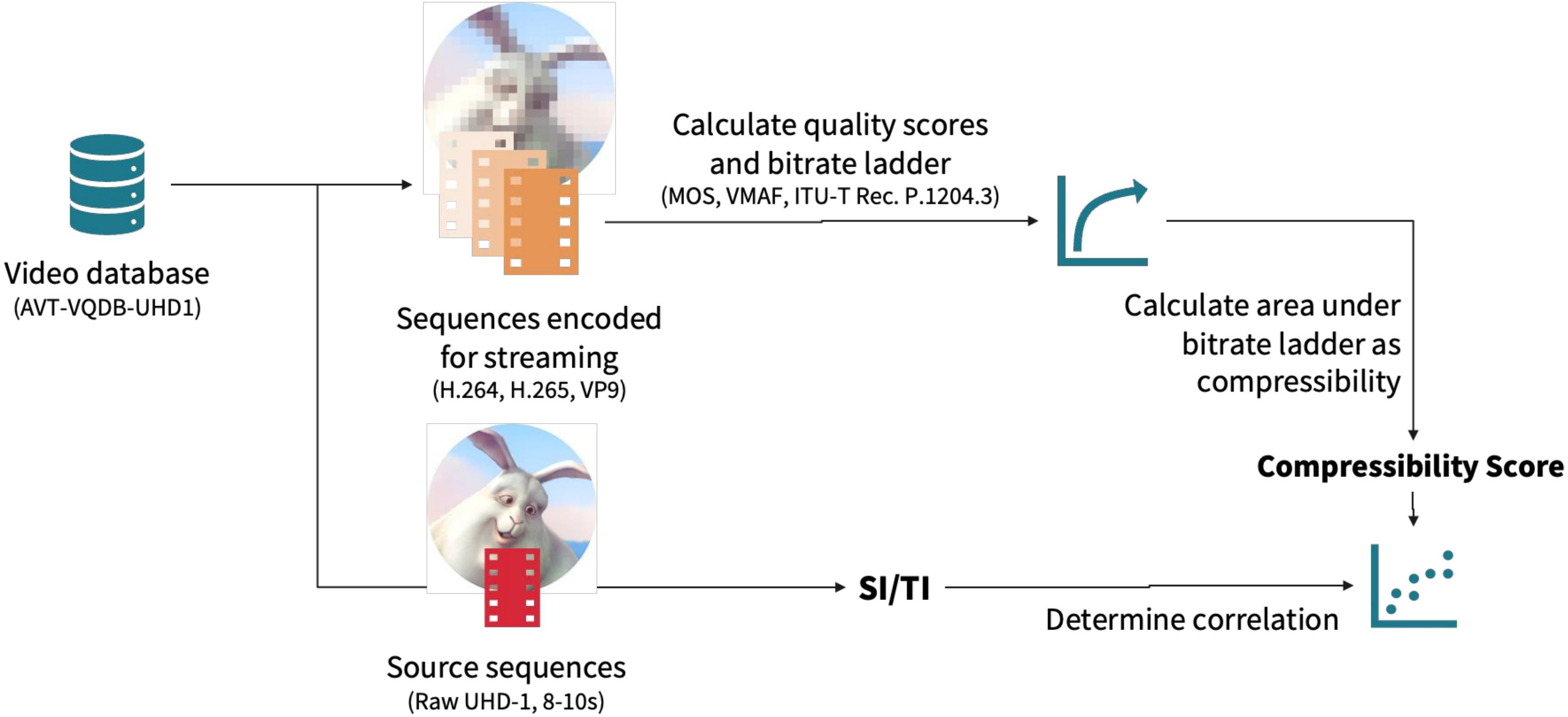
Videos with higher SI/TI have lower quality when compressed under bitrate constraint

Compressibility of a source == achievable quality under bitrate constraints

High SI/TI lead to lower quality and lower compressibility

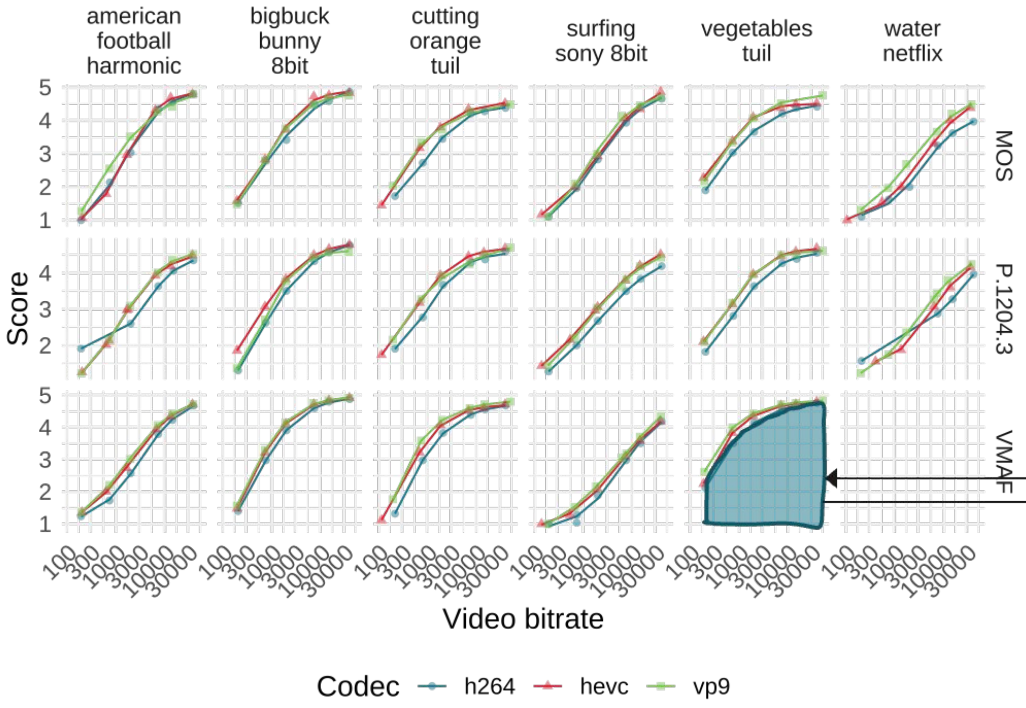
Quality == subjective or “objective” MOS

Approach



Video database
(AVT-VQDB-UHD1)
Sequences encoded
for streaming
(H.264, H.265, VP9)
Calculate quality scores
and bitrate ladder
(MOS, VMAF, ITU-T Rec. P.1204.3)
Calculate area under
bitrate ladder as
compressibility
Compressibility Score
Source sequences
(Raw UHD-1, 8-10s)
SI/TI
Determine correlation

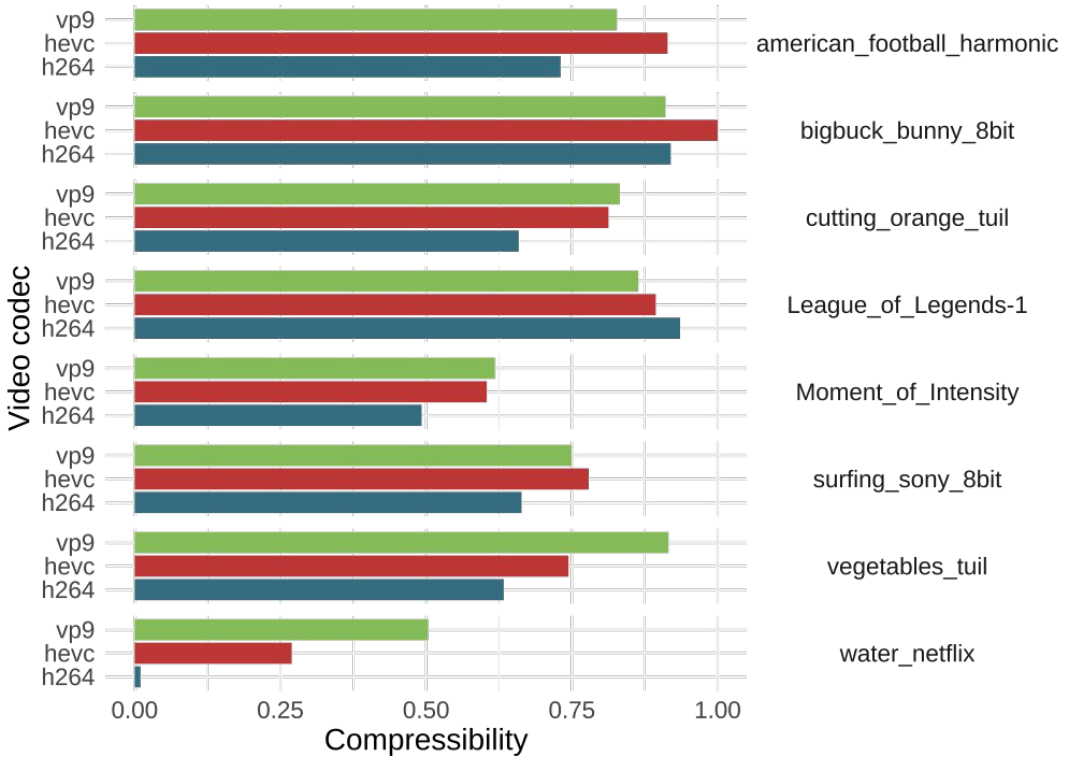
Pipeline



1. Create bitrate ladder for each SRC, i.e. MOS against bitrate
2. Construct convex hull of ideal points
3. Fit sigmoid function against hull
4. Calculate are under each curve as quality/compressibility for that SRC/codec

Compressibility
1 number for each SRC/codec combination

Compressibility Results



Each SRC and codec have different compressibility scores

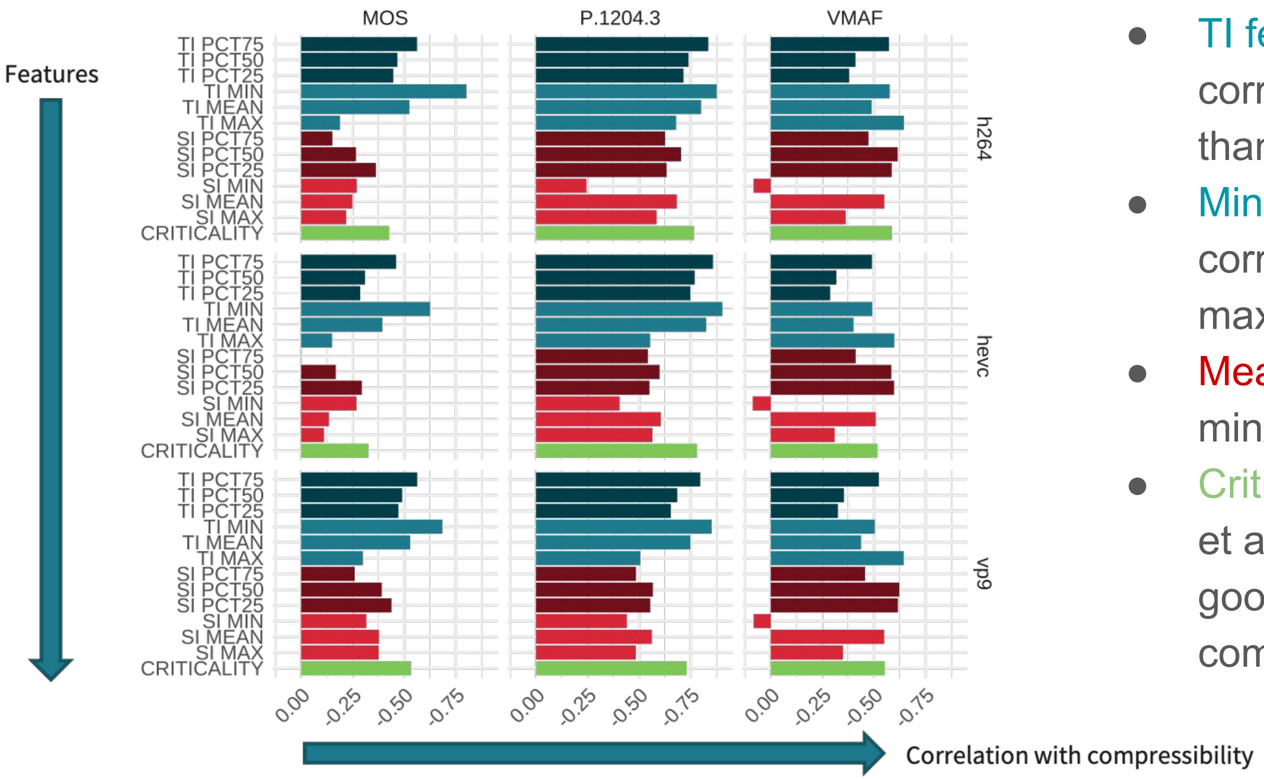
Note:

- Scores are normalized between 0 and 1 for this analysis (to be refined)
- Scores are based on MOS here

Examples:

- BBB is the easiest to compress, although used very often in tests
- Netflix Water sequence is the hardest

Correlation Results



- TI features have higher correlation with compressibility than SI features
- Minimum TI seems like a better correlated indicator than maximum TI
- Mean SI correlates better than min/max
- Criticality metric from Fenimore et al. (basically *log(SI × TI)*) has good correlation with compressibility

New Data

- Currently calculating additional explanatory metrics:
  - CRF-based content complexity indicator
  - VCA output / VCD from Hadi Amirpour, Vignesh V. Menon (see Presentation #105)
- Can implement any data source in the original analysis code
- Please share what you have, if you want something added

P.910 Updates

Existing Recommendation ITU-T Rec. P.910

- Current state:
  - Contains a description of how SI/TI can be used
  - Definition of Sobel operator
- Updated text:
  - https://docs.google.com/document/d/1pGqvifcoYk_nZ33Q-xnbOMTzQl4wpwgS/edit#heading=h.s8a3itid4k6q
  - Add SDR/HDR processing pipeline (pre- and post-processing steps)
  - Better description of edge cases, usage notes
  - Add bibliography for relevant research
- Planned to submit to ITU-T Study Group 12 meeting, deadline: 25 May 2022