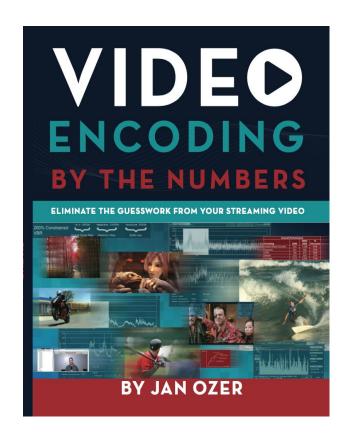# PSRN vs. VMAF vs. SSIMPlus
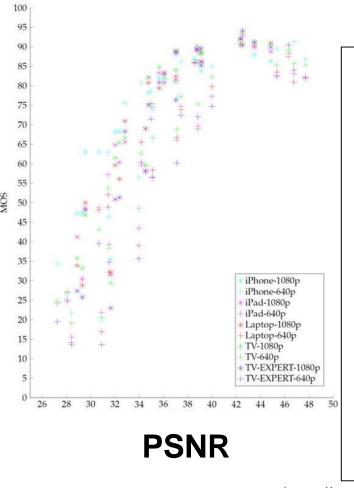
A Simple Visual Comparison
Part 1:

Introduction

- PSNR is widely criticized as a poor predictor of subjective ratings

- Netflix switched from PSNR to VMAF about 14 months ago (still sites PSNR results)

- SSIMwave recently released a study showing that SSIMplus was more accurate than VMAF

  - http://bit.ly/ssimplus_v_vmaf

- My book ties encoding decisions to video quality metrics, primarily PSNR
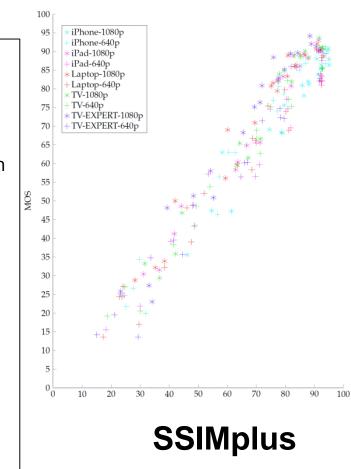
- So, how valid are the book results?

http://bit.ly/Ozer_msd

- According to University of Waterloo, PSNR (on left) has poor consistency with subjective Mean Opinion Scores (MOS) (human ratings)

- Consistency is much greater with SSIMplus (on the right)

So, how bad is PSNR?

PSNR

SSIMplus

https://ece.uwaterloo.ca/~z70wang/research/ssimplus/

Test Procedure - Group 1

- Encode four 720p test files using tests for profile, x264 preset, bitrate control, and B-frame interval

  - **Haunted** - movie like production https://youtu.be/bHan6KmhkyI

  - **Freedom** - music video - https://youtu.be/FqLF5lnewT4

  - **Tears of Steel** excerpt

  - **Sintel** excerpt

- All tests encoded with FFmpeg

- **PSNR** - Moscow University Video Quality Measurement Tool

  - http://bit.ly/VQMT_review

- **SSIMPlus** - SSIMWave Quality of Experience Monitor (SQM)

  - http://bit.ly/SQM_review

- **VMAF** - Hybrik Cloud Platform

  - www.hybrik.com

SSIMplus Much More Functional

- Can assess quality on different devices

  - Ran tests here on single 54" 1080p TV

- Can run cross-resolution/cross frame rate tests

- Can automatically align test and reference files

- Overall, a more comprehensive metric

- This functionality comes at a price; somewhere north of ~$3,000.

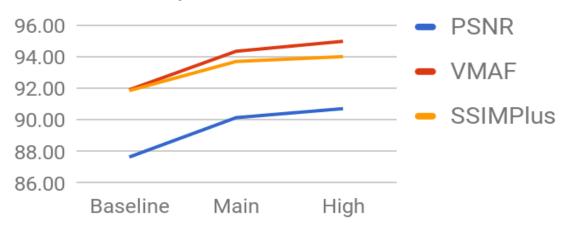- MSU VQMT is about $1,000, but PSNR available in FFmpeg and via cloud encoding facilities like Hybrik

Other Presentation Details

- **PSNR scores multiplied by 2.25 to get them in the same ballpark as VMAF/SSIMplus**

- VMAF/SSIMplus show actual results

H.264 Profile - Composite over 4 Videos

- Trendline very similar
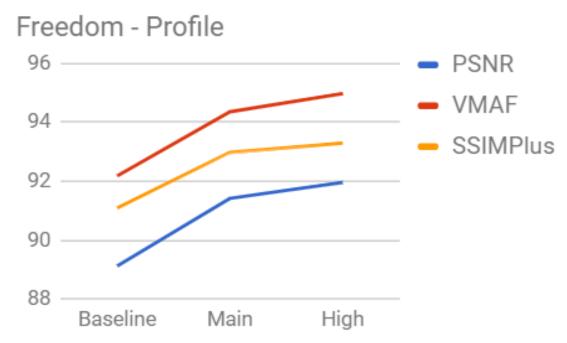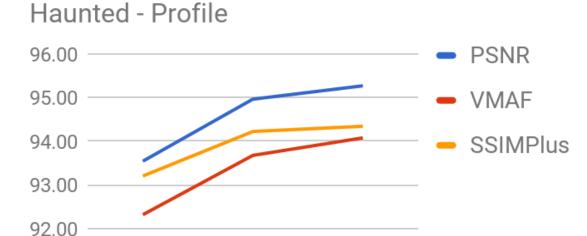- VMAF shows significantly lower quality at the start

Profile - Composite Over 4 Videos

H.264 Profile - Freedom

- Very consistent

Freedom - Profile

H.264 Profile - Haunted

- Overall results consistent
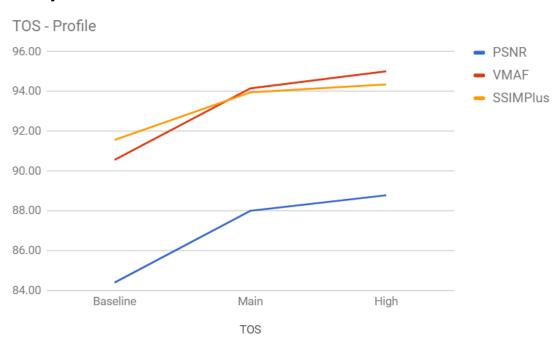- Different rate of change between Main and High
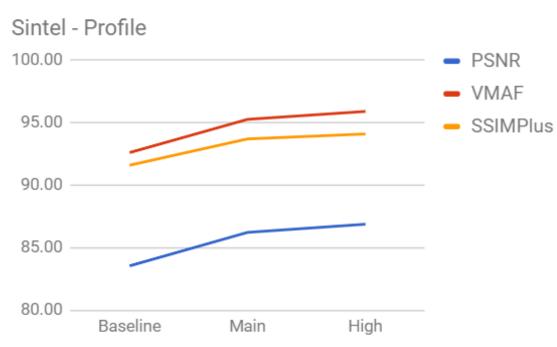

Haunted - Profile

Legend: PSNR, VMAF, SSIMPlus

H.264 Profile - Tears of Steel

- x264 preset
- Trendline very similar
- VMAF shows significantly lower quality at the start

TOS - Profile
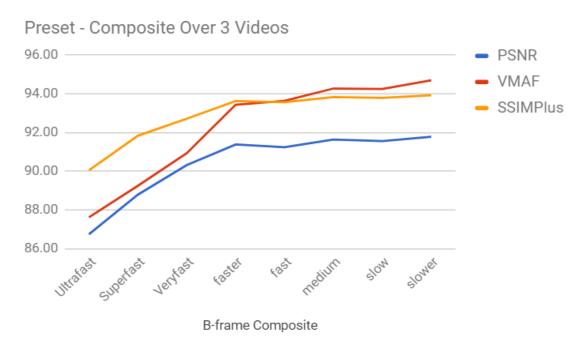
Legend: PSNR, VMAF, SSIMPlus

H.264 Profile - Sintel
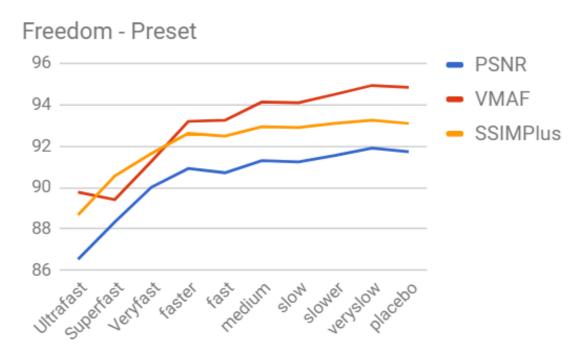
- Very similar

Sintel - Profile

Legend: PSNR, VMAF, SSIMPlus

x264 Preset - Composite Over Three Videos

- Trendline very similar
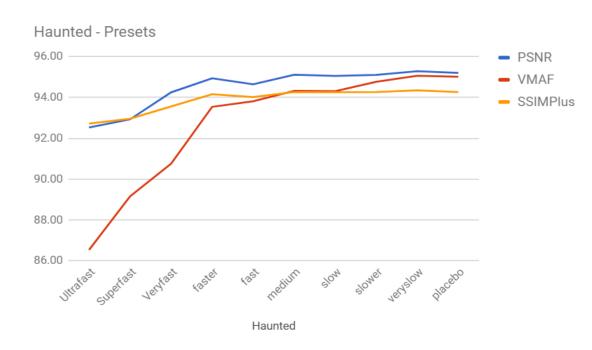- VMAF shows significantly lower quality at the start

Preset - Composite Over 3 Videos

x264 Preset - Freedom

- VMAF shows Superfast as lowest quality
- Otherwise consistent

Freedom - Preset

Legend: PSNR (blue), VMAF (red), SSIMPlus (orange)

X-axis: Ultrafast, Superfast, Veryfast, faster, fast, medium, slow, slower, veryslow, placebo
Y-axis: 86 to 96

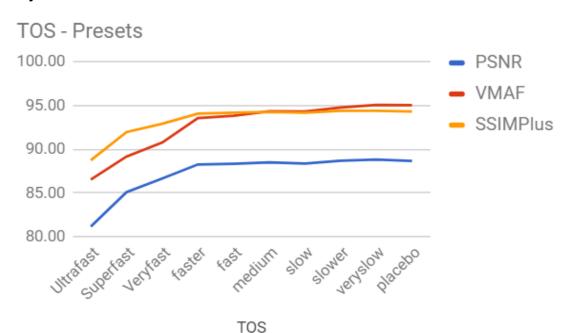# x264 Preset – Haunted

- VMAF shows significantly lower quality at the start
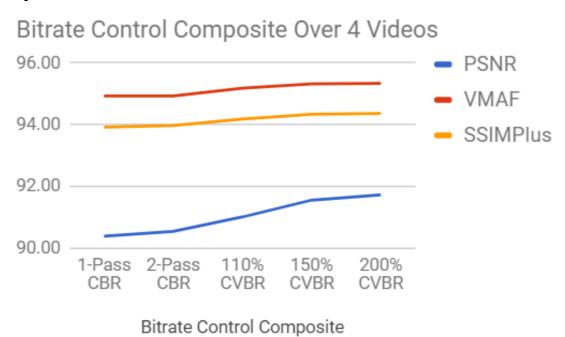- PSNR and SSIMPlus very consistent

Haunted - Presets

x264 Preset - Tears of Steel

- x264 preset
- Trendline very similar
- VMAF shows significantly lower quality at the start

TOS - Presets
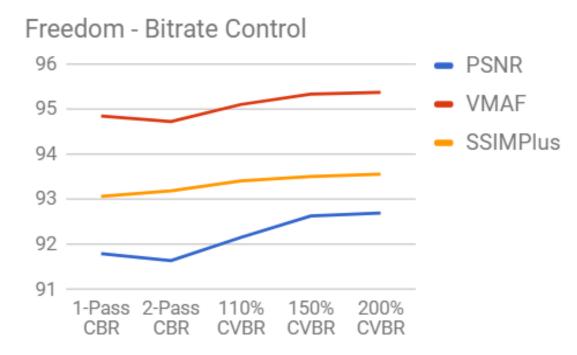
Bitrate Control – Composite over 4 Videos

- PSNR shows much great quality differential

Bitrate Control Composite Over 4 Videos

Legend: PSNR (blue), VMAF (red), SSIMPlus (orange)

X-axis: 1-Pass CBR, 2-Pass CBR, 110% CVBR, 150% CVBR, 200% CVBR

Y-axis: 90.00 – 96.00

Bitrate Control Composite

Bitrate Control - Freedom

- VMAF/PSNR show 1-pass CBR as higher; SSIMPlus no
- Constrained VBR higher overall with all three metrics

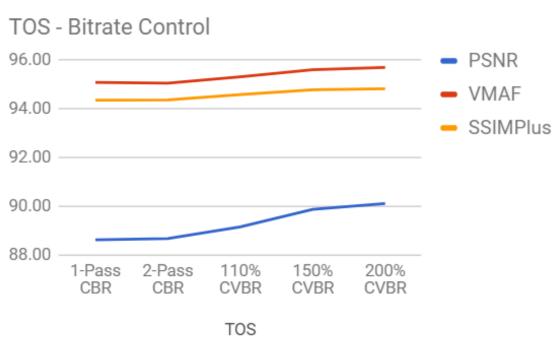Freedom - Bitrate Control

Bitrate Control - Haunted

- General trends similar but VMAF/SSIMPlus have 1-pass CBR as highest quality overall by slim margin
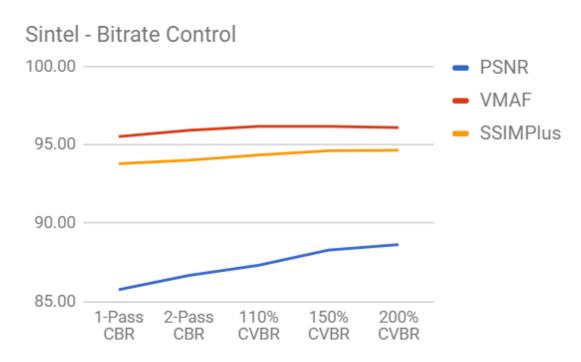

Haunted - Bitrate Control

Bitrate Control – Tears of Steel

- Order is restored
- Again, PSNR shows greater qualitative difference between techniques

TOS - Bitrate Control
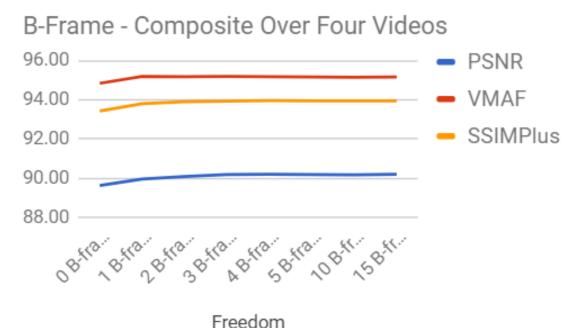
Bitrate Control - Sintel

- Very similar
- PSNR again with much greater differential

Sintel - Bitrate Control

Legend: PSNR (blue), VMAF (red), SSIMPlus (orange)

X-axis: 1-Pass CBR, 2-Pass CBR, 110% CVBR, 150% CVBR, 200% CVBR
Y-axis: 85.00 to 100.00

B-Frame Interval - Composite over 4 Videos

- Very similar

B-Frame - Composite Over Four Videos

B-Frame Interval - Freedom

- Very similar

Freedom - B-frames
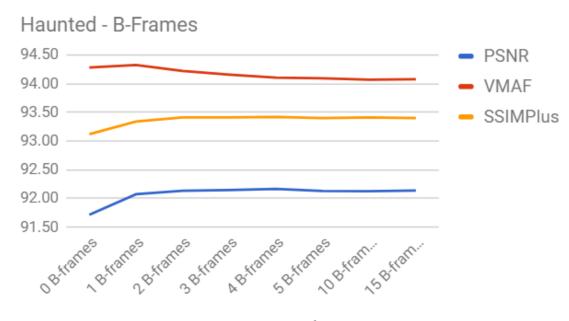
B-Frame Interval - Haunted

- VMAF fines highest quality at fewest number of B-frames
- SSIMPlus/PSNR very similar

Haunted - B-Frames

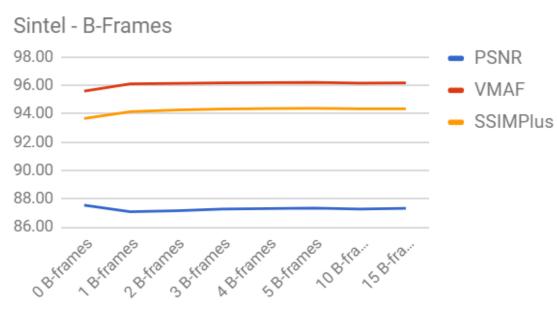B-Frames/Haunted

Legend: PSNR, VMAF, SSIMPlus

B-Frame Interval - Tears of Steel

- Order is restored

TOS - B-Frames

B-Frame Interval - Sintel

- Here, PSNR shows 0 B-Frames as the highest quality

Sintel - B-Frames

Legend: PSNR, VMAF, SSIMPlus

B-Frames/Sintel

Conclusions

- The key question is, how many times would I have reached a different recommendation by using VMAF or SSIMplus. The answer is not that often (and never where both SSIMplus and VMAF agreed).

- So, for simple configuration decisions, PSNR results were reasonably consistent with VMAF and SSIMPlus. If PSNR is the only tool you have affordable access to, it appears useful for these types of comparisons.

- Next time, I evaluate the results achieved when comparing different codecs with the same metrics. Preliminary results tend to indicate that the differentials will be much greater (and PSRN not so useful).

- For the record, I did not tune the H.264 encodes in these tests for PSNR. I probably should have for PSNR, but shouldn't have VMAF and SSIMplus. Then I would be testing with two different file sets, which makes no sense.