Perceptual video quality assessment for adaptive streaming encoding

Estêvão C. Monteiro, Ricardo E. P. Scholz, Carlos A. G. Ferraz, Tsang I. Ren, Roberto S. M. Barros

{ecm3, reps, cagf, tir, roberto}@cin.ufpe.br

Centro de Informática, Universidade Federal de Pernambuco, Brasil

IEEE VCIP 2015

Outline

- Encoding for adaptive streaming over HTTP
- Perceptual video quality assessment metrics
  - The Shifted Gradient Similarity index (SG-Sim)
  - Visual quality comparison
  - Feature pooling filters
- jVQA: Video Quality Assessment in Java
- Experiments
- Conclusions

Encoding for adaptive streaming over HTTP

- Narrow bandwidths in Internet limit video data rates, and small keyframe intervals reduce compression efficiency.
- **Efficient lossy encoding** is required to minimize **blurring** and **artifacts**.
- Rate-distortion optimization in encoder requires efficient metrics, for low latency.

Perceptual video quality assessment metrics

- The Structural Similarity (**SSIM**) index is a relevant and efficient full-reference perceptual IQA metric.
  - Multi-scale SSIM (MS-SSIM)
  - Gradient SSIM
  - Gradient-weighted SSIM (3-SSIM, 4-SSIM)
  - Fast SSIM
  - Gradient Magnitude Similarity Deviation (GMSD).

Perceptual video quality assessment metrics

- Fast SSIM and GMSD similarity equation for source *S* and version *V*:

$$FastSSIM(S,V) = \frac{2\nabla_S \nabla_V + C_1}{\nabla_S^2 + \nabla_V^2 + C_1}$$

  - Stabilized (and distorted) by constant C_1.

Perceptual video quality assessment metrics

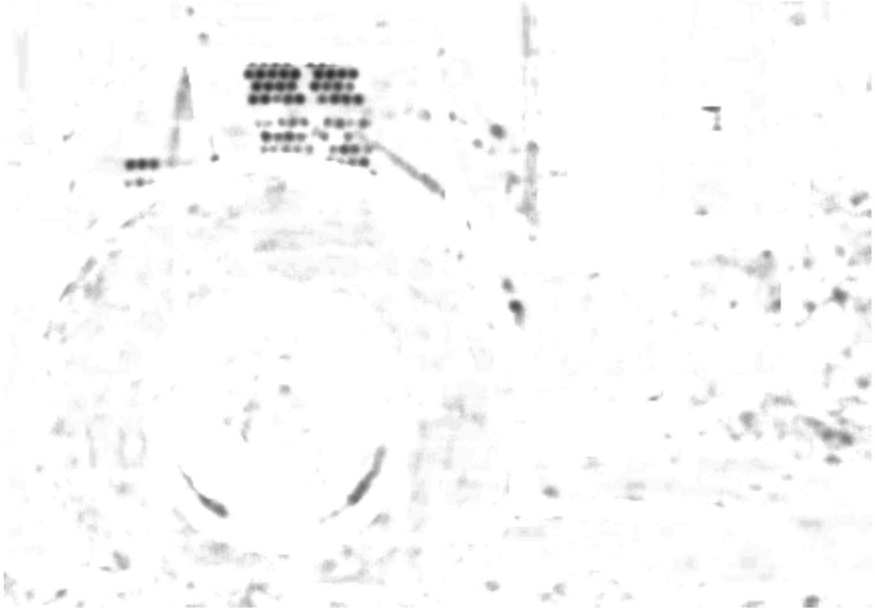
- Rouse & Hemami remove the stabilizing constant and apply logical treatment:
  - If $\nabla_S = \nabla_V$, then SSIM = 1;
  - else, if $\nabla_S = 0$ or $\nabla_V = 0$, then SSIM = 0;
  - else, compute normal SSIM (without constants).
  - Dynamic range is broader, but distortions are more dramatic due to loss of information (zeros).

Visual quality comparison

Original image detail

Compressed image detail

Visual quality comparison



Quality map by Fast SSIM: clean

Quality map by Fast SSIM with logical similarity stabilization: artifacts

Shifted gradient similarity

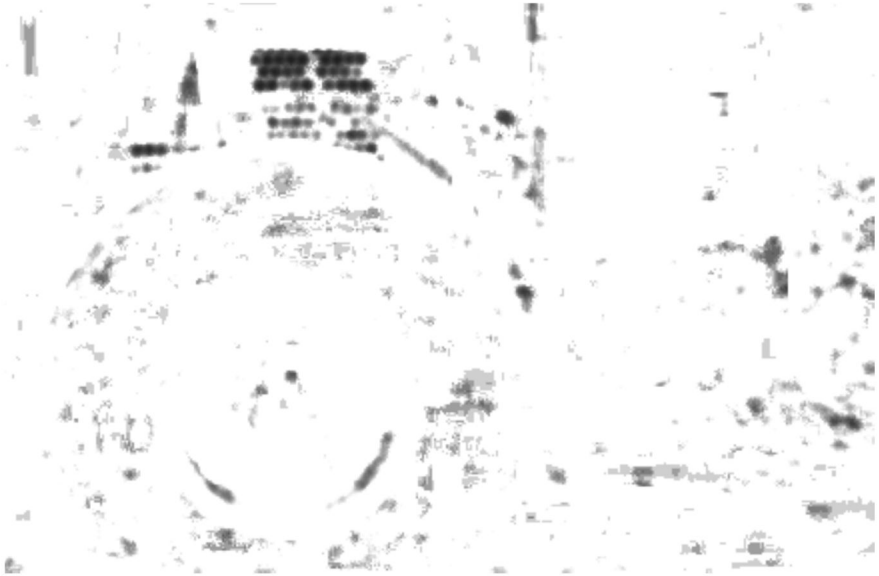
- Gradient magnitude approximation in Fast SSIM:

$$\nabla_n = \max(|\nabla i_n|, |\nabla j_n|) + {}^{1}/_{4} \min(|\nabla i_n|, |\nabla j_n|)$$

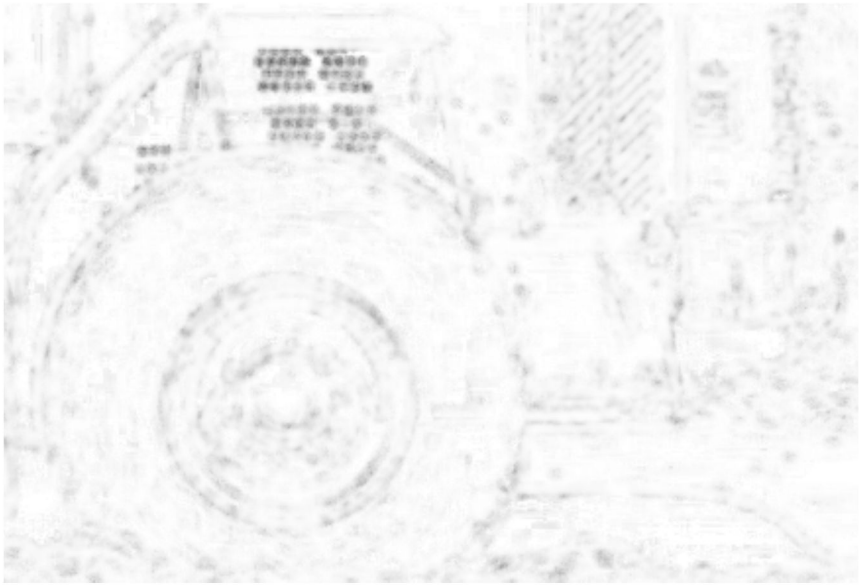
- **Shifted** gradient magnitude:

$$\nabla_n = \max(|\nabla i_n|, |\nabla j_n|) + {}^{1}/_{4} \cdot \min(|\nabla i_n|, |\nabla j_n|) + \mathbf{1}$$

Visual quality comparison

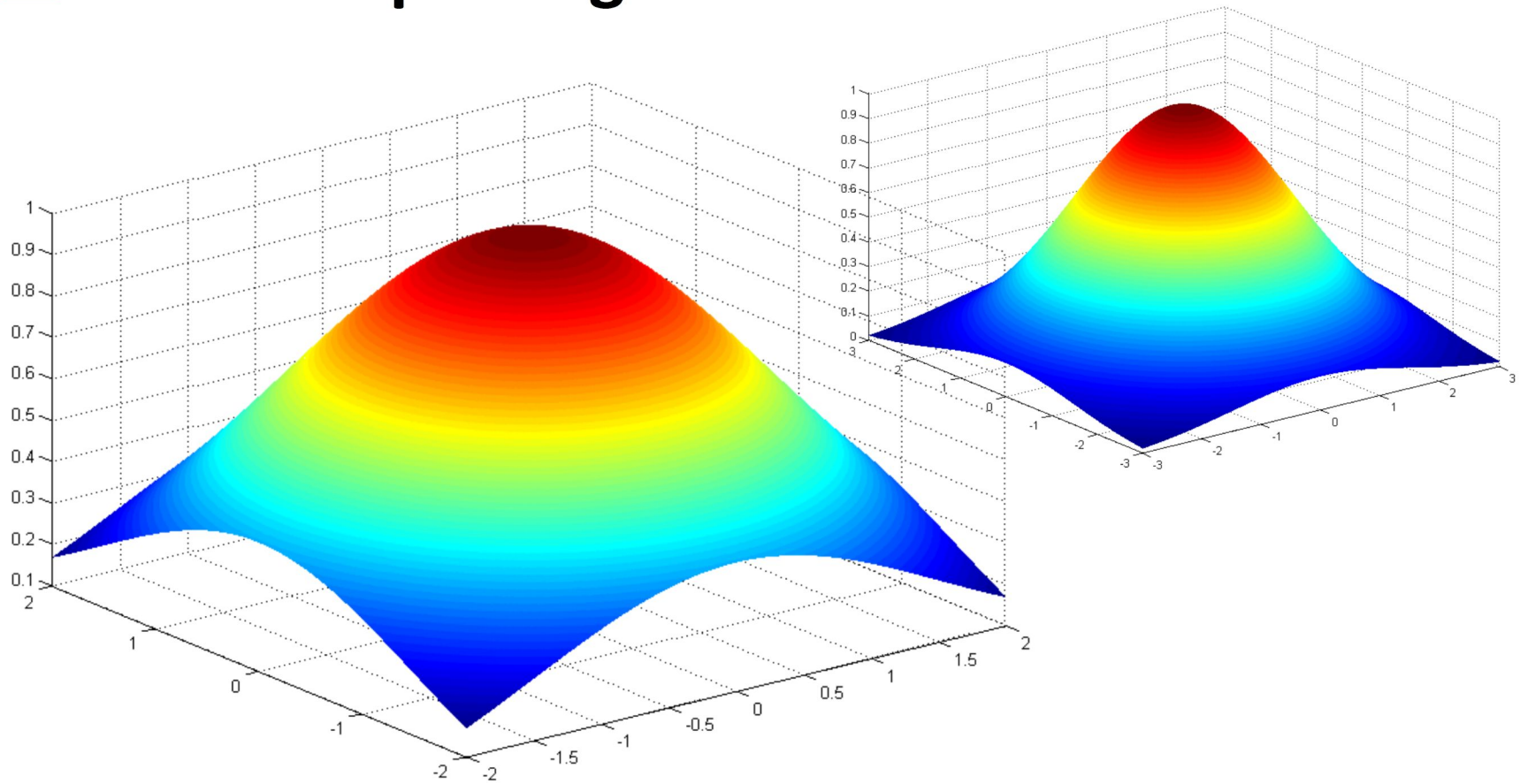


Quality map by SG-Sim:
wider dynamic range



Quality map by SSIM:
more monotonic

Feature pooling filters



3D representation of **2*σ*-**radius and **3*σ*-**radius Gaussian operator of σ = 1.5.

Feature pooling filters

0	0	0	1	1	0	0	0
0	0	1	2	2	1	0	0
0	1	2	4	4	2	1	0
1	2	4	8	8	4	2	1
1	2	4	8	8	4	2	1
0	1	2	4	4	2	1	0
0	0	1	2	2	1	0	0
0	0	0	1	1	0	0	0

8×8 integer operator from Fast SSIM

2	6	12	15	12	6	2

1D Gaussian integer operator

Feature pooling filters

- 7×7 Box downsampling instead of sliding the window:
  - **275% faster** than 11×11 Gaussian.
  - Resulting index is **98% similar** to 11×11 Gaussian.
  - Significant efficiency gain.
  - Optimized with integral image (summed area table).

jVQA: Video Quality Assessment in Java

- Multi-platform testing suite
  - Implements SSIM, MS-SSIM, 3-SSIM, 4-SSIM, Fast SSIM, Fast MS-SSIM, GMSD, and **SG-Sim**.
  - Facilitates customization.
  - Supports virtually any video file as input (Ffmpeg + AviSynth)
  - Free and open source software at http://sourceforge.net/projects/jvqa/

Experiments

- Differential mean opinion score (DMOS) prediction:
  - Spearman rank-order correlation coefficient (SROC);
  - Pearson linear correlation coefficient (PLC);
  - Root mean squared error (RMSE).
- LIVE Mobile VQA dataset

Visual quality index	SRC	PLC	RMSE	Time
4-scale SG-Sim, 2σ Gaussian	**0.917**	**0.907**	**0.48**	**0.37**
5-scale SG-Sim, 2σ Gaussian	**0.913**	**0.904**	**0.49**	0.99
4-scale SG-Sim, 5×5 downsampled	**0.908**	**0.901**	**0.49**	**0.19**
MS-SSIM, 2σ Gaussian	0.840	0.839	0.62	**1.40**
SG-Sim, 2σ Gaussian	0.823	0.812	0.67	0.69
Fast SSIM, 2σ Gaussian	0.807	0.803	0.68	0.68
SG-Sim, 7×7 downsampled *	0.805	0.807	0.67	**0.22**
GMSD	0.782	0.804	0.68	0.59
SG-Sim, 5×5 Box	0.781	0.797	0.69	0.43
Fast SSIM, 2σ Gaussian, logical-stabilized	**0.747**	**0.773**	**0.72**	0.68
3-SSIM, 2σ Gaussian	**0.731**	**0.761**	**0.74**	**1.48**
SSIM, 2σ Gaussian	**0.708**	**0.743**	**0.76**	**1.00**

* Updated result, relative to the paper.

Conclusions

- Shifted Gradient Similarity (SG-Sim)
  - Requires no additional similarity index stabilization.
  - Highest DMOS prediction (subjective quality).
  - 45% faster than SSIM.

Conclusions

- 7×7 Downsampled Pooling
  - 98% similar to 11×11 Gaussian.
  - 275% faster than 11×11 Gaussian.
  - Highest efficiency of all when combined with 4-scaling: **Fast MS-SG-Sim**.

Future work

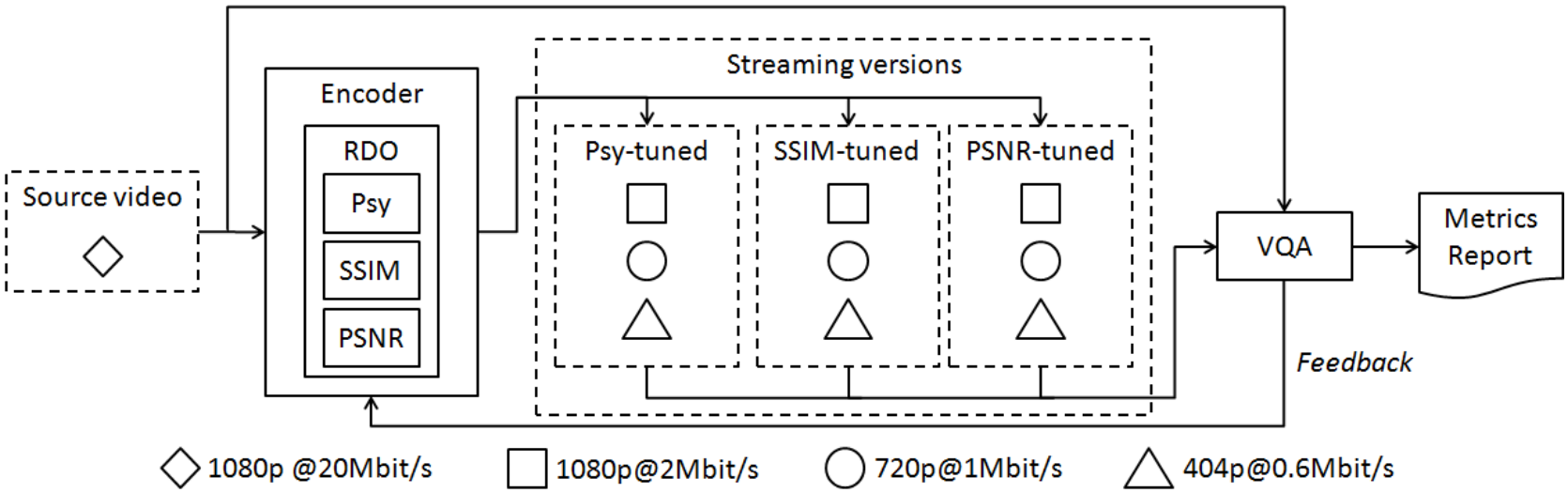
- SSIM-based indexes do not scale well
  - Lower resolution = low pass filter = blurring
  - Worse with cinema film grain noise
  - Adjustment for a universal index requires dynamic index adaptation to content
- General video and image quality assessment
  - Expand DMOS tests to general VQA and IQA datasets
  - Verify if SG-Sim is also effective outside of video encoding

Extra slides...

Encoding for adaptive streaming over HTTP

- Multiple content versions:
  - **Display resolutions**
    - QVGA, VGA, HD (720p), Full HD (1080p), Ultra HD (4K)...
  - **Decoding capabilities**
    - H.264 Baseline Profile, H.264 High Profile, H.265...
  - **Internet bandwidths**
    - 3G, 4G, DSL, cable...

Perceptual video quality assessment metrics



Feature pooling filters

11×11 Gaussian operator of σ = 1.5 (3.3σ radius).

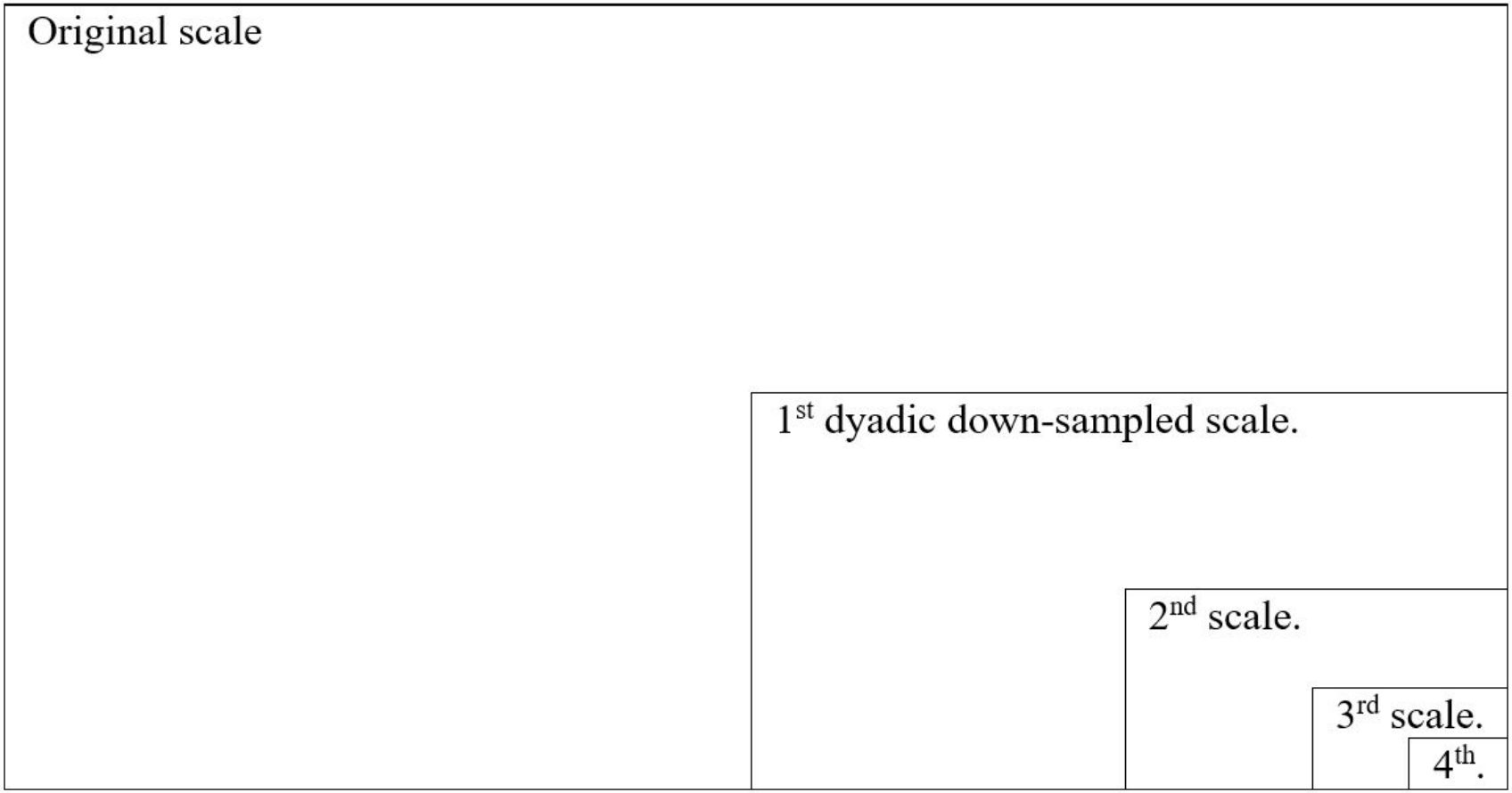
0	1	5	16	31	39	31	16	5	1	0
1	8	39	117	229	286	229	117	39	8	1
5	39	183	556	1084	1353	1084	556	183	39	5
16	117	556	1690	**3292**	4111	**3292**	1690	556	117	16
31	229	1084	**3292**	6412	**8007**	6412	**3292**	1084	229	31
39	286	1353	4111	**8007**	10000	**8007**	4111	1353	286	39
31	229	1084	**3292**	6412	**8007**	6412	**3292**	1084	229	31
16	117	556	1690	**3292**	4111	**3292**	1690	556	117	16
5	39	183	556	1084	1353	1084	556	183	39	5
1	8	39	117	229	286	229	117	39	8	1
0	1	5	16	31	39	31	16	5	1	0

96.6% of the total weights are within the emphasized 7×7 (2σ radius), requiring **40.5%** as many operations.

The 5×5 core of **20.7%** coefficients holds **86%** of the weights.

Smaller operators are more efficient due to spatial coherency.

Input scaling



JVQA — Video Quality Assessment in Java

Input

Select original video clip...

AviSynth script - original clip

```
DirectShowSource("bf_org.mp4")
ConvertToYV12()
```

Select distorted video clip...

AviSynth script - distorted clip

```
DirectShowSource("bf_r1.mp4")
ConvertToYV12()
```

Analysis

VQA metric composition

Structure scale:	Multi-scaled (4 half-scales)
Structure statistic:	Shifted gradient
Stabilization:	Logical (full dynamic range)
Pooling filter:	Downsampled (box)
Pooling size:	5
Luminance statistic:	None (ignore luminance)

(See tooltips for explanations of parameters.)

Operation

RUN

SAVE

Operation log

```
22:53:53 INFO  [Thread-16] estevaocm.jvqa.frameserver.FfmpegClip.getFormatContext(?), line 134:
Input #1, avisynth, from F:\jvqa\live\bf_org.avs:
  Duration: 0:0:18.0, bitrate: N/A
    Stream #1.0: Video: rawvideo (I420 / 0x30323449), yuv420p, 1280x720, 25.0 fps, 25.0 tbn, 25.0 tbc

22:53:54 INFO  [Thread-16] estevaocm.jvqa.frameserver.FfmpegClip.getFormatContext(?), line 134:
Input #2, avisynth, from F:\jvqa\live\bf_r1.avs:
  Duration: 0:0:18.0, bitrate: N/A
    Stream #2.0: Video: rawvideo (I420 / 0x30323449), yuv420p, 1280x720, 25.0 fps, 25.0 tbn, 25.0 tbc

22:53:54 INFO  [Thread-16] estevaocm.jvqa.cli.Experiment.run(?), line 104:
{multiscaled-4 shifted-gradient-structure downsampled 5x5 conditional-stabilized ignore-luma}

22:54:26 INFO  [Thread-16] estevaocm.jvqa.cli.Experiment.logResults(?), line 271:
Job completed. Visual quality index: 0.979067 (16.79db, 95.86%). Time elapsed: 31,166 ms. Total frames in video: 450.
```

Clear Log