Prediction of Satisfied User Ratio (SUR) for Compressed Video

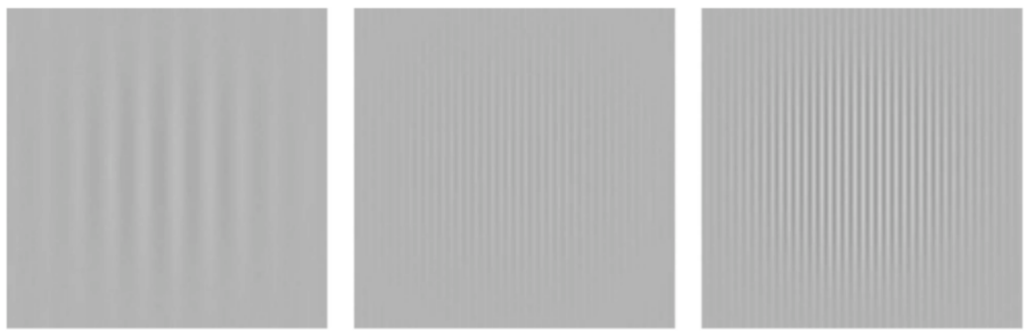
April 18, 2018

C.-C. Jay Kuo

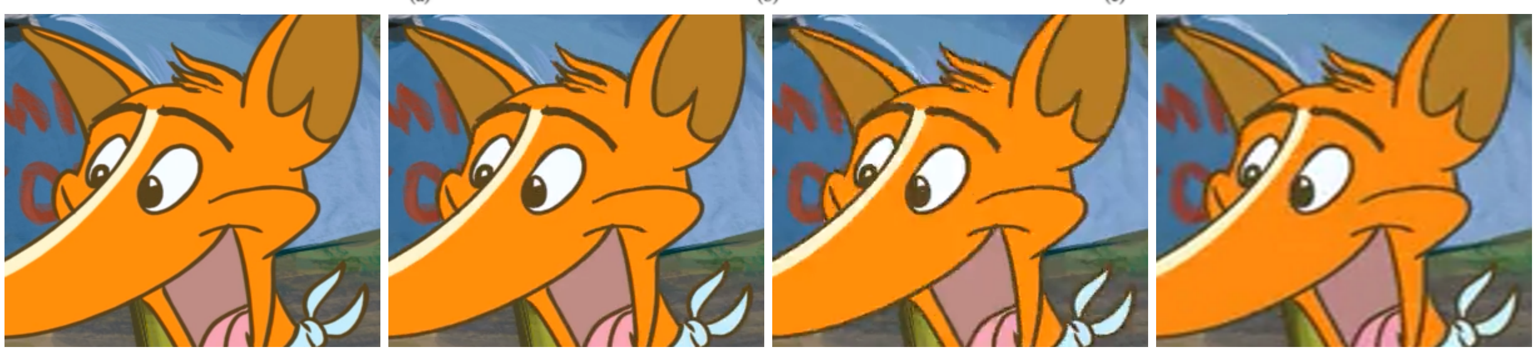
University of Southern California

Just Noticeable Difference (JND)

- Distortion visibility threshold
- CSF, luminance adaption, masking effect
- On real image/video rather than predefined patterns



(a) (b) (c)

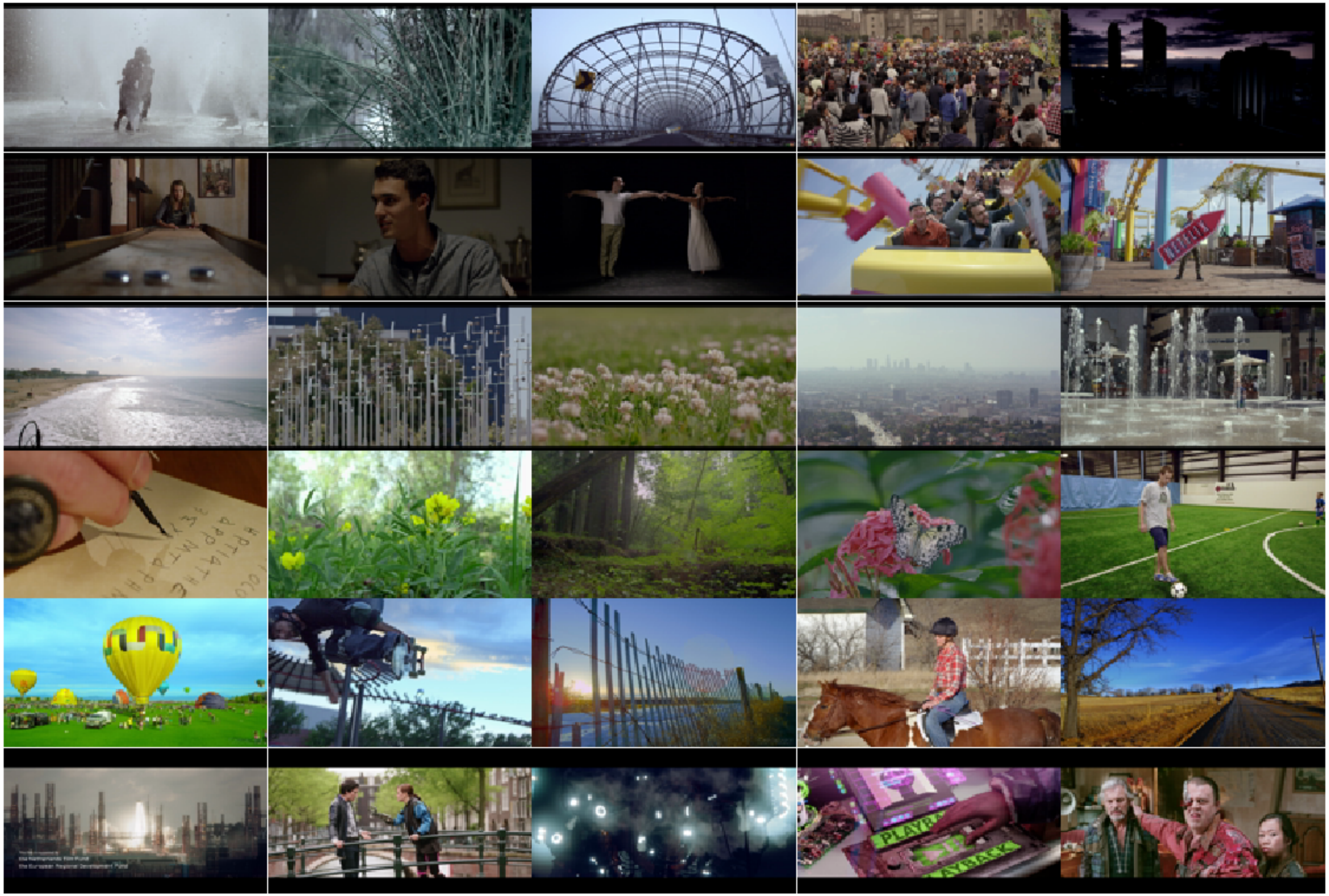


1. Anchor 2. Non-noticeable 3. Just Noticeable 4. Noticeable

VideoSet

- A Large Scale Subjective Test
  - Held in Shenzhen, China, 2016
  - 220 sequences, 5 seconds
  - 4 resolutions (1080p, 720p, 540p, 360p)
  - 3 JND points and around 35 samples per sequence-resolution
- Co-sponsored by
  - Netflix, Huawei, Samsung and Mediatek
- Available to the public:
  - https://ieee-dataport.org/documents/videoset

Representative Thumbnails from VideoSet



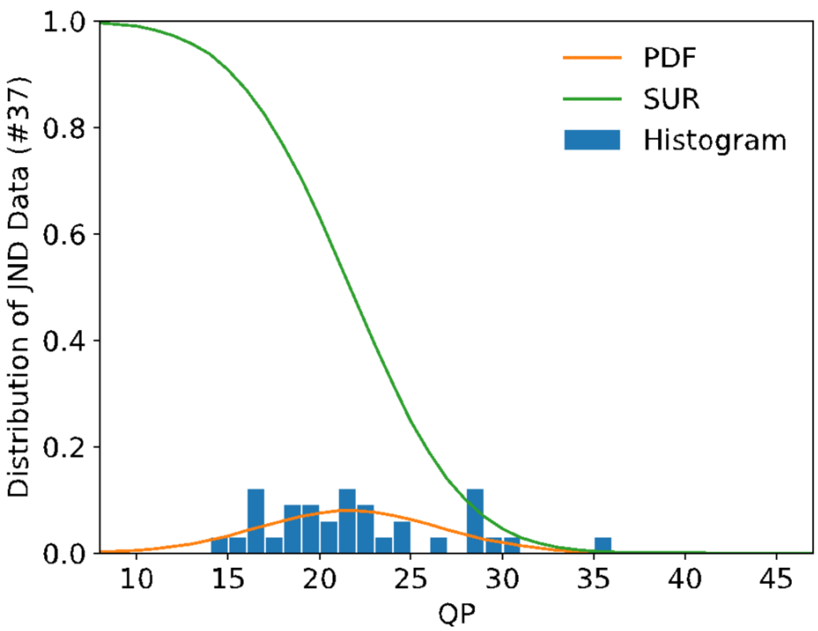
Properties of JND Points

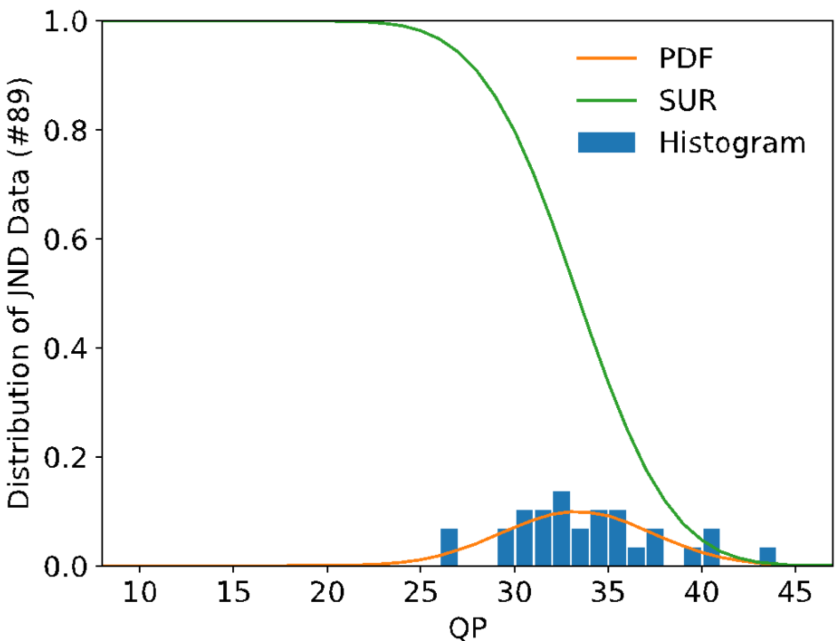
- JND position (QF or QP value) is a random variable
  - Follows the Gaussian distribution (Normality test)
  - Uniquely determined by the mean and variance
  - Integration of the JND curve gives the satisfied user ratio (SUR)

Resolution	The first JND (%)	The second JND (%)	The third JND (%)
1080p	95.9	95.9	93.2
720p	94.1	98.2	95.9
540p	94.5	97.7	96.4
360p	95.9	97.7	95.5

Explanation with Visual Examples

- From JND to SUR
  - The JND histogram (in blue)
  - The smoothed PDF curve (in orange)
  - The SUR curve (in green)





Universal Quality Metric

- We cannot compare quality measures for different video contents
  - Sequence A: 30 dB
  - Sequence B: 25 dB
  - Is Sequence A better than Sequence B?
- Satisfied User Ratio (SUR)
  - A viewer is satisfied if the compressed video appears to be perceptually the same as the reference
  - Ratio of users that a compressed video satisfies
- Change quality measures from PSNR to SUR

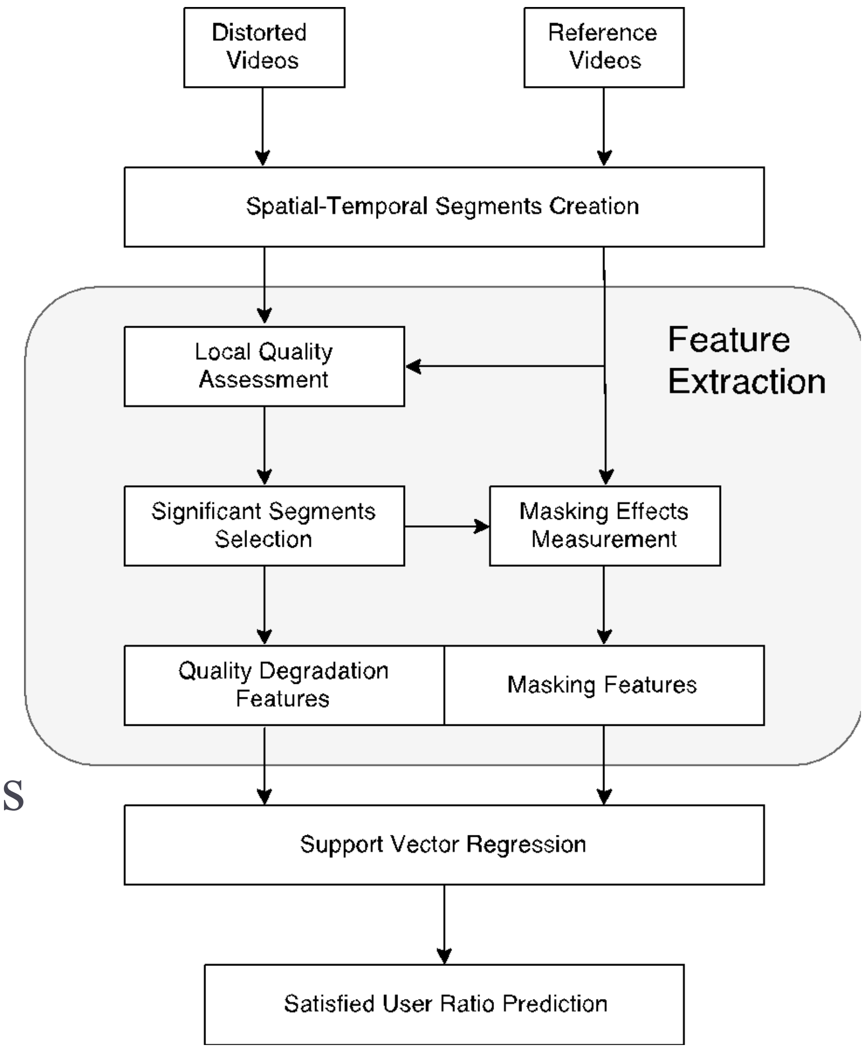
Strategy for SUR/JND Prediction

- Idea #1: What to predict?
  - Predicting the SUR curve first
  - Deriving the JND point accordingly

- Idea #2: How to predict SUR?
  - Quality degradation caused by coding – a full reference approach
  - Masking effect – a property of source video

Proposed SUR Prediction System

1. Spatial-Temporal Segments
   - 320x180x0.5s
2. Local Quality Assessment
   - VMAF
3. Significant Segments Selection
   - Not all segments contribute equally
4. Masking Effect Measure
   - Spatial and temporal randomness
5. Support Vector Regression
   - 5-fold validation



1. Spatial-Temporal Segments Creation

- Zoom in local quality
  - Rooted in psycho-visual studies
  - Dimension: 320x180x0.5s
  - Spatial dimension should be large enough for eye pursuit
  - Temporal dimension should be small enough to represent local quality

- Neighbor segments overlap 50% in spatial domain

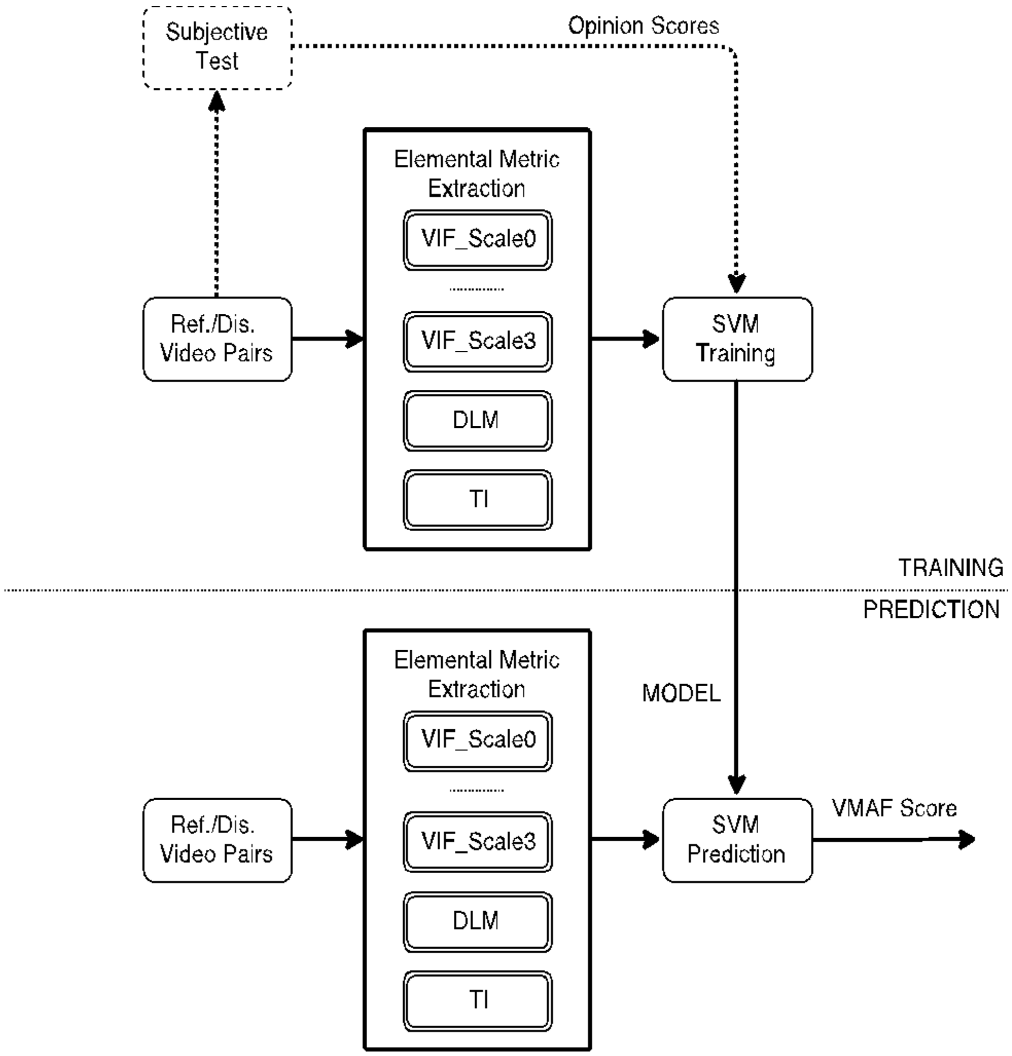
2. VMAF for Local Quality Assessment

- VMAF: Video Multimethod Assessment Fusion
- State-of-the-art VQA metric
- Developed and open-sourced by Netflix (collaboration with USC)
- Full reference video quality metric
- Combine multiple elementary quality metrics
- SVM to predict the final score

VMAF
Li, Zhi, Anne Aaron, Ioannis Katsavounidis, Anush Moorthy and Megha Manohara. "Toward a practical perceptual video quality metric." The Netflix Tech Blog 6 (2016).

VMAF for Local Quality Assessment

- Elemental Metric Scores
  - VIF: Visual Information Fidelity, information loss in wavelet domain
  - DLM: Detail Loss Measure, measuring the loss of details
  - TI: Temporal Information, temporal difference of adjacent frames



Visual Information Fidelity
H. Sheikh and A. Bovik, “Image Information and Visual Quality,” IEEE Transactions on Image Processing, vol. 15, no. 2, pp. 430–444, Feb. 2006.

Detail Loss Measure
S. Li, F. Zhang, L. Ma, and K. Ngan, “Image Quality Assessment by Separately Evaluating Detail Losses and Additive Impairments,” IEEE Transactions on Multimedia, vol. 13, no. 5, pp. 935–949, Oct. 2011.

Comparison of Video Quality Indices (1)

- Performance Comparison
  - PSNR, SSIM, FastSSIM, PSNR-HVS, VQM-VFD
  - SRCC, PCC, RMSE

NETFLIX-TEST Dataset (Compression + Scaling)

Metric	SRCC	PCC	RMSE
PSNR	0.746	0.725	24.577
SSIM	0.603	0.417	40.686
FastSSIM	0.685	0.605	31.233
PSNR-HVS	0.845	0.839	18.537
VQM-VFD	0.949	0.934	11.967
VMAF 0.3.1	**0.953**	**0.963**	**9.277**

LIVE Dataset (compression only impairments)

Metric	SRCC	PCC	RMSE
PSNR	0.416	0.394	16.934
SSIM	0.658	0.618	12.340
FastSSIM	0.566	0.561	13.691
PSNR-HVS	0.589	0.595	13.213
VQM-VFD	**0.763**	**0.767**	**9.897**
VMAF 0.3.1	0.690	0.655	12.180

Comparison of Video Quality Indices (2)

- Performance Comparison
  - PSNR, SSIM, FastSSIM, PSNR-HVS, VQM-VFD
  - SRCC, PCC, RMSE

VQEGHD3 Dataset (streaming impairments)

Metric	SRCC	PCC	RMSE
PSNR	0.772	0.759	0.738
SSIM	0.856	0.834	0.621
FastSSIM	0.910	0.922	0.415
PSNR-HVS	0.858	0.850	0.580
VQM-VFD	0.925	0.924	0.420
VMAF 0.3.1	**0.929**	**0.939**	**0.372**

LIVE Mobile Dataset

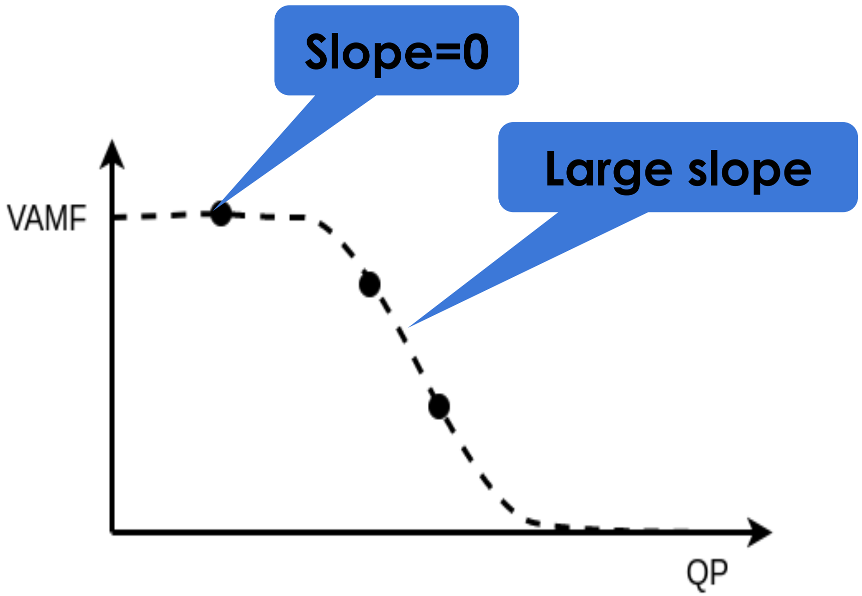
Metric	SRCC	PCC	RMSE
PSNR	0.632	0.643	0.850
SSIM	0.664	0.682	0.831
FastSSIM	0.747	0.745	0.718
PSNR-HVS	0.703	0.726	0.722
VQM-VFD	0.770	0.795	0.639
VMAF 0.3.1	**0.872**	**0.905**	**0.401**

3. Significant Segments Selection

- Also known as “Spatial/Temporal Pooling”
- Not all segments contribute equally to the final score
  - Salient regions? regions with lowest score? adaptive weighting?
- Slop of local quality score

$$\delta V(S_i) = \frac{V(S_{i-k}) - V(S_i)}{k},$$

V: VMAF score,
S: segment,
i: current QP,
k=2 (QP difference)



Quality Comparison: Seq #37

- Video content: still camera, dark background, salient male speaker

Quality Comparison: Seq #89

- Video content: Moving object, water drops in the background

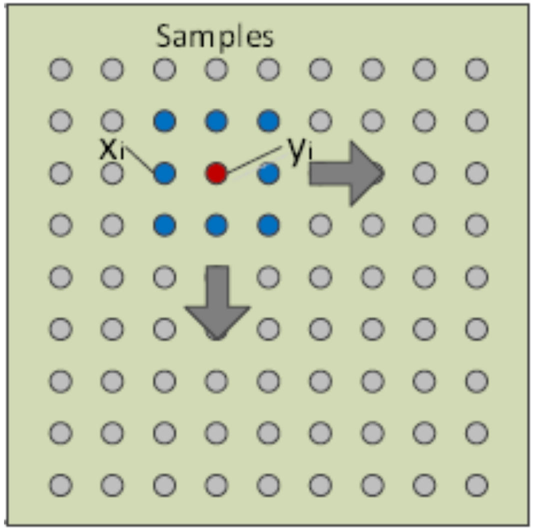
4. Masking Effects Measurement

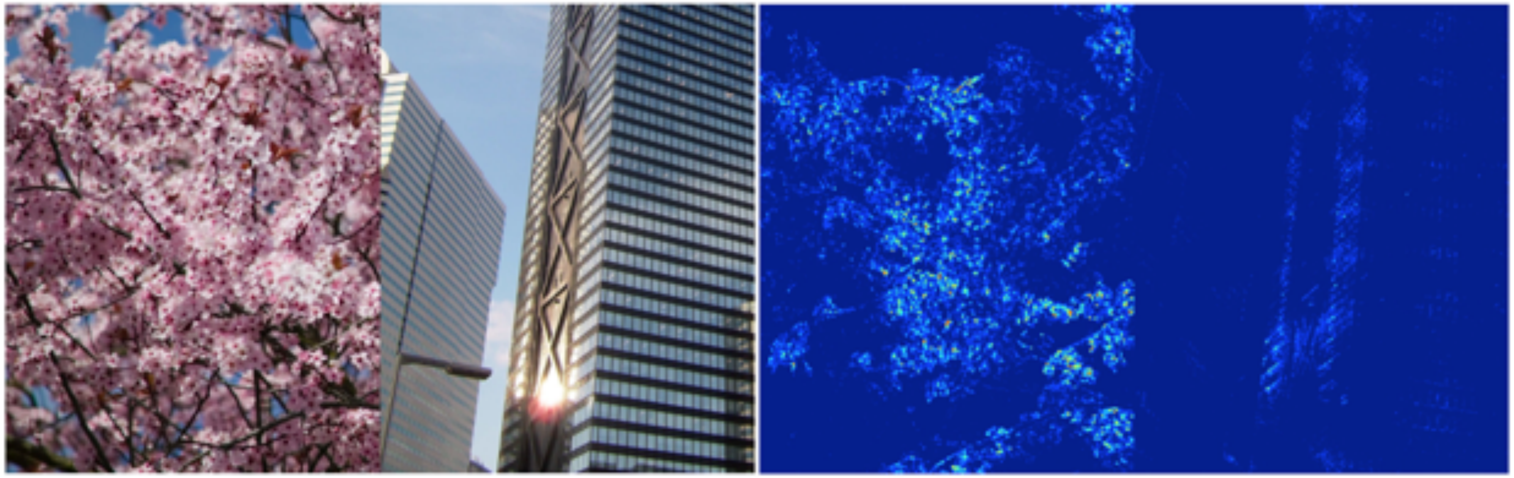
- Spatial randomness

- Temporal randomness

Spatial Randomness

- ➢ Spatial masking effect (spatial randomness)
- ➢ Predict center pixels from neighborhood





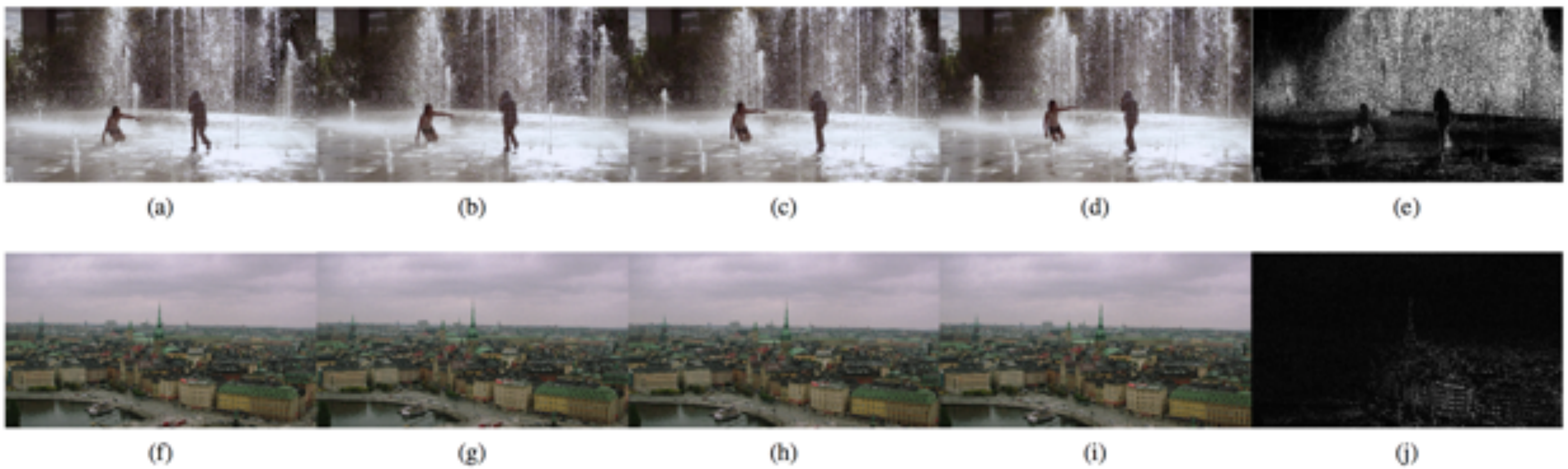
Original Image

Spatial Randomness

Hu, Sudeng, Lina Jin, Hanli Wang, Yun Zhang, Sam Kwong, and C-C. Jay Kuo. "Compressed image quality metric based on perceptually weighted distortion." *IEEE TIP,* 2015

Temporal Randomness Map (TRM)

- ➢ Temporal masking effect (temporal randomness)
- ➢ Motion regularity between frames



Hu, Sudeng, Lina Jin, Hanli Wang, Yun Zhang, Sam Kwong, and C-C. Jay Kuo. "Objective video quality assessment based on perceptually weighted mean squared error." *IEEE TCSVT,* 2017

5. Support Vector Regression

- Feature Representation
  - Feature dimension 40
  - 20-D local quality degradation feature vector
  - 20-D mask feature vector

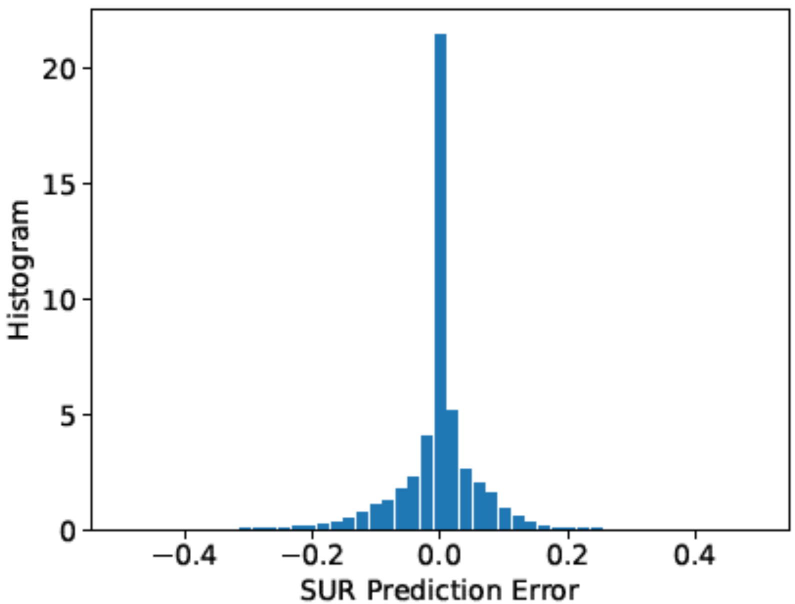
    - Spatial randomness: 10 dimensions
    - Temporal randomness: 10 dimensions

Quality Degradation Feature	Spatial Randomness	Temporal Randomness

- Support Vector Regression
  - Epsilon-SVR with radial basis function (rbf) kernel
  - 5-fold validation

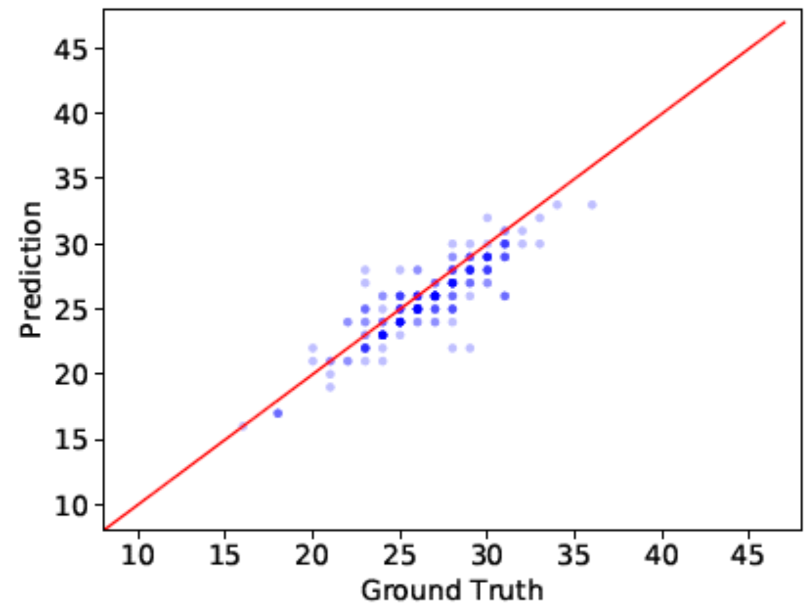
Experimental Results (1)

Prediction errors for 720p video



(a)

SUR Prediction Result



(b)

QP Prediction Result

Experimental Results (2)

- The first JND of 4 resolutions
  - ΔSUR: MAE of predicted SUR
  - ΔQP: MAE of predicted JND location (target at SUR=0.75)

	1080p	720p	540p	360p
ΔSUR	0.039	0.038	0.037	0.042
ΔQP	1.218	1.273	1.345	1.605

- Error increases as resolution becomes lower
- Probably due to fixed segment dimension

Future Work

- Prediction of the 2nd and the 3rd SUR curves and JND points
- Application to real world video streaming

Related Publications

- **JND Prediction**

Qin Huang, Haiqiang Wang, Sung Chang Lim, Hui Yong Kim, Se Yoon Jeong, C.-C. Jay Kuo, "Measure and Prediction of HEVC Perceptually Lossy/Lossless Boundary QP Values," DCC 2017.

- **VideoSet dataset**

Haiqiang Wang, Ioannis Katsavounidis, Jiantong Zhou, Jeonghoon Park, Shawmin Lei, Xin Zhou, Man-On Pun, Xin Jin, Ronggang Wang, Xu Wang, Yun Zhang, Jiwu Huang, Sam Kwong, C.-C. Jay Kuo. "VideoSet: A large-scale compressed video quality dataset based on JND measurement," JVCIR 2017.

- **Spatial and temporal randomness**

Hu, Sudeng, Lina Jin, Hanli Wang, Yun Zhang, Sam Kwong, and C-C. Jay Kuo. "Compressed image quality metric based on perceptually weighted distortion." TIP, 2015

- **Spatial and temporal randomness**

Hu, Sudeng, Lina Jin, Hanli Wang, Yun Zhang, Sam Kwong, and C-C. Jay Kuo. "Objective video quality assessment based on perceptually weighted mean squared error." TCSVT, 2017

Related Publications (2)

- **MCL-JCV dataset**

Haiqiang Wang, Weihao Gan, Sudeng Hu, Joe Yuchieh Lin, Lina Jin, Longguang Song, Ping Wang, Ioannis Katsavounidis, Anne Aaron and C.-C. Jay Kuo, “MCL-JCV: a JND-based H.264/AVC video quality assessment dataset,” ICIP, 2016.

- **GMM-based stair quality model**

Sudeng Hu, Haiqiang Wang and C.-C. Jay Kuo, “A GMM-based stair quality model for human perceived JPEG images,” ICASSP, 2016.

- **MCL-JCI dataset**

Lina Jin, Joe Yuchieh Lin, Sudeng Hu, Haiqiang Wang, Ping Wang, Ioannis Katsavounidis, Anne Aaron and C.-C. Jay Kuo. “Statistical Study on Perceived JPEG Image Quality via MCL-JCI Dataset Construction and Analysis,” HVEI, 2016.

- **JND-based quality measure of coded image/video**

Joe Yuchieh Lin, Lina Jin, Sudeng Hu, Ioannis Katsavounidis, Anne Aaron and C.-C. Jay Kuo. “Experimental Design and Analysis of JND Test on Coded Image/Video.” SPIE, 2015.