Prediction of Satisfied User Ratio for Compressed Video

Haiqiang Wang*, Ioannis Katsavounidis†, Qin Huang*, Xin Zhou‡, and C.-C. Jay Kuo*
* University of Southern California, Los Angeles, California, USA, † Netflix, Los Gatos, California, USA,
‡ Northwestern Polytechnical University, Xi'an, China.

USC Viterbi School of Engineering

Problem

We say that a viewer is satisfied if a compressed video appears to be perceptually the same as the reference. The phenomena is modeled by Just Noticeable Difference (JND), which is the transitional index that lies on the boundary of perceptually lossless and lossy visual experience for a subject. Satisfied User Ratio (SUR) of video clip d_i can be expressed as

$$S_i = 1 - \frac{1}{M}\sum_{m=1}^{M} \mathbb{1}_m(d_i), \qquad (1)$$

where M is the total number of subjects and $\mathbb{1}_m(d_i) = 1$ or 0 if the mth subject can or cannot see the difference between compressed clip d_i and its reference, respectively.

This is a challenging problem due to these aspects.

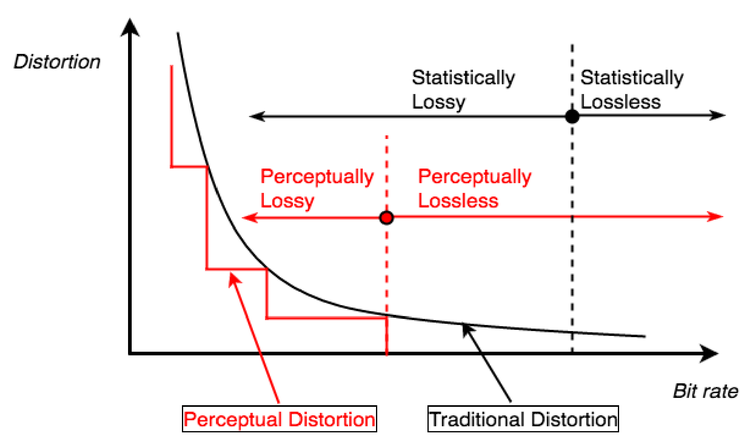
1. Diversified video contents
2. Limited understanding of the Human Visual System (HVS)
3. Subjective opinions vary from person to person

Contributions

We constructed a large-scale video quality dataset called the VideoSet [1] to measure human subjective experience of H.264/AVC coded video in terms of the just-noticeable-difference (JND). We propose a method to predict the satisfied-user-ratio (SUR) curves using a machine learning framework.

1. A consistent visual quality metric
2. Construction of a large-scale JND-based video quality dataset
3. A method to predict SUR curves with machine learning tool

Background



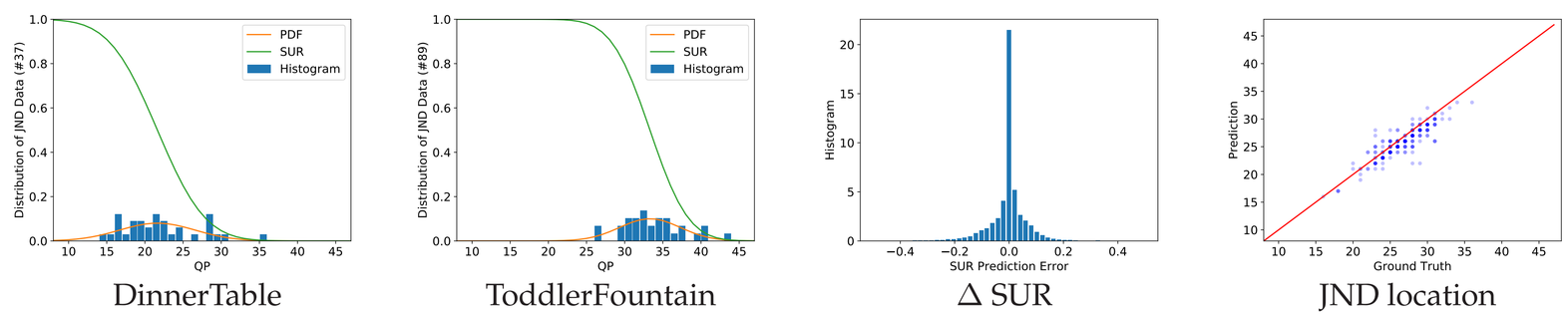
The peak signal-to-noise ratio (PSNR) has been used as an objective measure in video coding standards for years, it is generally agreed that it is a poor visual quality metric that does not correlate with human visual experience well.

Objective Video Quality Assessment (VQA) metrics try to address this problem. Typically, calculating the similarity/distance between compressed video and reference.

Visual Examples

- DinnerTable: the masking effect is weak (still camera, salient face region and dark background) and the JND point arrives earlier (i.e. a smaller QP value).
- ToddlerFountain: the masking effect is strong (water drops in background, fast motion). Compression artifacts are difficult to perceive.

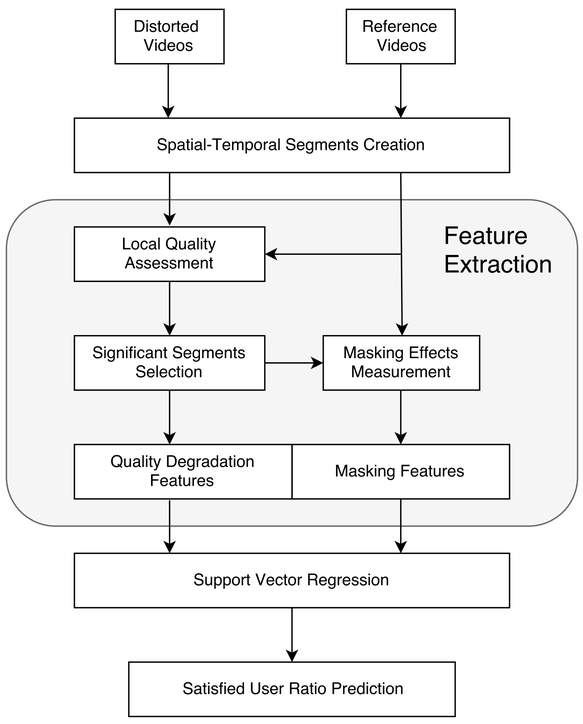
Results



DinnerTable | ToddlerFountain | Δ SUR | JND location

- DinnerTable and ToddlerFountain: SUR modeling from JND samples. The JND histogram (in blue), the smoothed PDF curve (in orange) and the SUR curve (in green)
- Δ SUR and JND location: the histogram of SUR prediction error and the predicted VS. the ground truth JND location (with % 75 SUR).

Method



The SUR curve is primarily determined by two factors: 1) quality degradation due to compression and 2) the masking effect. When a subject evaluates a pair of video clips, different spatial-temporal segments of the two video clips are successively assessed.

The proposed method takes both the local quality degradation (VMAF [2]) as well as the masking effect ([3,4]) into consideration. It extracts a compact feature vector and feeds it into the support vector regressor (SVR) to obtain the predicted SUR curve.

Discussion

Summary of averaged prediction errors for video clips in four resolutions.

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

The system achieves good performance in all resolutions. The averaged prediction errors were summarized above. We see that prediction errors increase as the resolution becomes lower. This is probably due to the use of fixed dimensions in generating spatial-temporal segments. We will adopt the same framework to predict locations of the second and the third JND points in the near future.

Reference

[1] Haiqiang Wang, et al., Videoset: A Large-scale Compressed Video Quality Dataset Based on JND Measurement. In *JVCIR, 2017.*

[2] Zhi Li, et al., Toward a practical perceptual video quality metric. *The Netflix Tech Blog, 2016.*

[3] Sudeng Hu, et al., Compressed image quality metric based on perceptually weighted distortion. *TIP, 2015.*

[4] Sudeng Hu, et al., Objective Video Quality Assessment based on Perceptually Weighted Mean Squared Error. *TCSVT, 2016.*