

Transcoding quality based on SSIM for DCT-based compressed images and video

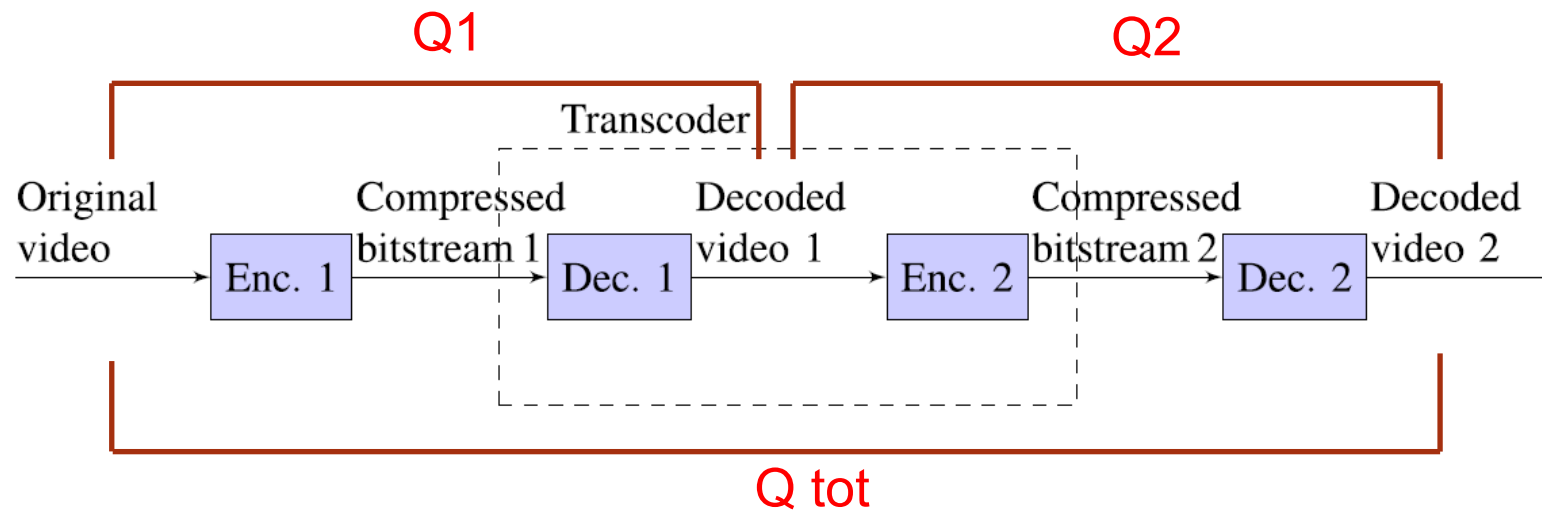
Prof Maria Martini
Kingston University London

VQEG meeting May 2025

Kingston
University
London

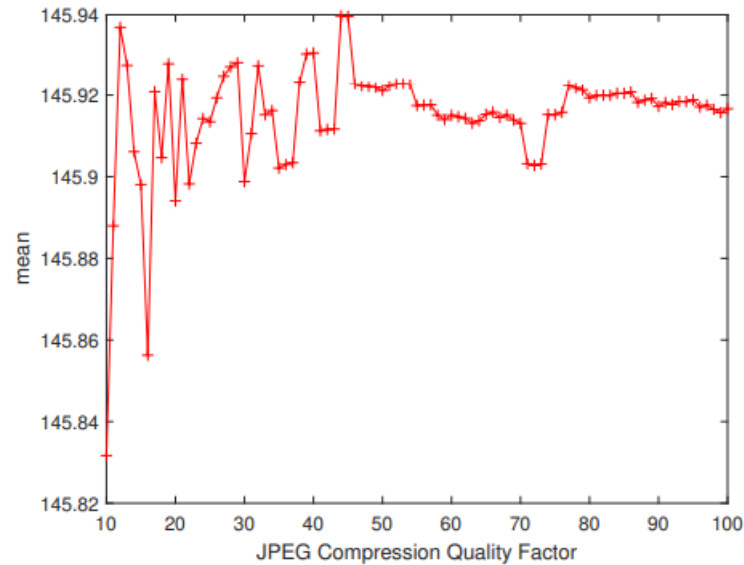
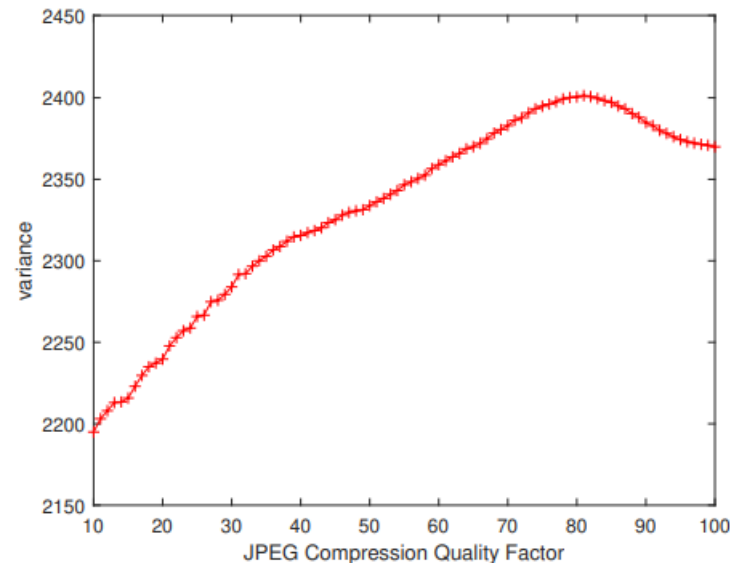
Motivation

- Transcoding: conversion of already compressed video into a different format / using a different encoder (e.g., to create multiple representations in video streaming)
- Transcoding quality via objective quality metrics is mainly calculated with **reference to the already encoded source image/video**, but this does not provide information on the actual quality of the transcoded video vs. **original** uncompressed one
- The calculation of the total quality is not obvious, in particular for metrics such as SSIM

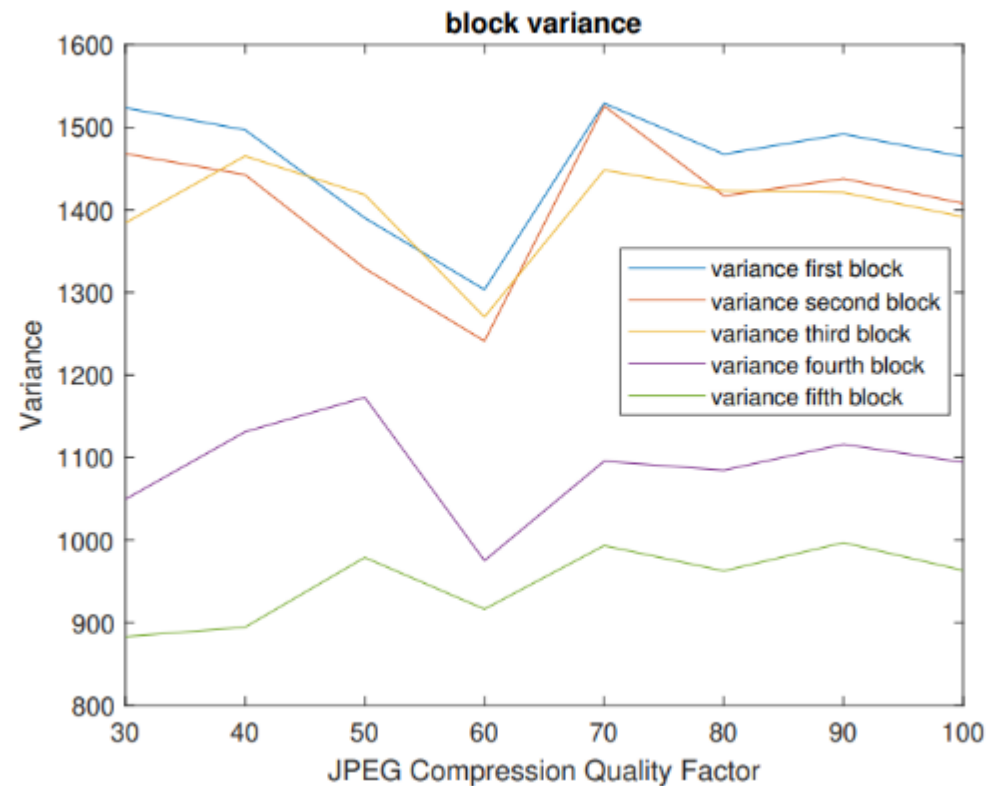


Assumptions for DCT compressed images / video

$$\mu_e = 0, \mu_x = \mu_y \text{ and } \sigma_x^2 = \sigma_y^2$$

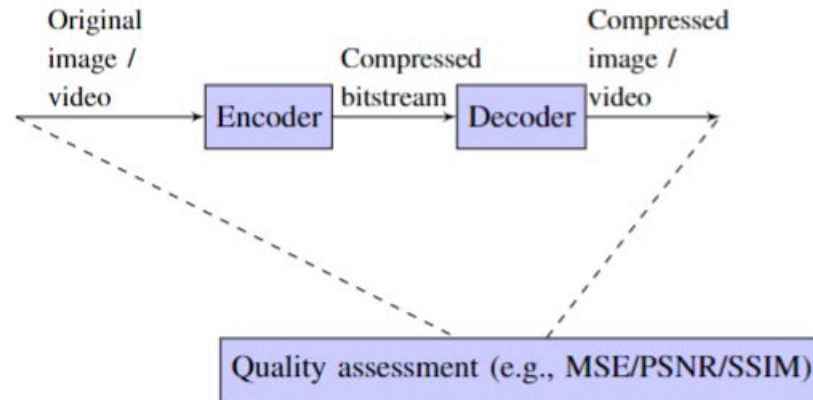


Assumption on variance - Block variance



Example for *Baboon* image, first five blocks

Reminder



$$PSNR = 10 \log_{10} \frac{(2^b - 1)^2}{MSE}$$

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

Relationship derivation part 1 (SSIM vs. MSE)

With these assumptions: [reminder: $SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$]

$$SSIM(x, y) = \frac{(2\mu_y^2 + C_1)(2\sigma_{xy} + C_2)}{(2\mu_y^2 + C_1)(2\sigma_y^2 + C_2)}$$

Assumption of
equal mean
and equal
variance

$$SSIM(x, y) = \frac{(2\sigma_{xy} + C_2)}{(2\sigma_y^2 + C_2)}$$

$$e = x - y \quad \sigma_e^2 = \sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}$$

$$2\sigma_{xy} = \sigma_x^2 + \sigma_y^2 - \sigma_e^2 = 2\sigma_y^2 - \sigma_e^2$$

$$SSIM(x, y) = \frac{(2\sigma_y^2 - \sigma_e^2 + C_2)}{(2\sigma_y^2 + C_2)} = 1 - \frac{\sigma_e^2}{(2\sigma_y^2 + C_2)}$$

With the previous assumption $\mu_e = 0$ we have

$$\sigma_e^2 = \mathcal{E}(e^2) = MSE$$

Note: with assumption of equal
mean only

$$SSIM(x, y) = 1 - \frac{MSE}{(2\sigma_y^2 + C_2)}$$

$$SSIM(x, y) = 1 - \frac{MSE}{(\sigma_x^2 + \sigma_y^2 + C_2)}$$

$$PSNR = 10 \log_{10} \frac{(2^b - 1)^2}{MSE}$$

$$SSIM(x, y) = 1 - \frac{(2^b - 1)^2}{10^{PSNR/10} (2\sigma_y^2 + C_2)}$$

$$MSSIM(X, Y) = \frac{1}{M} \sum_j \left[1 - \frac{(2^b - 1)^2}{10^{PSNR_j/10} (2\sigma_{y_j}^2 + C_2)} \right]$$

$$MSSIM(X, Y) = \frac{1}{M} \sum_j SSIM(x_j, y_j)$$

Relationship derivation part 2

Mean

$$\mathcal{E}(e_1) = \frac{1}{N} \sum_{i=1}^N e_{1i} \quad \text{Encoder 1}$$

$$\mathcal{E}(e_2) = \frac{1}{N} \sum_{i=1}^N e_{2i} \quad \text{Encoder 2}$$

$$\mathcal{E}(e) = \frac{1}{N} \sum_{i=1}^N e_i. \quad \text{Total}$$

The encoding error of a DCT-based video codec has approximately zero mean, Hence all above are approx. 0.

Variance

$$\sigma_{e_1}^2 = \mathcal{E}(e_1^2) - \mathcal{E}^2(e_1) = \mathcal{E}(e_1^2) = \text{MSE}_1$$

$$\sigma_{e_2}^2 = \mathcal{E}(e_2^2) - \mathcal{E}^2(e_2) = \mathcal{E}(e_2^2) = \text{MSE}_2$$

$$\sigma_e^2 = \mathcal{E}(e^2) - \mathcal{E}^2(e) = \mathcal{E}(e^2) = \text{MSE}_{\text{tot}}.$$

The variance of the total error can be written as

$$\sigma_e^2 = \sigma_{e_1+e_2}^2 = \sigma_{e_1}^2 + \sigma_{e_2}^2 + 2\sigma_{e_1 e_2}$$

where the last term is the covariance.

With the assumption that e_1 and e_2 are independent

$$\sigma_e^2 = \sigma_{e_1}^2 + \sigma_{e_2}^2$$

$$\text{SSIM}_1 = 1 - \frac{\text{MSE}_1}{(2\sigma_y^2 + C_2)}$$

$$\text{SSIM}_2 = 1 - \frac{\text{MSE}_2}{(2\sigma_y^2 + C_2)}$$

$$\text{MSE}_{\text{tot}} = \text{MSE}_1 + \text{MSE}_2$$

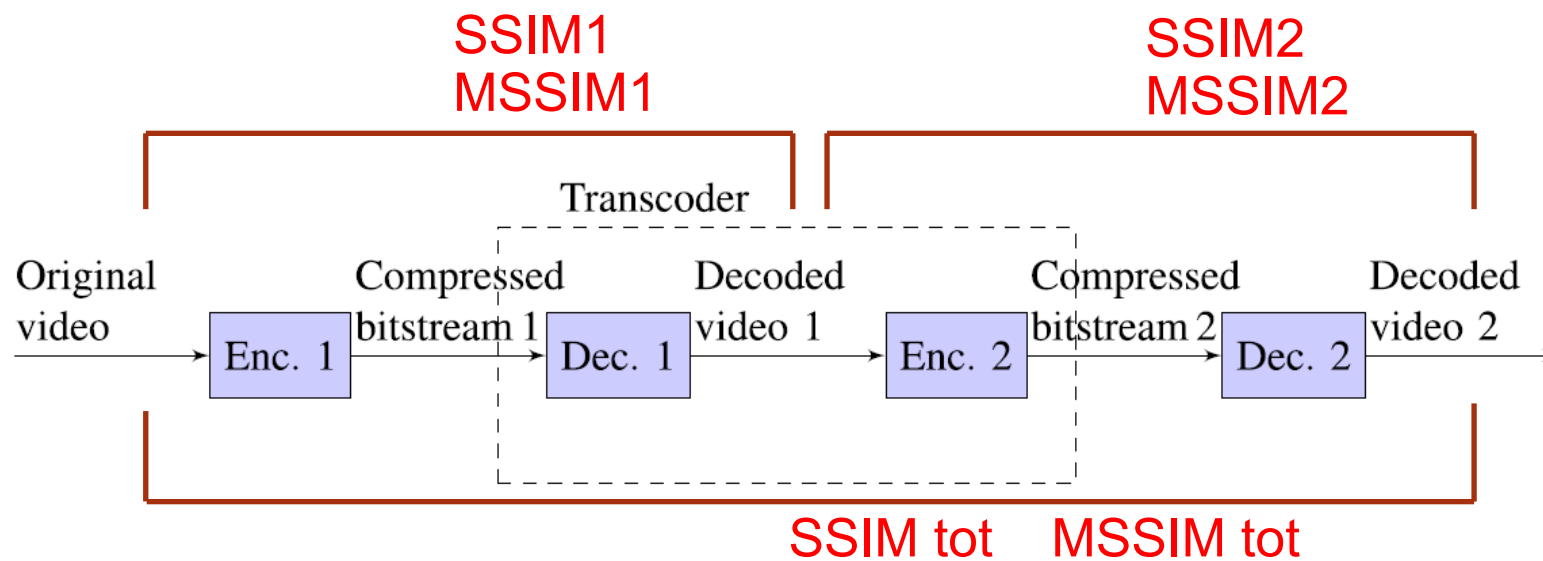
$$\text{SSIM}_{\text{tot}} = 1 - \frac{\text{MSE}_{\text{tot}}}{(2\sigma_y^2 + C_2)} = 1 - \frac{\text{MSE}_1 + \text{MSE}_2}{(2\sigma_y^2 + C_2)}$$

$$\begin{aligned} \text{SSIM}_{\text{tot}} &= 1 - \frac{\text{MSE}_1}{(2\sigma_y^2 + C_2)} - \frac{\text{MSE}_2}{(2\sigma_y^2 + C_2)} \\ &= \text{SSIM}_1 + \text{SSIM}_2 - 1. \end{aligned}$$

Final relationship:

$$\text{SSIM tot} = \text{SSIM1} + \text{SSIM2} - 1$$

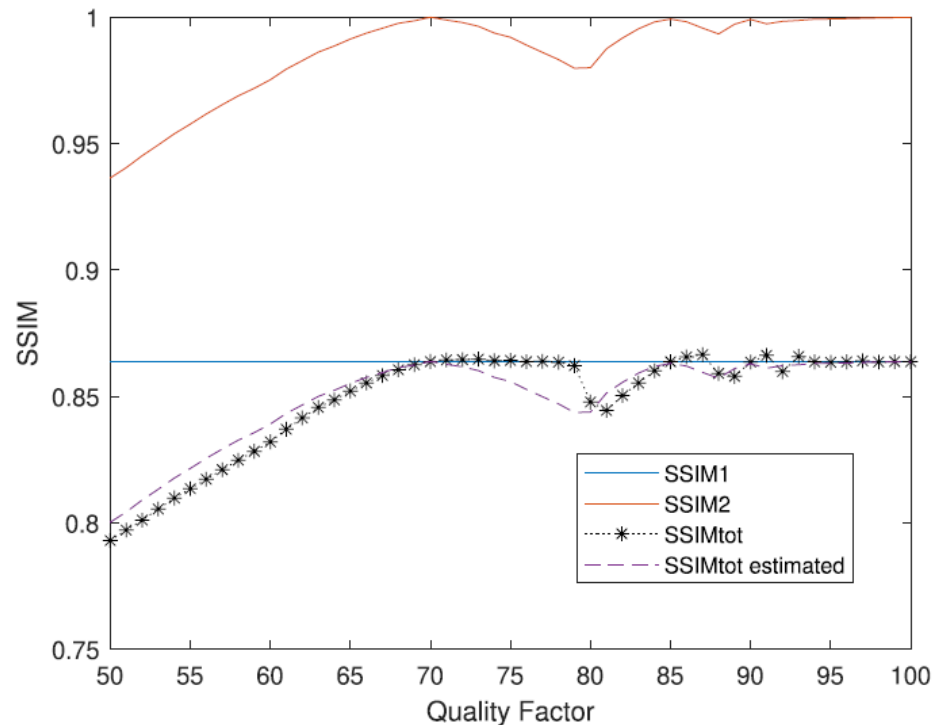
$$\text{MSSIM tot} = \text{MSSIM1} + \text{MSSIM2} - 1$$



Note: MSSIM1 can be read from metadata when available or estimated in no-reference or reduced-reference modality based on bitrate and other information (in metadata / in the bitstream, such as motion vectors and QP values).

Example

SSIM for the Baboon image, encoded first with JPEG at QF 70, then transcoded with QF from 50 to 100, default Gaussian window.



Discussion:
High accuracy

It was assumed that:

- the mean of the encoding error in first encoder and transcoder is zero
- the coding errors produced by the two encoders are independent

The second assumption may be the main responsible of the minor inaccuracies

Conclusion

- Simple relationship between SSIM based quality of original and transcoded image/video
- Assumptions reasonably good for quality range of typical interest
- Useful relationship for optimization problems
- High accuracy

Associated paper:

M.G. Martini, "Measuring objective image and video quality: On the relationship between SSIM and PSNR for DCT-based compressed images." *IEEE Transactions on Instrumentation and Measurement*, vol. 74, Jan 2025.

Transcoding quality based on SSIM for DCT-based compressed images and video

Thank you

Prof Maria Martini
Kingston University London

VQEG meeting May 2025