On the relationship between SSIM and PSNR for DCT-based compressed images and video: SSIM as content-aware PSNR

Maria Martini [1]

[1]Affiliation not available

October 30, 2023

Abstract

The aim of this paper is to show how the structural similarity metric SSIM for image quality assessment can be seen in many cases, such as DCT-based compressed images and video, as a content-aware version of the peak signal-to-noise ratio (PSNR). In fact, under some assumptions described in the paper, the first can be derived directly from the latter based on a single content-dependent parameter, i.e. the variance of the image / video frame.

A simple relationship between SSIM and PSNR for DCT-based compressed images and video: SSIM as content-aware PSNR

Maria G. Martini
Wireless and Multimedia Networking Research Group
School of Computer Science and Mathematics
Kingston University, London, UK
m.martini@kingston.ac.uk

***Abstract*—The aim of this paper is to show how the structural similarity metric SSIM for image quality assessment can be seen in many cases, such as DCT-based compressed images and video, as a content-aware version of the peak signal-to-noise ratio (PSNR). In fact, under some assumptions described in the paper, the first can be derived directly from the latter based on a single content-dependent parameter, i.e. the variance of the image / video frame.**

***Index Terms*—Quality assessment, objective quality metrics, PSNR, SSIM, image and video compression**

I. Introduction

Lossy image/video compression results in a reduction of image/video quality, that can be assessed objectively based on quality metrics. Figure 1 shows via a block diagram the process of lossy compression and the relevant objective, full reference quality assessment based on original and reconstructed image or video.

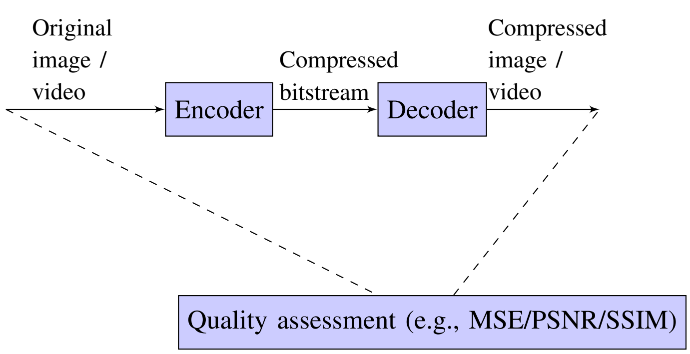

Fig. 1: Quality assessment as comparison between original and compressed video.

Focusing on a single image or video frame, the mean square error (MSE) between the original (X) and the compressed version (Y) is calculated as follows:

$$MSE = \frac{1}{MN} \sum_{j=1}^{M \times N} (x_j - y_j)^2 = \mu_{e^2} \tag{1}$$

where M and N are the number of pixels in horizontal and vertical direction, e represents the error and μ is used to identify the mean.

The Peak Signal to Noise Ratio (PSNR) is derived from MSE as follows:

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

where b is the bit depth (number of bits per pixel).

While MSE and PSNR are easy to calculate, they are not always good indicators of the actual quality as perceived by the users. For this reason, other quality metrics have been developed. In particular, the structural similarity metric (SSIM) [1] has become very popular since it has been shown to correlate well with the quality as perceived by humans for different types of distortions.

The SSIM metric is defined as below:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{3}$$

where μ_x and μ_y are the mean of the original and the impaired grayscale image, respectively, while σ_x^2 and σ_y^2 are the variance of the original and the impaired grayscale image, respectively; σ_{xy} represents the covariance between original and impaired images (grayscale) and C_1 and C_2 are constant values to make sure the metric is a real number. A similar index can be calculated on the color components of an image.

SSIM is a full reference metric and it can be decomposed in terms addressing structure, contrast, and luminance comparison.

There have been a number of attempts to compare PSNR and SSIM and both metrics with the results of subjective tests summarised via the Mean Opinion Score (MOS). Raw approximations, not taking image / video content into account, resulted in tables (see e.g. [2]), used for the design and optimization of multimedia systems. For this purpose, a mathematically tractable metric is required, hence subjective quality metrics cannot be used directly and the same applies to objective quality metrics based on machine learning (e.g., Video Multimethod Assessment Fusion (VMAF) [3]). In this paper we propose to derive SSIM directly from PSNR or MSE in the case of DCT-based image and video compression. In this case, in fact, we can assume that the mean of the luminance/chrominance values do not vary with the compression ratio and the same assumption can be made for the variance.

The main contribution of this paper is a new analytical relationship between PSNR and SSIM for DCT-based compressed images and video. This involves a simplified way to calculate SSIM based on only MSE or PSNR and the variance of the impaired image as only content-dependent factor. To the author's knowledge, this is the first work to highlight that in the considered use case the two are related via a simple content dependent parameter, hence SSIM can be seen as a content-aware PSNR.

This model, being easily mathematically tractable can be used in the formulation of system optimization for the compression and transmission of images and video. An extra advantage of the method is that, assuming PSNR information but not SSIM is provided, for instance as metadata in a received bitstream, SSIM can be calculated easily with no need for other reference to the original video.

The remainder of this paper is organized as follows. After a summary of related work in Section II, an analysis of the relationship between SSIM and PSNR is presented in Section III, including examples to validate the assumptions and comparative results with an example image. Section IV concludes the work along with a brief discussion of potential applications.

II. Related work

The relationship between SSIM [1] and PSNR has been analysed in [4] [5] where analytical expressions and approximations are provided for different use cases. According to the approximations considered, the calculation of SSIM from PSNR still requires some joint processing of original and compressed image, beyond what required for PSNR. In particular, in [4] the proposed relationship involves the evaluation of the covariance between original and impaired image and a linear approximation is proposed for SSIM values between 0.2 and 0.8. Such relationship is explored with examples in [5]. This work also highlights how the luminance comparison component in SSIM has a marginal impact on the final value for the examples considered, while the structure comparison term has a higher impact than the contrast comparison one. The work in [6], after recognising that the SSIM expression can be simplified to a correlation coefficient in the case $\mu_x = \mu_y$, propose an alternative quality metric based on similar statistics. Other works (such as [7]) aim at estimating the SSIM metric in absence of a reference from bitstream and/or reconstructed video and compare the estimation with PSNR estimation, but not establishing a relationship between the two. Two of the authors of the SSIM metric discuss in [8] the properties of MSE and SSIM, but they do not focus explicitly on their relationship.

III. Analysis of the relationship between SSIM and PSNR

In order to identify a simple relationship between SSIM and PSNR, we assume that for DCT based compression distortion we have $\mu_e = 0, \mu_x = \mu_y$ and $\sigma_x^2 = \sigma_y^2$ [9] [10]. To show the validity of this assumption with an example, Figures 2 and 3 report mean and variance for the Baboon reference image (Figure 4) compressed according to the JPEG standard at different compression ratios (corresponding to the quality factors reported in the horizontal axis).

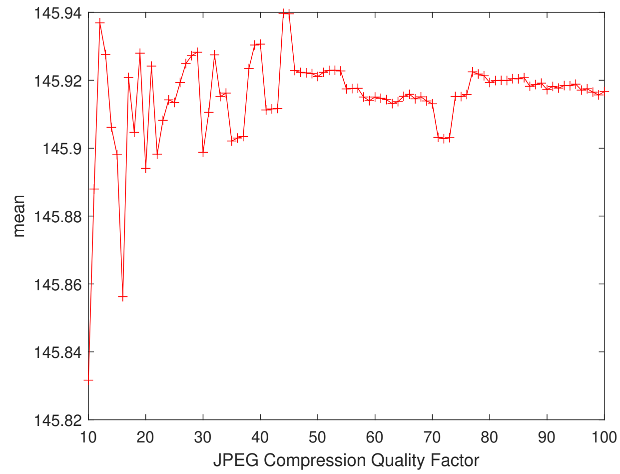

Fig. 2: Mean for the example Baboon image at different JPEG compression quality factors.

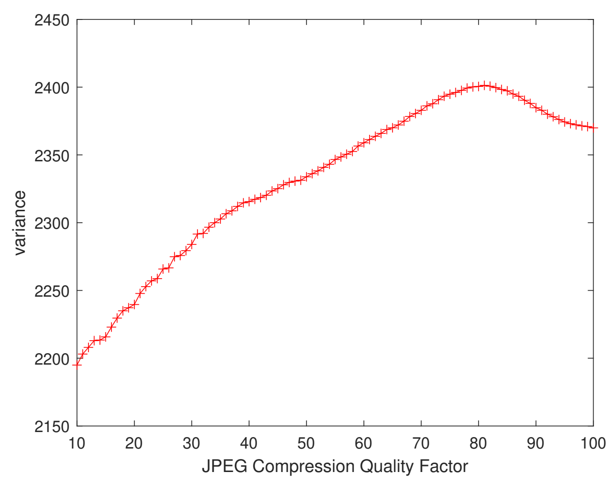

Fig. 3: Variance for the example Baboon image at different JPEG compression quality factors.

We can observe that indeed the mean is almost constant across a very wide range of compression ratios / quality factors (ranging from 145.83 to 145.94 with a variation within 0.075%) while the variation of the global variance is limited between 2200 and 2400 (variation within 9%).

With these assumptions we have from (3):

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_y^2 + C_1)(2\sigma_{xy} + C_2)}{(2\mu_y^2 + C_1)(2\sigma_y^2 + C_2)} \tag{4}$$

hence:

$$\mathrm{SSIM}(x,y) = \frac{(2\sigma_{xy} + C_2)}{(2\sigma_y^2 + C_2)}. \tag{5}$$

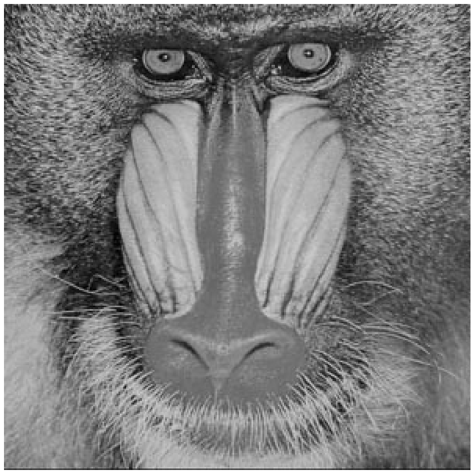

Fig. 4: Baboon grayscale image considered in the example.

We also have:

$$e = x - y \tag{6}$$

$$\sigma_e^2 = \sigma_x^2 + \sigma_y^2 - 2\sigma_{xy} \tag{7}$$

hence:

$$2\sigma_{xy} = \sigma_x^2 + \sigma_y^2 - \sigma_e^2 = 2\sigma_y^2 - \sigma_e^2. \tag{8}$$

Using (8) in (5):

$$\mathrm{SSIM}(x, y) = \frac{(2\sigma_y^2 - \sigma_e^2 + C_2)}{(2\sigma_y^2 + C_2)} = 1 - \frac{\sigma_e^2}{(2\sigma_y^2 + C_2)}. \tag{9}$$

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

$$\sigma_e^2 = \mathcal{E}(e^2) = MSE \tag{10}$$

hence:

$$\mathrm{SSIM}(x, y) = 1 - \frac{MSE}{(2\sigma_y^2 + C_2)}. \tag{11}$$

Considering PSNR

$$PSNR = 10log_{10}\frac{(2^b - 1)^2}{MSE} \tag{12}$$

we can write

$$MSE = \frac{(2^b - 1)^2}{10^{PSNR/10}} \tag{13}$$

and using (13) in (11) we have:

$$\mathrm{SSIM}(x, y) = 1 - \frac{\frac{(2^b-1)^2}{10^{PSNR/10}}}{(2\sigma_y^2 + C_2)}. \tag{14}$$

We note that this is also a way to calculate SSIM with reduced computational complexity: rather than needing to calculate the mean and variance of both original and impaired image, as well as the covariance, only the variance of the impaired image needs to be calculated, in addition to PSNR (14) or MSE (11). Since PSNR is content-independent, the variance of the impaired image is the only element taking into account the content of the image under assessment.

We observe that the global structural similarity index is typically calculated as the mean of the Structural Similarity (SSIM) of sub-windows composing the image:

$$\mathrm{MSSIM}(X, Y) = \frac{1}{M}\sum_j \mathrm{SSIM}(x_j, y_j) \tag{15}$$

where X and Y are the reference and the distorted images, respectively, x_j and y_j are the image contents at the $j - th$ local window, and M is the number of samples in the quality map [1].

Hence,

$$\mathrm{MSSIM}(X, Y) = \frac{1}{M}\sum_j [1 - \frac{\frac{(2^b-1)^2}{10^{PSNR_j/10}}}{(2\sigma_{yj}^2 + C_2)}]. \tag{16}$$

In order to verify the accuracy of the derivation with the assumptions considered, Figure 5 reports the comparison between the MSSIM value (blue curve) and the corresponding value obtained from PSNR / MSE values with (16) (red curve) for the Baboon example image compressed with JPEG. We observe that, for the typical quality range of interest, (16) approximates the SSIM score with high accuracy.

For a better understanding of the quality range corresponding to the considered quality factors in terms of PSNR, we report in Figure 6 the PSNR vs. JPEG quality factor for the same example image.

While these results are obtained as a mean across the image as in (16), we show local results in Figures 7 and 8. Figure 7 shows the comparison between actual SSIM and SSIM obtained with (14) for the first block of the Baboon image, while Figure 8 shows the scatter plot for local PSNR value vs. local MMSIM value for the same block of the same image.

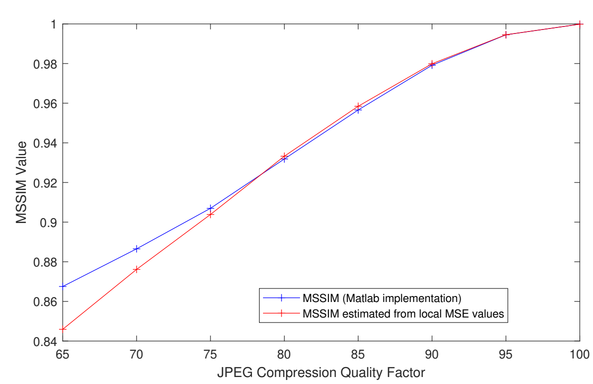

Fig. 5: MSSIM and MSSIM estimated from PSNR vs. JPEG compression Quality Factor.

We can observe how at local level the relationship is less accurate and more oscillating, but the approximation is still very good for high quality values (above PSNR 30).

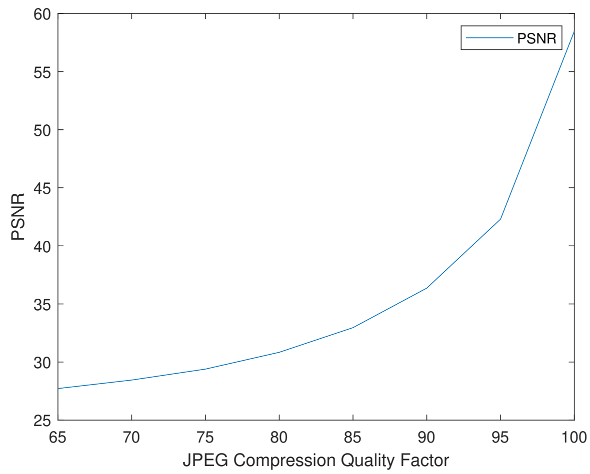

Fig. 6: PSNR vs. JPEG compression quality factor for the Baboon example image.

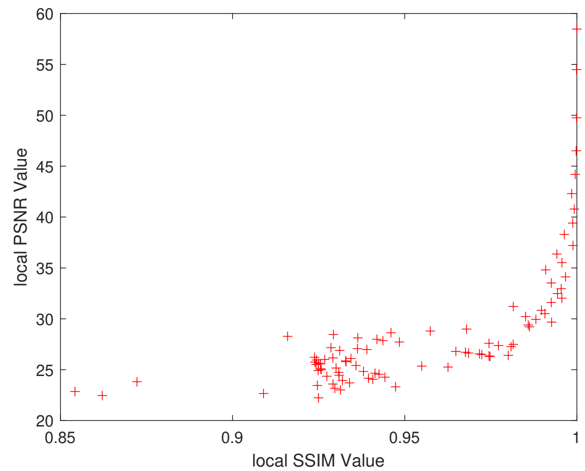

Fig. 8: PSNR vs. SSIM for the first block of the Baboon example image at different JPEG compression level (Quality Factor from 10 to 100).

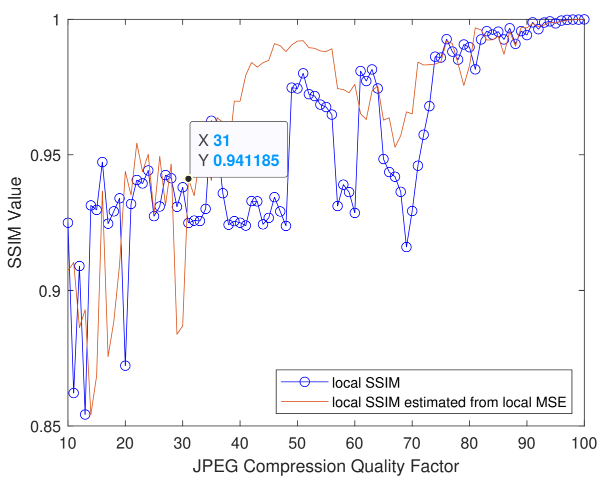

Fig. 7: SSIM vs. JPEG compression Quality Factor for the first block of the Baboon example image.

IV. Conclusion

This paper highlighted how the structural similarity metric SSIM for image quality assessment can be seen in many cases, such as DCT-based compressed images and video, as a content-aware version of the peak signal-to-noise ratio (PSNR) and SSIM can be obtained from PSNR via the variance of the impaired image/video, hence with no further reference to the original content. This is expected to support the optimization of image/video compression and transmission systems, enabling mathematically tractable quality based optimization based on just one parameter beyond MSE/PSNR. This will also support an easy comparison of different compression methods based on SSIM. In fact, BD-rate and BD-quality [11] were defined based on PSNR and their translation to different metrics such as SSIM is not obvious [12].

References

[1] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[2] T. Zinner, O. Abboud, O. Hohlfeld, T. Hossfeld, and P. Tran-Gia, "Towards QoE management for scalable video streaming," in *21th ITC specialist seminar on multimedia applications-traffic, performance and QoE*. Citeseer, 2010, pp. 64–69.

[3] Z. Li, C. Bampis, J. Novak, A. Aaron, K. Swanson, A. Moorthy, and J. Cock, "VMAF: The journey continues," *Netflix Technology Blog*, vol. 25, 2018.

[4] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *2010 20th international conference on pattern recognition*. IEEE, 2010, pp. 2366–2369.

[5] A. Horé and D. Ziou, "Is there a relationship between peak-signal-to-noise ratio and structural similarity index measure?" *IET Image Processing*, vol. 7, no. 1, pp. 12–24, 2013.

[6] G. Palubinskas, "Image similarity/distance measures: what is really behind MSE and SSIM?" *International Journal of Image and Data Fusion*, vol. 8, no. 1, pp. 32–53, 2017.

[7] T. Shanableh, "Prediction of structural similarity index of compressed video at a macroblock level," *IEEE Signal Processing Letters*, vol. 18, no. 5, pp. 335–338, 2011.

[8] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? a new look at signal fidelity measures," *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, 2009.

[9] J. Yang, G. Zhu, and Y.-Q. Shi, "Analyzing the effect of JPEG compression on local variance of image intensity," *IEEE Transactions on Image Processing*, vol. 25, no. 6, pp. 2647–2656, 2016.

[10] X. Shang, H. Zhao, G. Wang, X. Zhao, and Y. Zuo, "A novel objective quality assessment method for transcoded videos from H. 264/AVC to H. 265/HEVC utilizing probability theory," *IEEE Transactions on Broadcasting*, vol. 65, no. 4, pp. 777–781, 2019.

[11] G. Bjontegaard, "Calculation of average PSNR differences between rd-curves," *VCEG-M33*, 2001.

[12] N. Barman, M. G. Martini, and Y. Reznik, "Revisiting Bjontegaard delta bitrate (BD-BR) computation for codec compression efficiency comparison," in *Proceedings of the 1st Mile-High Video Conference*, 2022, pp. 113–114.