The ubiquitous Structural Similarity Index or SSIM can be dramatically simplified, straightened and re-interpreted as a normalized visibility function, or dissimilarity quotient. Explicitly writing SSIM in a symmetric formulation immediately reveals the previously enigmatic structural covariance as a difference of variances. A dramatic simplification ensues.

Although the SSIM was a milestone in the recent history of Image Quality Assessment (IQA) its immense success (in citation terms) has now become more of a millstone. Among thousands of interested researchers only a few cognoscenti seem to be aware that there is much simpler concept at the core of SSIM; a concept that is, arguably, the real reason for its explanatory power. But as the citation juggernaut rolls onward most researchers seem destined to implementing endless variations of the unintuitively complex SSIM formula. A more hard-nosed evaluation of SSIM might conclude that it is only an indirect way of using the simplest of all perceptually-masked image quality metrics (namely normalized RMSE), and that SSIM works coincidentally since the covariance term is actually the MSE in disguise.

Perhaps the search for an ideal image quality index can now formally split into:

1. Intricate HVS informed models that closely correlate with human observers,

2. Simple, viewing condition invariant, low-computational models with tractable mathematical properties yet tolerable HVS correlation.
Symmetric Reformulation

A few simple substitutions dramatically simplify the SSIM of Wang and Bovik [1]. Because of the partial cancellation of the variance and covariance terms, the SSIM for two images $S$ can be written as the product of 2 partial indices, $S_L$ and $S_V$:

$$SSIM\{f_1, f_2\} = S = \frac{2\mu_1\mu_2}{\mu_1^2 + \mu_2^2} \times \frac{2\sigma_{12}}{\sigma_1^2 + \sigma_2^2} \times \frac{\sigma_{12}}{\sigma_1\sigma_2}$$

$$S = \frac{2\mu_1\mu_2}{\mu_1^2 + \mu_2^2} \times \frac{2\sigma_{12}}{\sigma_1^2 + \sigma_2^2} = S_L \cdot S_V$$  \hspace{1cm} (1)

The first index $S_L$ contains local luminance and the second $S_V$ local covariance/variance.

SSIM symmetry/antisymmetry

Observe that SSIM is a symmetric measure relative to image ordering $\Gamma\{f_1, f_2\} = \Gamma\{f_2, f_1\}$, so that it seems natural to formulate in terms of symmetric/antisymmetric image constructs:

$$f_+(x) = f_2(x) + f_1(x) \quad \quad 2f_1(x) = f_+(x) - f_-(x)$$
$$f_-(x) = f_2(x) - f_1(x) \quad \quad 2f_2(x) = f_+(x) + f_-(x)$$  \hspace{1cm} (2)

These can be interpreted simply as the (even) sum image $f_+$ and the (odd) difference $f_-$ image respectively. Now introduce the means $\mu$ and variances $\sigma$ for the even and odd images in exactly the same way as for original images. The local statistics are straightforward and covered in all the conventional texts, resulting in:

$$\mu_+ = \mu_2 + \mu_1, \quad \mu_- = \mu_2 - \mu_1$$
$$4\mu_1\mu_2 = \mu_2^2 - \mu_1^2 \quad \quad 4\sigma_{12} = \sigma_+^2 - \sigma_-^2$$
$$2\mu_1^2 + 2\mu_2^2 = \mu_1^2 + \mu_2^2 \quad \quad 2\sigma_1^2 + 2\sigma_2^2 = \sigma_+^2 + \sigma_-^2$$  \hspace{1cm} (3)

Which means the (anti)symmetrical image statistics are simply related to the original image statistics. Crucially the numerators can be represented as the difference of squares; numerators as the sum of squares.

$$SSIM\{f_1, f_2\} = \frac{2\mu_1\mu_2}{\mu_1^2 + \mu_2^2} \cdot \frac{2\sigma_{12}}{\sigma_1^2 + \sigma_2^2} = \left( \frac{\mu_2^2 - \mu_1^2}{\mu_1^2 + \mu_2^2} \right) \left( \frac{\sigma_+^2 - \sigma_-^2}{\sigma_1^2 + \sigma_2^2} \right)$$  \hspace{1cm} (4)
Accordingly the SSIM now has a covariance free formulation: The terms on the RHS can be interpreted directly as a (squared) luminance contrast and a variance contrast (in the sense of a Michelson contrast or visibility as described by Peli [2]).

Many important types of image distortion (jpeg compression, additive noise, blurring etc) induce negligible variation in the luminance term, and we ignore it henceforward:

$$\begin{align*}
\mu_o &= 0 \\
\mu_1 = \mu_2 = \mu_c \\
\mu_1^2 + \mu_2^2 &= 1,
\end{align*}$$

SSIM \{f_1, f_2\} = S_p = \frac{2\sigma_{12}}{\sigma_1^2 + \sigma_2^2} = \frac{\sigma_+^2 - \sigma_-^2}{\sigma_+^2 + \sigma_-^2}

(5)

Applying the linear operations of negation and offset to SSIM [3] gives the classic form of a normalised dissimilarity index (RHS):

$$1 - \text{SSIM} \{f_1, f_2\} = 1 - \frac{\sigma_+^2 - \sigma_-^2}{\sigma_+^2 + \sigma_-^2} = \frac{2\sigma_-^2}{\sigma_+^2 + \sigma_-^2}

(7)

In most published comparisons of the SSIM versus human opinion (MOS and DMOS) the highly curved scatterplots are straightened up with a heuristic logistic curve remapping. Yet it seems a trivial square root operation can achieve much the same effect:

$$\sqrt{1 - \text{SSIM} \{f_1, f_2\}} = \frac{\sqrt{2}\sigma_-}{\sqrt{\sigma_+^2 + \sigma_-^2}}

(8)

The above formulation is ubiquitous: for example the Divisive Normalisation of Laparra’s [4] image quality measure, the Noise Visibility Function (NVF) of Voloshynovskiy [5] the complex image contrast of Peli [2] and more generally the idea of relative noise or Normalized Root Mean Square Error (NRMSE) [6]. Laparra’s analysis for two image databases LIVE and TID2008 (figs 6 & 7 in [4]) shows that the Divisive Normalisation outperforms SSIM in terms of correlation with Difference Mean Opinion Scores (DMOS) of image distortion. Of the measures investigated only the VIF [7] slightly outperforms Laparra’s divisive norm, and VIF achieves this through a significantly more computational, multi-channel HVS model.

Both the concept and the computation of SSIM can be replaced by the far simpler concept and computation of a universal noise visibility function or Dissimilarity Quotient (DQ):

$$D = \sqrt{\frac{1 - S_p}{2}} = \frac{\sigma_-}{\sqrt{\sigma_+^2 + \sigma_-^2}}

(8a)

Here we have taken the liberty of naming the quantity DQ so as to distinguish the notational efficiency afforded by the symmetric formulation.
**Discussion**
Applying Occam's Razor to our current understanding of SSIM inevitably leads to a concept like Dissimilarity Quotient with equivalent explanatory power and a simple, tractable, and perceptually pleasing mathematical formulation.

**Notes**

**Avoidance of zero division**
SSIM, NRMSE and Laparra's divisive norm suffer from potential blow-up caused by a zero denominator. The problem is conventionally addressed by adding a small regularizing constant to the denominator in each case.

**Aggregation/Pooling**
The spatially localized value of DQ can be aggregated into a single, overall value representing a complete image measurement. SSIM (quadratically related to DQ) typically uses spatial averaging which corresponds to Minkowski pooling of DQ with exponent of 2. Laparra's optimization gave a spatial exponent of 2.2. However, given that a exponent of 1 corresponds to a more perceptually linear effect, it may be appropriate for DQ. Nevertheless, optimal spatial pooling remains to be determined.

**Tuning, Frequency Response, Scale invariant Measures**
Models of image quality which more closely correspond to human observer DMOS ratings must inevitably be tuned to the specific HVS responses applied under specific viewing conditions and geometry. However many researchers are really looking for a robust image quality measure that applies for a large (and perhaps unspecified) range of viewing conditions and geometry. This means that such a measure must be gain, scale, and rotation invariant at the very least. The dissimilarity quotient fits the criteria, as do suitable invariant transforms of the image (applied before the variances are computed) such a gradient magnitude, Riesz transform and Laplacian. Of these only the Riesz transform maintains a neutral frequency effect and hence does not enhance high spatial frequency features.

**Extensions**
Virtually all the proposed extensions to SSIM (e.g. gradient, Riesz transform, gradient magnitude, wavelet, multi-scale, curvelet, DCT, FFT, multi-channel HVS, etcetera) can be applied directly to the dissimilarity quotient.

**Previous work**
Several other researchers have reached similar conclusions about SSIM versus normalized Mean Square Error over the last half dozen years. They have performed detailed comparisons and analyses, whereas I have tried to be more succinct than is actually possible.

**Minimal References**
[1] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image Quality Assessment: From Error Visibility to Structural Similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, Apr. 2004.
[2] E. Peli, “Contrast in complex images,” *JOSA A*, vol. 7, no. 10, pp. 2032–2040, 1990.
[3] D. Brunet, E. R. Vrscay, and Z. Wang, “Structural similarity-based approximation of signals and images using orthogonal bases,” International Conference on Image Analysis and Recognition, Portugal, 2010.
[4] V. Laparra, J. Muñoz-Mari, and J. Malo, “Divisive normalization image quality metric revisited,” *JOSA A*, vol. 27, no. 4, pp. 852–864, 2010.
[5] S. Voloshynovskiy, A. Herrigel, N. Baumgaertner, and T. Pun, “A stochastic approach to content adaptive digital image watermarking,” in *Information Hiding*, 2000, pp. 211–236.
[6] D. Brunet, E. R. Vrscay, and Z. Wang, “On the mathematical properties of the structural similarity index,” *Image Process. IEEE Trans. On*, vol. 21, no. 4, pp. 1488–1499, 2012.
[7] H. R. Sheikh and A. C. Bovik, “Image information and visual quality,” *Image Process. IEEE Trans. On*, vol. 15, no. 2, pp. 430–444, 2006.