Max and min values of the structural similarity (SSIM) function $S(x,a)$ on the L^2 sphere $S_R(a), a \in \mathbb{R}^N$

Devin Glew and Edward R. Vrscay

Department of Applied Mathematics

University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

ervrscay@uwaterloo.ca

"Waterloo Fractal Coding and Analysis Group:"

http://links.uwaterloo.ca

Research supported by Natural Sciences and Engineering Research Council of Canada (NSERC)

Introduction

Given a reference signal $a \in \mathbb{R}^N$, we analytically solve for the critical points that maximize/minimize the Structural Similarity (SSIM) function $S(x, a)$ while restricting ourself to points x that lie on an L^2 sphere in $\mathbb{R}^N$ centered at a with fixed radius R.

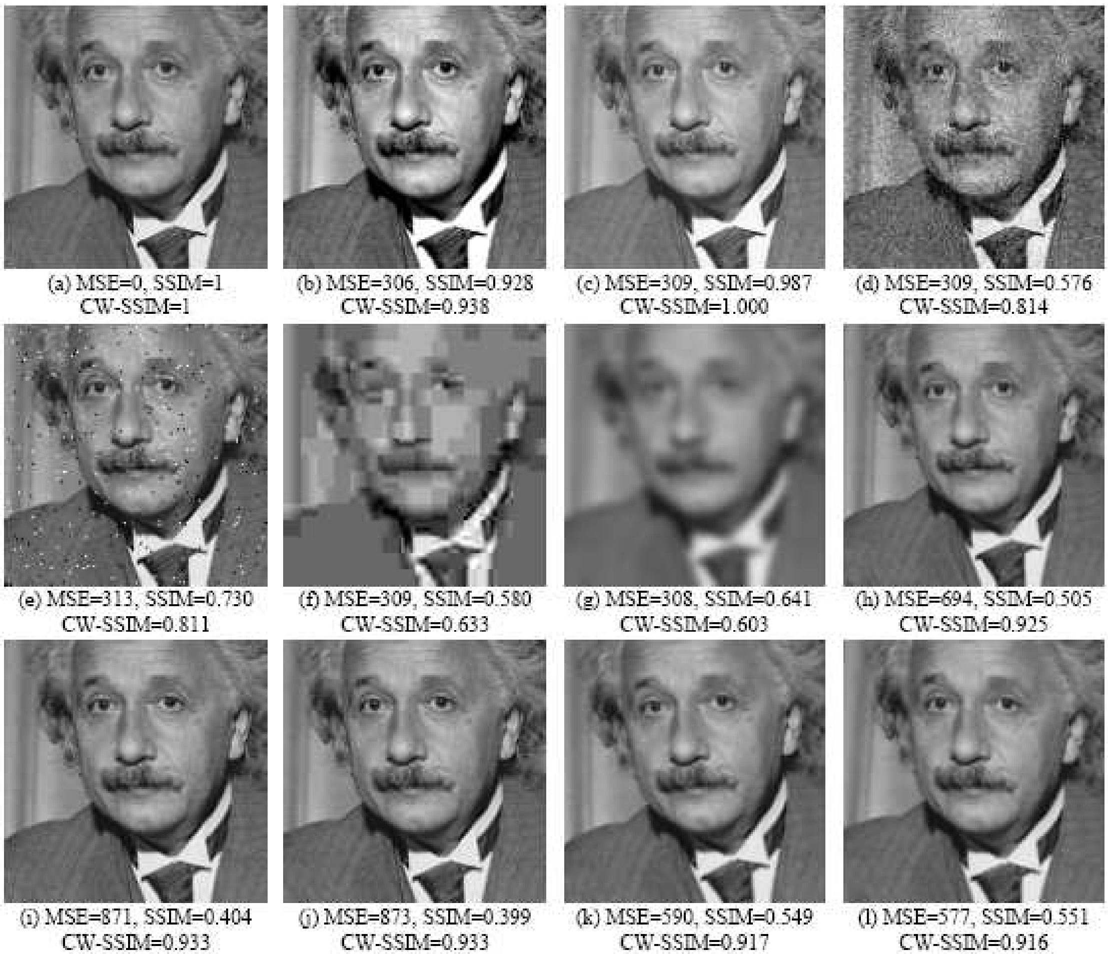
The main contribution of this paper is a "math result." It was was motivated by a very convincing example: a collection of *Einstein* images presented in a paper by by Wang and Bovik [3]. The 8 bit-per-pixel *Einstein* test image is shown along with a number of distorted versions – some obtained by adding noise, some by blurring and some by shifting.

Important points in this example:

1. The mean squared error (MSE) of these distorted images – hence there L^2 distance from the original – is roughly the same. Visually, someimage appear much closer to the undistorted image than others.
2. The Wang-Bovik structural similarity index (SSIM, and some variations) can assess the visual quality of the distorted images on the sphere: Images that are visually closer have higher SSIM values.

Primary motivation for this study

The inability of L^2 (and ability of SSIM) to measure visual closeness/distance



(a) *Einstein* test image x (upper left). All other images y_i are distorted versions. L^2 distance (MSE) between images (b)-(g) and x are roughly the same, i.e., $\|x - y_i\| \approx r$, constant. Visually, which of (b)-(g) is (i) best, (ii) worst? Images (g)-(l) illustrate effectiveness of complex-wavelet form of SSIM to detect translational distortion. From Wang *et al.* (2004).

Original formulation of SSIM index

Given two N-dimensional signals/images (or corresponding patches),

$$x = (x_1, \cdots, x_N) \quad \text{and} \quad y = (y_1, \cdots, y_N),$$

SSIM examines similarities between their **luminance**, **contrast** and **structure**.

1. For **luminance**, $l(x, y)$, use mean values, e.g.,
$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i.$$
2. For **contrast**, $c(x, y)$, use variances, e.g.,
$$s_x^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2.$$
3. For **structure**, $s(x, y)$, use normalized signals (unit standard deviation), e.g.,
$$x' = \frac{x - \bar{x}}{s_x}$$

Then combine these components (in some way!) to yield an overall similarity measure, i.e.,

$$S(x, y) = F(l(x, y), c(x, y), s(x, y)).$$

Original formulation of SSIM index (cont'd)

Given two N-dimensional signals/images (or corresponding patches),

$$x = (x_1, \cdots, x_N) \quad \text{and} \quad y = (y_1, \cdots, y_N),$$

SSIM examines similarities between their **luminance**, **contrast** and **structure**.

1. For **luminance**, $l(x,y)$, define

$$l(x,y) = \frac{2\bar{x}\bar{y} + \epsilon_1}{\bar{x}^2 + \bar{y}^2 + \epsilon_1}.$$

2. For **contrast**, $c(x,y)$, define

$$c(x,y) = \frac{2s_x s_y + \epsilon_2}{s_x^2 + s_y^2 + \epsilon_2}.$$

3. For **structure**, $s(x,y)$, compute correlation between normalized signals x' and y'. But this is equal to correlation between x and y, so define

$$s(x,y) = \frac{s_{xy} + \epsilon_3}{s_x s_y + \epsilon_3},$$

   where

$$s_{xy} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}).$$

The ϵ_i are **numerical stability coefficients** to prevent zero denominators. As well, they can model the deviation from Weber's Law as intensities approach 0.

Original formulation of SSIM index (cont'd)

Given two N-dimensional signals/images (or corresponding patches),

$$x = (x_1, \cdots, x_N) \quad \text{and} \quad y = (y_1, \cdots, y_N),$$

general form of SSIM index (Wang et al. 2004):

$$S(x, y) = [l(x, y)]^\alpha [c(x, y)]^\beta [s(x, y)]^\gamma,$$

where $\alpha, \beta, \gamma > 0$.

Simplified version used in literature as well in this study:

Set $\alpha = \beta = \gamma = 1$ and $\epsilon_3 = \epsilon_2/2$ to give

$$S(x, y) = S_1(x, y)\, S_2(x, y) = \left[\frac{2\bar{x}\bar{y} + \epsilon_1}{\bar{x}^2 + \bar{y}^2 + \epsilon_1}\right] \left[\frac{2s_{xy} + \epsilon_2}{s_x^2 + s_y^2 + \epsilon_2}\right]$$

Note:

$$-1 \leq S(x, y) \leq 1 \quad \text{and} \quad S(x, y) = 1 \;\text{ iff }\; x = y.$$

$$\text{If } \;\bar{x} = \bar{y}\; \text{ then } \;S_1(x, y) = 1$$

THE MATHEMATICAL PROBLEM:

Given a point $a = (a_1, \cdots, a_N) \in \mathbb{R}^N$, let $S_R(a)$ denote the L^2 sphere of radius R centered at a, i.e.,

$$S_R(a) = \{x \in \mathbb{R}^N \mid \|x - a\|_2 = R\}.$$

Find and classify the critical points of the SSIM function,

$$S(x, a) = \frac{4\bar{x}\bar{a}s_{xa}}{(\bar{x}^2 + \bar{a}^2)(s_x^2 + s_a^2)},$$

on $S_R(a)$. Here,

$$\bar{x} = \frac{1}{N}\sum_{k=1}^{N} x_k, \quad s_{xa} = \frac{1}{N-1}\sum_{k=1}^{N}(x_k - \bar{x})(a_k - \bar{a}),$$

and the formula for $s_x^2 = s_{xx}$ follows.

Note that, for simplicity, we have set the so-called SSIM stability constants to zero, simplifying the form of the SSIM function and making possible an analytic solution of the problem.

THE SOLUTION

Here, we simply present some of the most important points. Details are presented in Paper 206 of the Proceedings.

The Lagrangian function associated with this problem is given by

$$L(x) = S(x, a) + \lambda g(x),$$

where

$$g(x) = \sum_{k=1}^{N} (x_k - a_k)^2 - R^2$$

represents the constraint and λ denotes the Lagrange multiplier. As usual, we impose the stationary constraints,

$$\frac{\partial L}{\partial x_p} = 0, \quad 1 \le p \le N.$$

The necessary partial derivatives are

$$\begin{aligned}\frac{\partial S}{\partial x_p} &= \frac{4\bar{a}}{N(\bar{x}^2 + \bar{a}^2)^2 (s_x^2 + s_a^2)^2} \Big[s_{xa}(s_x^2 + s_a^2)(\bar{a}^2 - \bar{x}^2) \\ &+ \frac{N}{N-1} \bar{x}(\bar{x}^2 + \bar{a}^2)(s_x^2 + s_a^2)(a_p - \bar{a}) - \frac{2N}{N-1} \bar{x} s_{xa} (\bar{x}^2 + \bar{a}^2)(x_p - \bar{x}) \Big].\end{aligned}$$

The Lagrangian stationarity constraints yield the equations

$$\frac{4\bar{a}}{N(\bar{x}^2+\bar{a}^2)^2(s_x^2+s_a^2)^2}\left[s_{xa}(s_x^2+s_a^2)(\bar{a}^2-\bar{x}^2)+\right.$$
$$\frac{N}{N-1}\bar{x}(\bar{x}^2+\bar{a}^2)(s_x^2+s_a^2)(a_p-\bar{a})$$
$$\left.-\frac{2N}{N-1}\bar{x}s_{xa}(\bar{x}^2+\bar{a}^2)(x_p-\bar{x})\right]+2\lambda(x_p-a_p)=0, \quad 1\le p\le N.$$

Summing up both sides of the above equation for $1 \le p \le N$ yields the following equality,

$$\frac{4\bar{a}}{(\bar{x}^2+\bar{a}^2)^2(s_x^2+s_a^2)^2}s_{xa}(s_x^2+s_a^2)(\bar{a}^2-\bar{x}^2)+2N\lambda(\bar{x}-\bar{a})=0. \qquad (1)$$

This equation is obviously satisfied if $\bar{x}=\bar{a}$ but we must also examine the case $\bar{x}\neq\bar{a}$.

Case 1: $\bar{x} = \bar{a}$

This condition,

$$x_1 + x_2 + \cdots + x_N = a_1 + a_2 + \cdots + a_N,$$

defines a **hyperplane** in $\mathbb{R}^N$ that passes through the center of the sphere $S_R(a)$.

Critical points and associated SSIM values are

$$\begin{aligned} S_\beta^{(1)} &= \frac{s_a^2 + s_a \frac{R}{\sqrt{N-1}}}{s_a^2 + s_a \frac{R}{\sqrt{N-1}} + \frac{R^2}{2(N-1)}} \quad \text{at} \quad x = a + R\hat{a}' \\ S_\beta^{(2)} &= \frac{s_a^2 - s_a \frac{R}{\sqrt{N-1}}}{s_a^2 - s_a \frac{R}{\sqrt{N-1}} + \frac{R^2}{2(N-1)}} \quad \text{at} \quad x = a - R\hat{a}' \end{aligned} \tag{2}$$

Here,

$$a' = a - \bar{a}\underline{1}$$

denotes the zero-mean component of a and $\hat{a}'$ is the unit vector in the direction of a'.

A little algebra shows that

$$S_\beta^{(1)} > S_\beta^{(2)}.$$

Case 2: $\bar{x} \neq \bar{a}$

The following two local extrema necessarily exist:

$$S_\alpha^{(1)} = \frac{\bar{a}\left(\bar{a} + \frac{R}{\sqrt{N}}\right)}{\bar{a}\left(\bar{a} + \frac{R}{\sqrt{N}}\right) + \frac{R^2}{2N}} \quad \text{at} \quad x = a + R\underline{\hat{1}}$$

$$S_\alpha^{(2)} = \frac{\bar{a}\left(\bar{a} - \frac{R}{\sqrt{N}}\right)}{\bar{a}\left(\bar{a} - \frac{R}{\sqrt{N}}\right) + \frac{R^2}{2N}} \quad \text{at} \quad x = a - R\underline{\hat{1}}.$$

The unit vector $\underline{\hat{1}} = \frac{1}{\sqrt{N}}(1, 1, \cdots, 1)$ is normal to the hyperplane $\bar{x} = \bar{a}$.

The following local extrema,

$$S_\alpha^{(3)} = -\frac{\bar{a}\left(\bar{a} + \sqrt{\frac{\Delta}{N}}\right)}{\bar{a}\left(\bar{a} + \sqrt{\frac{\Delta}{N}}\right) + \frac{\Delta}{2N}} \quad \text{at} \quad x = 2\bar{a}\underline{1} - a + \sqrt{\Delta}\underline{\hat{1}},$$

$$S_\alpha^{(4)} = -\frac{\bar{a}\left(\bar{a} - \sqrt{\frac{\Delta}{N}}\right)}{\bar{a}\left(\bar{a} - \sqrt{\frac{\Delta}{N}}\right) + \frac{\Delta}{2N}} \quad \text{at} \quad x = 2\bar{a}\underline{1} - a - \sqrt{\Delta}\underline{\hat{1}},$$

exist provided that

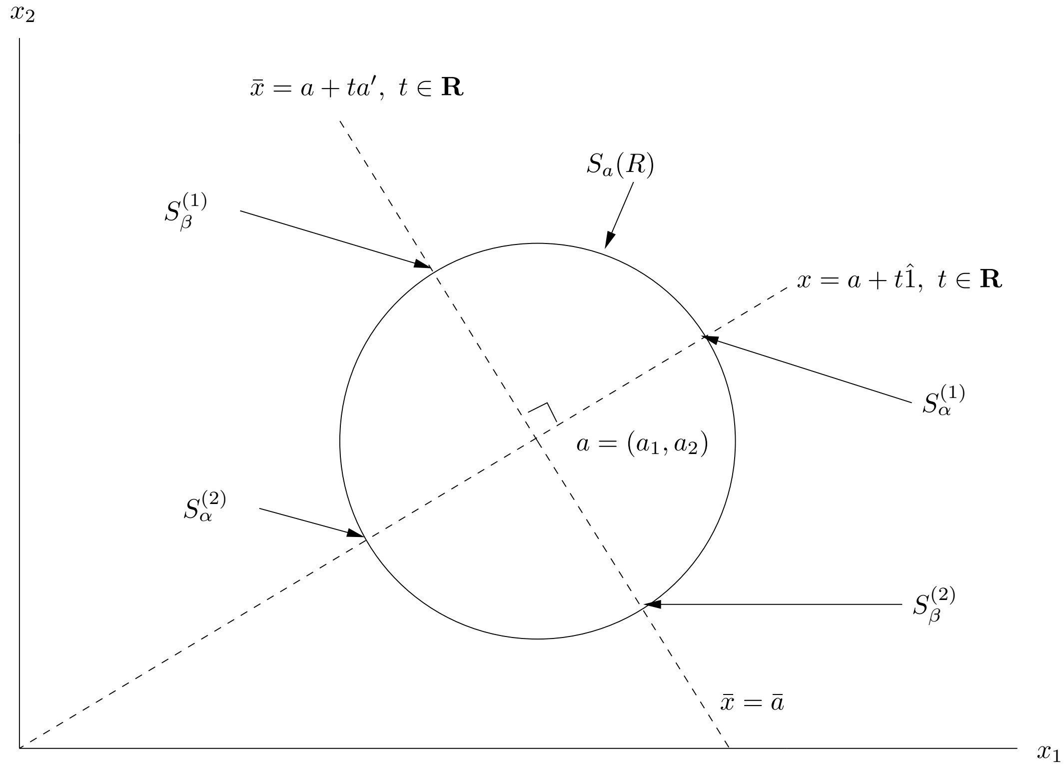
$$\Delta = R^2 - 4(N-1)s_a^2 \geq 0.$$

All extremum points for Case 2 represent constant greyscale shifts of the reference image a.

Furthermore,

$$S_\alpha^{(1)} > S_\alpha^{(2)} \text{ and } S_\alpha^{(4)} > S_\alpha^{(3)} \quad \text{if} \quad \bar{a} > 0.$$

Illustration of results (2D case)



These formulas have been verified numerically

Numerical results also show that global maxima and minima of $S(x, a)$ on the sphere $S_R(a)$ can occur both on and off the hyperplane $\bar{x} = \bar{a}$.

Unfortunately, further analysis appears to be quite complicated.

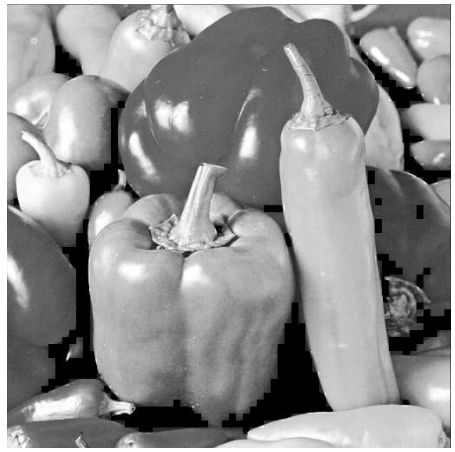
Illustration of results with some computations on images

A local block-based approach with the 512×512 pixel, 8 bit-per-pixel, *Lena* and *Peppers* test images.

Let B_i denote an image subblock, considered as an N-vector. For each B_i, the best and worst approximations to B_i, according to SSIM, while being constrained on an L^2-sphere of radius R centered at B_i were computed. (This was done by evaluating all critical points and selecting the maximum and minimum values.)

In the following calculations, 8×8 non-overlapping image subblocks were used and the radius $R = 300$. The next figure shows the results of these experiments. The SSIM values reported in the figure are the averages of the $(64 \times 64 = 4096)$ non-overlapping block SSIM values.

((a)) Best: SSIM = 0.9408



((b)) Best: SSIM = 0.9121

((c)) Original: SSIM = 1

((d)) Original: SSIM = 1

((e)) Worst: SSIM = -0.8466

((f)) Worst: SSIM = -0.7783

Best and worst 8×8-pixel block approximations to *Lena* and *Peppers* constrained on L^2-spheres of radius $R = 300$ centered at each block.

References

[1] D. Glew, *Self-similarity of images and non-local image processing*, M.Math. Thesis, Department of Applied Mathematics, University of Waterloo (2011).

[2] Z. Wang, A.C. Bovik, H.R. Sheikh and E.P. Simoncelli, Image quality assessment: From error visibility to structural similarity, IEEE Trans. Image Proc. **13**, 600-612 (2004).

[3] Z. Wang and A.C. Bovik, Mean squared error: Love it or leave it? A new look at signal fidelity measures, IEEE Sig. Proc. Mag., **26**, 98-117 (2009).

[4] Z. Wang and E.P. Simoncelli, Maximum differentiation (MAD) competition: A methodology for comparing computational models of perceptual quantities, J. Vision **8** (12) 1-13 (2008).