How to explain empirically successful structural similarity index

Gerson Cruz[1], Olga Kosheleva[2], and Vladik Kreinovich[1]

Departments of [1]Computer Science and [2]Teacher Education, University of Texas at El Paso, 500 W. University, El Paso, TX 79968, USA

gcruz10@miners.utep.edu, olgak@utep.edu, vladik@utep.edu

Formulation of the problem

- How can you gauge perception-based similarity of two images – or segments of images – x and y?
- It turns out that we can get a good description of this similarity based on the first two moments of the joint distribution:
  - the means μ_x and μ_y,
  - the standard deviations σ_x and σ_y, and the covariance σ_{xy}.
- Specifically, we need to use a combination of the following three characteristics:

$$\frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1},\ \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2},\ \text{and}\ \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3}.$$

- Why these three and not other possible characteristics?

Explanation of the third characteristic

- Let us start by explaining why, based on σ_{xy}, σ_x, and σ_y, we get the third characteristic.
- We are looking for a characteristic $F(\sigma_{xy}, \sigma_x, \sigma_y)$ that would not change if we change the lighting of the images.
- This is equivalent to multiplying each image by the corresponding constant c_x or c_y.
- Under such change, σ_x changes to $c_x \cdot \sigma_x$, σ_y to $c_y \cdot \sigma_y$, and σ_{xy} to $c_x \cdot c_y \cdot \sigma_{xy}$.
- In these terms, invariance means

$$F(\sigma_{xy}, \sigma_x, \sigma_y) = F(c_x \cdot c_y \cdot \sigma_{xy}, c_x \cdot \sigma_x, c_y \cdot \sigma_y).$$

- In particular, for $c_x = 1/\sigma_x$ and $c_y = 1/\sigma_y$, we conclude that

$$F(\sigma_{xy}, \sigma_x, \sigma_y) = f\left(\frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}\right),\ \text{where}\ f(x) \stackrel{\text{def}}{=} F(x, 1, 1).$$

- The simplest case to compute is to simply take $f(x) = x$.
- In this case, for identical images, we get 1.

Explanation of the third characteristic (cont-d)

- There is one problem.
- Two almost blank pages, for which $\sigma_x = \sigma_y \approx 0$, should be very similar, with the ratio equal to 1.
- However, the expression $\dfrac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$ is not continuous around $\sigma_x = 0$.
- So, for small σ_x and σ_y we can get many different value instead of the desired 1.
- The computationally simplest way to make it continuous is to add a constant to the denominator.
- This way we only add 1 addition – of this constant – to the algorithm.
- We also want the characteristic to be equal to 1 when the images are identical.
- This is true for the original ratio $\dfrac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$.
- However, this is not true if we add a constant to the denominator.
- To make it 1 again, the computationally simplest way is to add the same constant to the numerator.
- Thus, we get the third characteristic.

Explanation of the first characteristic

- What characteristic $G(\mu_x, \mu_y)$ can we construct based on the means?
- We cannot make it fully scale-invariant – as in the above explanation.
- Indeed, as one can show, the only scale-invariant characteristic is the identical constant.
- However, we can make it invariant with respect to similar change to both images, when we replace μ_x and μ_y with $c \cdot \mu_x$ and $c \cdot \mu_y$.
- This implies that we cannot avoid division, so we can have a characteristic $\dfrac{P(\mu_x, \mu_y)}{Q(\mu_x, \mu_y)}$.

Explanation of the first characteristic (cont-d)

- We want this characteristic to be 1 if and only if $\mu_x = \mu_y$ and smaller than 1 in all other cases.
- One can show that this excludes the case of computationally simplest case of linear P and Q.
- So P and Q should be at least quadratic – and the simplest are quadratic.
- Due to symmetry between x and y and scale-invariance:
  - we should have $P(a,b) = k_1 \cdot a \cdot b + k_2 \cdot (a^2 + b^2)$ and
  - similarly $Q(a,b) = \ell_1 \cdot a \cdot b + \ell_2 \cdot (a^2 + b^2)$.
- There should be zero similarity between empty $x = 0$ and a non-empty image $y \neq 0$, so $k_2 = 0$.
- We can now divide both numerator and denominator by $k_1/2$ and get

$$P(a,b) = 2a \cdot b.$$

- For polarized images, negative amplitude makes sense.
- So, it is reasonable to require that for x and $-x$, we should have -1 – thus $\ell_1 = 0$.
- The condition that for $x = y$ we should have 1 leads to $\ell_2 = 1$, so we get

$$\frac{2\mu_x \cdot \mu_y}{\mu_x^2 + \mu_y^2}.$$

- Similarly to the previous case, this expression is not continuous for

$$\mu_x = \mu_y = 0.$$

- To make it continuous, the computationally simplest way is:
  - to add a constant to the denominator – and then
  - to add the same constant to the numerator to make sure that the characteristic if 1 when $x = y$.

Explanation of the second characteristic

- The second characteristic can be explained similarly.
- The only difference is that we cannot have negative σ's.