

A Class of Image Metrics Based on the Structural Similarity Quality Index

Dominique Brunet¹, Edward R. Vrscay¹, and Zhou Wang²

¹ Department of Applied Mathematics, Faculty of Mathematics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

{dbrunet, ervrscay}@uwaterloo.ca

² Department of Electrical and Computer Engineering, Faculty of Engineering, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

zhouwang@ieee.org

Abstract. We derive mathematically a class of metrics for signals and images, considered as elements of $\mathbf{R}^N$, that are based upon the structural similarity (SSIM) index. The important feature of our construction is that we consider the two terms of the SSIM index, which are normally multiplied together to produce a scalar, as components of an ordered pair. Each of these terms is then used to produce a normalized metric, one of which operates on the means of the signals and the other of which operates on their zero-mean components. We then show that a suitable norm of an ordered pair of metrics defines a metric in $\mathbf{R}^N$.

Keywords: structural similarity index, normalized metrics, extended metrics, image quality assessment.

1 Introduction

Image quality assessment consists in modeling the perceptual fidelity between an original (ideal) image and a distorted version of it. The goal is not only to evaluate or compare the performance of image processing algorithms, but also to design an objective function to be optimized in order to develop better algorithms [1]. Traditionally, mean squared error (MSE) is used for this task, due to its simplicity and its many nice mathematical properties [1]. However, it is well known [1] that L^2 -based measures, e.g., mean squared error (MSE), are not necessarily good measures of visual quality.

Several image quality measures have been proposed in the literature as candidates to replace MSE [2]. While they generally outperform MSE in psycho-visual experiments, they are not known to share the mathematical properties of the MSE, making optimization very difficult to achieve. One concern is that these quality measures are not metrics in the strict mathematical sense since they do not satisfy the triangle inequality. As such, they are not amenable to standard procedures of mathematical analysis that may establish important properties, e.g., convergence, contractivity of operators.

An example of an application where these properties are important is collective sensing as described by Li in [3]. The main idea is to model an image as the

fixed point of a non-local operator, such as non-local means [4], BM3D [5] or a simplified version of non-local total variation [6]. One of the ideas of Li in [3] is to use an image representation that, contrary to cosine transforms and wavelets, is not based on a Hilbert space structure, but only on a metric space. Still, in his examples it was assumed implicitly that the metric used is the one associated with the L^2 -norm (i.e., MSE), which in fact is an inner product norm.

The structural similarity (SSIM) index [7] is an example of an image quality measure designed to provide better assessments of visual distortions between two images. The original formulation of the SSIM measure $S(\mathbf{x}, \mathbf{y})$ between two signals or images, $\mathbf{x}, \mathbf{y} \in \mathbf{R}_+^N$, involves a product of three terms, each of which measures a particular aspect of two images or image patches being compared, namely (i) the similarity of their mean values, (ii) the similarity between their contrasts and (iii) their correlation. The final two terms, however, can be collapsed into a single term. The resulting SSIM measure represents a combination of two pieces of information to produce a single number that characterizes the visual similarity of two image blocks. Such a procedure is known as *scalarization*. The question arises, however, whether it might be desirable to keep the two components, $S_1(\mathbf{x}, \mathbf{y})$ and $S_2(\mathbf{x}, \mathbf{y})$, of the SSIM separate, i.e., to treat the SSIM measure as a **vector**, an ordered pair, as opposed to a **scalar**. In this way, for example, their contributions could be weighted.

We show in this paper an example of a class of metrics for images derived from the SSIM index for which are associated neither norms nor inner products. This is done by first decomposing a signal $\mathbf{x}$ into two orthogonal components, a one-dimensional space, $\mathbf{R}_1^N$, which involves only $\bar{\mathbf{x}}$, the mean of $\mathbf{x}$, and an $(N-1)$ -dimensional space, $\mathbf{R}_2^N$, containing the zero-mean component of $\mathbf{x}$. We then show that if d_1 and d_2 are any two metrics on the spaces $\mathbf{R}_1^N$ and $\mathbf{R}_2^N$, respectively, then the L^p norm of the ordered pair $\mathbf{d} = (d_1, d_2)$ is a metric on $\mathbf{R}^N$. Finally, we employ SSIM-based metrics for d_1 and d_2 in order to obtain our desired class of image metrics.

2 The Structural Similarity (SSIM) Quality Measure

In what follows, we let $\mathbf{R}_+^N$ denote the space of non-negative N -dimensional signal/image blocks, i.e., $\mathbf{x} \in \mathbf{R}_+^N$ implies that $\mathbf{x} = (x_1, x_2, \dots, x_N)$, with $x_k \geq 0$, $1 \leq k \leq N$. We also consider the L^2 distance between two such signals $\mathbf{x}, \mathbf{y} \in \mathbf{R}_+^N$ to be the usual root mean squared error (RMSE), denoted as follows,

$$\|\mathbf{x} - \mathbf{y}\|_2 = \left[\frac{1}{N} \sum_{k=1}^N (x_k - y_k)^2 \right]^{1/2}. \quad (1)$$

The original definition of the SSIM measure between $\mathbf{x}$ and $\mathbf{y}$ is as follows,

$$S(\mathbf{x}, \mathbf{y}) = \left[\frac{2\bar{\mathbf{x}}\bar{\mathbf{y}} + \epsilon_1}{\bar{\mathbf{x}}^2 + \bar{\mathbf{y}}^2 + \epsilon_1} \right] \left[\frac{2s_{\mathbf{x}}s_{\mathbf{y}} + \epsilon_2}{s_{\mathbf{x}}^2 + s_{\mathbf{y}}^2 + \epsilon_2} \right] \left[\frac{s_{\mathbf{x}\mathbf{y}} + \epsilon_3}{s_{\mathbf{x}}s_{\mathbf{y}} + \epsilon_3} \right]. \quad (2)$$

where

$$\begin{aligned}\bar{\mathbf{x}} &= \frac{1}{N} \sum_{i=1}^N x_i, & \bar{\mathbf{y}} &= \frac{1}{N} \sum_{i=1}^N y_i, \\ s_{\mathbf{x}}^2 &= \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{\mathbf{x}})^2, & s_{\mathbf{y}}^2 &= \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{\mathbf{y}})^2, \\ s_{\mathbf{xy}} &= \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{\mathbf{x}})(y_i - \bar{\mathbf{y}}).\end{aligned}\tag{3}$$

The small positive constants $\epsilon_1, \epsilon_2 \ll 1$ are added for numerical stability along with an effort to accomodate the perception of the human visual system.

In the special case that $\epsilon_3 = \epsilon_2/2$, the above formula simplifies to the following product of two terms,

$$S(\mathbf{x}, \mathbf{y}) = S_1(\mathbf{x}, \mathbf{y}) S_2(\mathbf{x}, \mathbf{y}) = \left[\frac{2\bar{\mathbf{x}}\bar{\mathbf{y}} + \epsilon_1}{\bar{\mathbf{x}}^2 + \bar{\mathbf{y}}^2 + \epsilon_1} \right] \left[\frac{2s_{\mathbf{xy}} + \epsilon_2}{s_{\mathbf{x}}^2 + s_{\mathbf{y}}^2 + \epsilon_2} \right], \tag{4}$$

It is this form of SSIM, which is frequently used in applications, that will be examined in this paper. The extension to the three-term formulation in (2), if desired, is straightforward.

The component S_1 in (4) measures the similarity of the mean values, $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ of, respectively, $\mathbf{x}$ and $\mathbf{y}$. Its functional form was originally chosen in an effort to accomodate Weber's law of perception [7]. The component S_2 in (4) is a combination of the correlation and a measure of contrast distortion (similarity between the variances) between $\mathbf{x}$ and $\mathbf{y}$. Its functional form follows the idea of divisive normalization [8].

Since we are working with signals in $\mathbf{x}, \mathbf{y} \in \mathbf{R}_+^N$, it follows that $0 \leq S_1 \leq 1$ and $S_1(\mathbf{x}, \mathbf{y}) = 1$ if and only if $\bar{\mathbf{x}} = \bar{\mathbf{y}}$. Note also that $-1 \leq S_2(\mathbf{x}, \mathbf{y}) \leq 1$ and $S_2 = 1$ if and only if $\mathbf{x} - \bar{\mathbf{x}} = \mathbf{y} - \bar{\mathbf{y}}$. It implies that $-1 \leq S(\mathbf{x}, \mathbf{y}) \leq 1$ and that, for non-negative signals, $S(\mathbf{x}, \mathbf{y}) = 1$ if and only if $\mathbf{x} = \mathbf{y}$. (A negative value of $S(\mathbf{x}, \mathbf{y})$ implies that $\mathbf{x}$ and $\mathbf{y}$ are negatively correlated.) This suggests that the function,

$$T(\mathbf{x}, \mathbf{y}) = 1 - S(\mathbf{x}, \mathbf{y}), \tag{5}$$

could act as some kind of distance function, since $\mathbf{x} = \mathbf{y}$ implies that $T(\mathbf{x}, \mathbf{y}) = 0$. Note also that $0 \leq T(\mathbf{x}, \mathbf{y}) \leq 2$.

We now examine the components, S_1 and S_2 in (4), in this way. For S_1 ,

$$\begin{aligned}1 - S_1(\mathbf{x}, \mathbf{y}) &= 1 - \frac{2\bar{\mathbf{x}}\bar{\mathbf{y}} + \epsilon_1}{\bar{\mathbf{x}}^2 + \bar{\mathbf{y}}^2 + \epsilon_1} \\ &= \frac{|\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2}{\bar{\mathbf{x}}^2 + \bar{\mathbf{y}}^2 + \epsilon_1}.\end{aligned}\tag{6}$$

The RHS of (6) may be viewed as a *normalized* squared L^2 distance between the mean values $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$. For S_2 ,

$$1 - S_2(\mathbf{x}, \mathbf{y}) = 1 - \frac{2s_{\mathbf{xy}} + \epsilon_2}{s_{\mathbf{x}}^2 + s_{\mathbf{y}}^2 + \epsilon_2}$$

$$= \frac{s_{\mathbf{x}}^2 + s_{\mathbf{y}}^2 - 2s_{\mathbf{xy}}}{s_{\mathbf{x}}^2 + s_{\mathbf{y}}^2 + \epsilon_2} . \quad (7)$$

In the special case $\bar{\mathbf{x}} = \bar{\mathbf{y}} = 0$,

$$1 - S_2(\mathbf{x}, \mathbf{y}) = \frac{\|x - y\|^2}{\|x\|^2 + \|y\|^2 + \frac{N-1}{N}\epsilon_2} , \quad (8)$$

which is also a *normalized* squared L^2 distance between $\mathbf{x}$ and $\mathbf{y}$. Equations (6) and (8) suggest that it is natural to consider SSIM-based metrics which operate on a decomposition of signals into their means and zero-mean components. This will be done in the next section.

3 A Class of SSIM-Based Metrics

3.1 Orthogonal Decomposition of the Signal/Image Space

Here, we shall work in the space $\mathbf{R}^N$ of N -dimensional signals/image blocks. We also let $\mathbf{R}_2^N \subset \mathbf{R}^N$ denote the $(N - 1)$ -dimensional subspace (hyperplane) of zero-mean signals, i.e.,

$$\mathbf{x} = (x_1, x_2, \dots, x_N) \in \mathbf{R}_2^N \Rightarrow \bar{\mathbf{x}} = 0 \text{ or } \sum_{k=1}^N x_k = 0 . \quad (9)$$

Finally, define the one-dimensional subspace $\mathbf{R}_1^N = \text{span}\{(1, 1, \dots, 1)\}$, i.e.,

$$\mathbf{R}_1^N = \{\mathbf{y} = (y_1, y_2, \dots, y_n) \mid \mathbf{y} = c(1, 1, \dots, 1) \text{ for some } c \in \mathbf{R}\} . \quad (10)$$

$\mathbf{R}_1^N$ and $\mathbf{R}_2^N$ are orthogonal complements of each other since $\mathbf{x} \in \mathbf{R}_2^N$ and $\mathbf{y} \in \mathbf{R}_1^N$ implies that

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{k=1}^N x_k y_k = 0 . \quad (11)$$

Moreover,

$$\mathbf{R}^N = \mathbf{R}_1^N \oplus \mathbf{R}_2^N . \quad (12)$$

We shall denote the orthogonal decomposition of an element $\mathbf{x} \in \mathbf{R}^N$ in terms of these two subspaces as follows,

$$\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2, \quad \mathbf{x}_1 \in \mathbf{R}_1^N, \quad \mathbf{x}_2 \in \mathbf{R}_2^N . \quad (13)$$

The component $\mathbf{x}_1$ is the projection of $\mathbf{x}$ onto the subspace $\mathbf{R}_1^N$, i.e.,

$$\mathbf{x}_1 = \langle \mathbf{x}, \hat{\mathbf{e}}_1 \rangle \hat{\mathbf{e}}_1, \quad \text{where } \hat{\mathbf{e}}_1 = \frac{1}{\sqrt{N}}(1, 1, \dots, 1) . \quad (14)$$

Therefore,

$$\mathbf{x}_1 = (\bar{\mathbf{x}}, \bar{\mathbf{x}}, \dots, \bar{\mathbf{x}}) = \bar{\mathbf{x}}(1, 1, \dots, 1) . \quad (15)$$

where $\bar{\mathbf{x}}$ is the mean of $\mathbf{x}$ defined in (3). It follows that the zero-mean component, $\mathbf{x}_2$, of $\mathbf{x}$ in $\mathbf{R}_2^N$ is given by

$$\mathbf{x}_2 = \mathbf{x} - \mathbf{x}_1 . \quad (16)$$

3.2 A Class of Two-Dimensional Metrics

The next step is to consider metrics on these orthogonal spaces. Let d_1 be a metric on $\mathbf{R}$ and d_2 a metric on $\mathbf{R}^{N-1}$. Then for any two elements $\mathbf{x}, \mathbf{y} \in \mathbf{R}^N$, define the corresponding ordered pair,

$$\mathbf{d} = (d_1(\bar{\mathbf{x}}, \bar{\mathbf{y}}), d_2(\mathbf{x}_2, \mathbf{y}_2)) \in \mathbf{R}^2. \quad (17)$$

It is clear that $\mathbf{x} = \mathbf{y}$ implies that $\mathbf{d} = \mathbf{0}$. The following result shows that $\mathbf{d}$ can be used to define a metric on $\mathbf{R}^N$.

Theorem 1. *Let $\|\cdot\|$ be a norm in $\mathbf{R}^2$ that satisfies the following increasing property in $\mathbf{R}_+^2$: For any $\mathbf{a} \in \mathbf{R}_+^2$ and any positive ordered pair $\mathbf{b} = (b_1, b_2)$, with $b_1, b_2 > 0$,*

$$\|\mathbf{a} + \mathbf{b}\| \geq \|\mathbf{a}\|. \quad (18)$$

Then for $\mathbf{d}$ defined in (17),

$$d(\mathbf{x}, \mathbf{y}) := \|\mathbf{d}\| \quad (19)$$

is a metric in $\mathbf{R}^N$.

Note 1. This theorem can be generalized for a combination of M metrics on $\mathbf{R}^N$.

Before proving this theorem we state that for any $p \geq 1$, the L^p norm in $\mathbf{R}^2$ satisfies the above increasing property. It also applies to the case $p = \infty$, i.e., the L_∞ norm. This can be checked by using Taylor's Theorem for multivariable functions.

Proof. It is quite straightforward to show that $d(\mathbf{x}, \mathbf{y})$ in (19) satisfies following necessary properties of a metric:

1. $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$ (symmetry),
2. $d(\mathbf{x}, \mathbf{y}) \geq 0$ (positivity),
3. $d(\mathbf{x}, \mathbf{y}) = 0$ if and only if $\mathbf{x} = \mathbf{y}$ (strict positivity).

It remains to prove that $d(\mathbf{x}, \mathbf{y})$ satisfies the triangle inequality, i.e., for any $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbf{R}^N$,

$$d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y}). \quad (20)$$

This result follows from the assumptions that d_1 and d_2 are metrics and that the $\|\cdot\|$ norm satisfies the increasing property:

$$\begin{aligned} d(\mathbf{x}, \mathbf{y}) &= \|(d_1(\bar{\mathbf{x}}, \bar{\mathbf{y}}), d_2(\mathbf{x}_2, \mathbf{y}_2))\| \\ &\leq \|(d_1(\bar{\mathbf{x}}, \bar{\mathbf{z}}) + d_1(\bar{\mathbf{z}}, \bar{\mathbf{y}}), d_2(\mathbf{x}_2, \mathbf{z}_2) + d_2(\mathbf{z}_2, \mathbf{y}_2))\| \\ &= \|(d_1(\bar{\mathbf{x}}, \bar{\mathbf{z}}), d_2(\mathbf{x}_2, \mathbf{z}_2)) + (d_1(\bar{\mathbf{z}}, \bar{\mathbf{y}}), d_2(\mathbf{z}_2, \mathbf{y}_2))\| \\ &\leq \|(d_1(\bar{\mathbf{x}}, \bar{\mathbf{z}}), d_2(\mathbf{x}_2, \mathbf{z}_2))\| + \|(d_1(\bar{\mathbf{z}}, \bar{\mathbf{y}}), d_2(\mathbf{z}_2, \mathbf{y}_2))\| \\ &= d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y}). \end{aligned} \quad (21)$$

□

Note 2. The increasing property in (18) also holds for suitably weighted L^p norms, e.g.,

$$\|\mathbf{x}\| = \left[\sum_{k=1}^N w_{k,p} |x_k|^p \right]^{1/p}, \quad (22)$$

where $w_{k,p} > 0$ for $1 \leq k \leq N$. But (18) does not hold for **all** norms. That being said, the validity for L^p and weighted norms is sufficient for most, if not all, practical purposes.

3.3 The Normalized Metric Relevant to SSIM

We now return to the results of (6) and (8) in order to construct a SSIM-based metric. The following result will be necessary.

Theorem 2. For $M \geq 1$, let $\|\cdot\|_2$ the L^2 norm in $\mathbf{R}^M$. Then for $\epsilon \geq 0$, $\bar{d} : \mathbf{R}^M \times \mathbf{R}^M \rightarrow \mathbf{R}$, given by

$$\bar{d}(\mathbf{x}, \mathbf{y}) = \begin{cases} \frac{\|\mathbf{x} - \mathbf{y}\|_2}{\sqrt{\|\mathbf{x}\|_2^2 + \|\mathbf{y}\|_2^2 + \epsilon}}, & (\mathbf{x}, \mathbf{y}) \neq (\mathbf{0}, \mathbf{0}), \\ 0, & \mathbf{x} = \mathbf{y} = \mathbf{0}, \end{cases} \quad (23)$$

is a metric.

This theorem was proved for the case $\epsilon = 0$ in [9]. The proof for the case $\epsilon > 0$ will appear elsewhere [10].

Note 3. The metric $\bar{d}$ is an example of a *normalized metric*. The range of values assumed by $\bar{d}$ is the bounded interval $[0, \sqrt{2}]$: $\bar{d}(\mathbf{x}, \mathbf{y}) = 0$ when $\mathbf{x} = \mathbf{y}$ and, for $\epsilon = 0$, $\bar{d}(\mathbf{x}, \mathbf{y}) = \sqrt{2}$ when $\mathbf{y} = -\mathbf{x}$.

Note 4. For every $\epsilon \geq 0$, $\bar{d}(\mathbf{x}, \mathbf{0})$ is not a norm, since $\bar{d}(\alpha\mathbf{x}, \mathbf{0}) \neq \alpha\bar{d}(\mathbf{x}, \mathbf{0})$ for any $\alpha > 0$.

Note 5. The following is an interesting property of this metric: In the case $\epsilon = 0$,

$$\bar{d}(\mathbf{x}, \mathbf{0}) = 1 \quad \text{for all } \mathbf{x} \in \mathbf{R}^M. \quad (24)$$

This implies that no sequences $\{\mathbf{x}_n\}$ can converge to $\mathbf{0}$ in this metric: Even if $\mathbf{x}_n \rightarrow \mathbf{0}$ in the metric defined by the $\mathbf{R}^M$ norm $\|\cdot\|$, i.e. $\lim_{n \rightarrow \infty} \|\mathbf{x}_n - \mathbf{0}\| = 0$, it cannot converge to $\mathbf{0}$ in $\bar{d}$ metric since $\lim_{n \rightarrow \infty} \bar{d}(\mathbf{x}_n, \mathbf{0}) = 1$. This is not a major problem since, in general, we are concerned only with non-zero signals.

Nevertheless, this nonconvergence of sequences to $\mathbf{0}$ in the $\bar{d}$ metric disappears when $\epsilon > 0$. This parameter will, in fact, appear if we consider nonzero stability constants in the SSIM function of (4).

Note 6. Once again in the case $\epsilon = 0$, we have a scale invariance property: For any $\alpha \in \mathbf{R}$, $\bar{d}(\alpha\mathbf{x}, \alpha\mathbf{y}) = \bar{d}(\mathbf{x}, \mathbf{y})$, which is consistent with (24).

Unlike the L^2 case (Euclidean metric), the level sets associated with this metric are nonconcentric (hyper)spheres. To illustrate, we consider the simple $\mathbf{R}^2$ case with $\epsilon = 0$. Let $\mathbf{a} = (a_1, a_2)$ denote a reference point in $\mathbf{R}^2$. The C -level set of the metric $\bar{d}$ is the set of $\mathbf{x} = (x_1, x_2)$ for which $\bar{d}(\mathbf{x}, \mathbf{a}) = C$, where $C \in [0, \sqrt{2}]$, i.e.,

$$\frac{\|\mathbf{x} - \mathbf{a}\|_2}{\sqrt{\|\mathbf{x}\|_2^2 + \|\mathbf{a}\|_2^2}} = C \Rightarrow \|\mathbf{x} - \mathbf{a}\|_2^2 = C^2\|\mathbf{x}\|_2^2 + C^2\|\mathbf{a}\|_2^2. \quad (25)$$

After a little algebra, we found that the level sets may be classified into the following cases:

Case 1: $0 \leq C < 1$. For each C -value, the corresponding C -level set is composed of the points $\mathbf{x} = (x_1, x_2)$ that satisfy the equation,

$$\left[x_1 - \frac{a_1}{1 - C^2}\right]^2 + \left[x_2 - \frac{a_2}{1 - C^2}\right]^2 = \frac{C^2(2 - C^2)}{(1 - C^2)^2}(a_1^2 + a_2^2). \quad (26)$$

This is a circle centered at $\frac{1}{1 - C^2}(a_1, a_2)$ with radius $r = \frac{C\|\mathbf{a}\|_2}{1 - C^2}\sqrt{2 - C^2}$.

The centers of these circles lie on the line that extends from the origin $\mathbf{0}$ to the point $\mathbf{a}$. They start at $\mathbf{a}$ ($C=0$) and travel outward to infinity as $C \rightarrow 1^-$.

Case 2: $C = 1$. The level set is the line $a_1x_1 + a_2x_2$ which contains the origin $(0, 0)$. This line is perpendicular to the line that supports the centers of the level sets in Case 1.

Case 3: $1 < C \leq \sqrt{2}$. For each C -value the corresponding C -level set is composed of the points $\mathbf{x} = (x_1, x_2)$ that satisfy the equation,

$$\left[x_1 + \frac{a_1}{C^2 - 1}\right]^2 + \left[x_2 + \frac{a_2}{C^2 - 1}\right]^2 = \frac{C^2(2 - C^2)}{(C^2 - 1)^2}(a_1^2 + a_2^2). \quad (27)$$

This is a circle centered at $\frac{1}{C^2 - 1}(-a_1, -a_2)$ with radius $r = \frac{C\|\mathbf{a}\|_2}{C^2 - 1}\sqrt{2 - C^2}$.

Their centers of these circles lie on the line that extends from the origin $\mathbf{0}$ to the point $-\mathbf{a}$. They are coming in from infinity ($C = \sqrt{2}$) and travel toward $-\mathbf{a}$ as $C \rightarrow \sqrt{2}$. At $C = \sqrt{2}$, the level set is the single point $-\mathbf{a}$.

In Fig. 1 are plotted some level sets associated with the point $\mathbf{a} = (1, 1)$.

3.4 Construction of the SSIM-Based Metric

We may now define the SSIM-based metric that results from the above constructions. The normalized metric $\bar{d}$ will be used in each of the subspaces $\mathbf{R}_1^N$ and $\mathbf{R}_2^N$ defined in Sect. 3.1.

Given $\mathbf{x}, \mathbf{y} \in \mathbf{R}^N$, we now define the following vector of metrics,

$$\mathbf{d}(\mathbf{x}, \mathbf{y}) = (d_1(\bar{\mathbf{x}}, \bar{\mathbf{y}}), d_2(\mathbf{x}_2, \mathbf{y}_2)) \in \mathbf{R}^2, \quad (28)$$

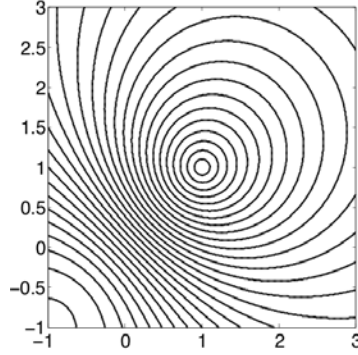


Fig. 1. Level sets $\bar{d}(\mathbf{x}, \mathbf{a}) = C$ about the reference point $\mathbf{a} = (1, 1)$ for $C = \frac{1}{20}k$, $1 \leq k \leq 28$, over the region $(x_1, x_2) \in [-1, 3] \times [-1, 3]$

where

$$\begin{aligned} d_1(\bar{\mathbf{x}}, \bar{\mathbf{y}}) &= \bar{d}(\bar{\mathbf{x}}, \bar{\mathbf{y}}) = \frac{|\bar{\mathbf{x}} - \bar{\mathbf{y}}|}{\sqrt{\bar{\mathbf{x}}^2 + \bar{\mathbf{y}}^2 + \epsilon_1}} \\ d_2(\mathbf{x}_2, \mathbf{y}_2) &= \bar{d}(\mathbf{x}_2, \mathbf{y}_2) = \frac{\|\mathbf{x}_2 - \mathbf{y}_2\|_2}{\sqrt{\|\mathbf{x}_2\|_2^2 + \|\mathbf{y}_2\|_2^2 + \frac{N-1}{N}\epsilon_2}} \quad , \end{aligned} \quad (29)$$

The components, $\mathbf{x}_2$ and $\mathbf{y}_2$ of, respectively, $\mathbf{x}$ and $\mathbf{y}$ were defined in Sect. 3.1.

In the particular case of two-dimensional signals, i.e., $N = 2$, which was illustrated in Fig. 1, we may view the d_1 metric as operating on the line $x_1 - x_2 = 0$ and the d_2 metric operator as operating on the orthogonal space $x_1 + x_2 = 0$ (zero-mean signals).

Now let $\|\cdot\|$ denote any norm in $\mathbf{R}^2$ satisfying the increasing property defined in Theorem 1. From that theorem, we have the resulting metric on $\mathbf{R}^N$:

$$D(\mathbf{x}, \mathbf{y}) = \|(\bar{d}(\bar{\mathbf{x}}, \bar{\mathbf{y}}), \bar{d}(\mathbf{x}_2, \mathbf{y}_2))\| \quad . \quad (30)$$

In the case that $\|\cdot\| = \|\cdot\|_p$, the weighted L^p norm on $\mathbf{R}^2$, with $p \geq 1$, the metric is given explicitly as

$$\begin{aligned} D_p(\mathbf{x}, \mathbf{y}) &= \|(\bar{d}(\bar{\mathbf{x}}, \bar{\mathbf{y}}), \bar{d}(\mathbf{x}_2, \mathbf{y}_2))\|_p \\ &= (w_{1,p} [\bar{d}(\bar{\mathbf{x}}, \bar{\mathbf{y}})]^p + w_{2,p} [\bar{d}(\mathbf{x}_2, \mathbf{y}_2)]^p)^{1/p} \quad . \end{aligned} \quad (31)$$

The cases $p = 1$ and $p = 2$ will probably be most relevant to standard image processing procedures:

$$D_1 = \bar{d}(\bar{\mathbf{x}}, \bar{\mathbf{y}}) + \bar{d}(\mathbf{x}_2, \mathbf{y}_2) \quad , \quad (32)$$

$$D_2 = \sqrt{\bar{d}^2(\bar{\mathbf{x}}, \bar{\mathbf{y}}) + \bar{d}^2(\mathbf{x}_2, \mathbf{y}_2)} \quad . \quad (33)$$

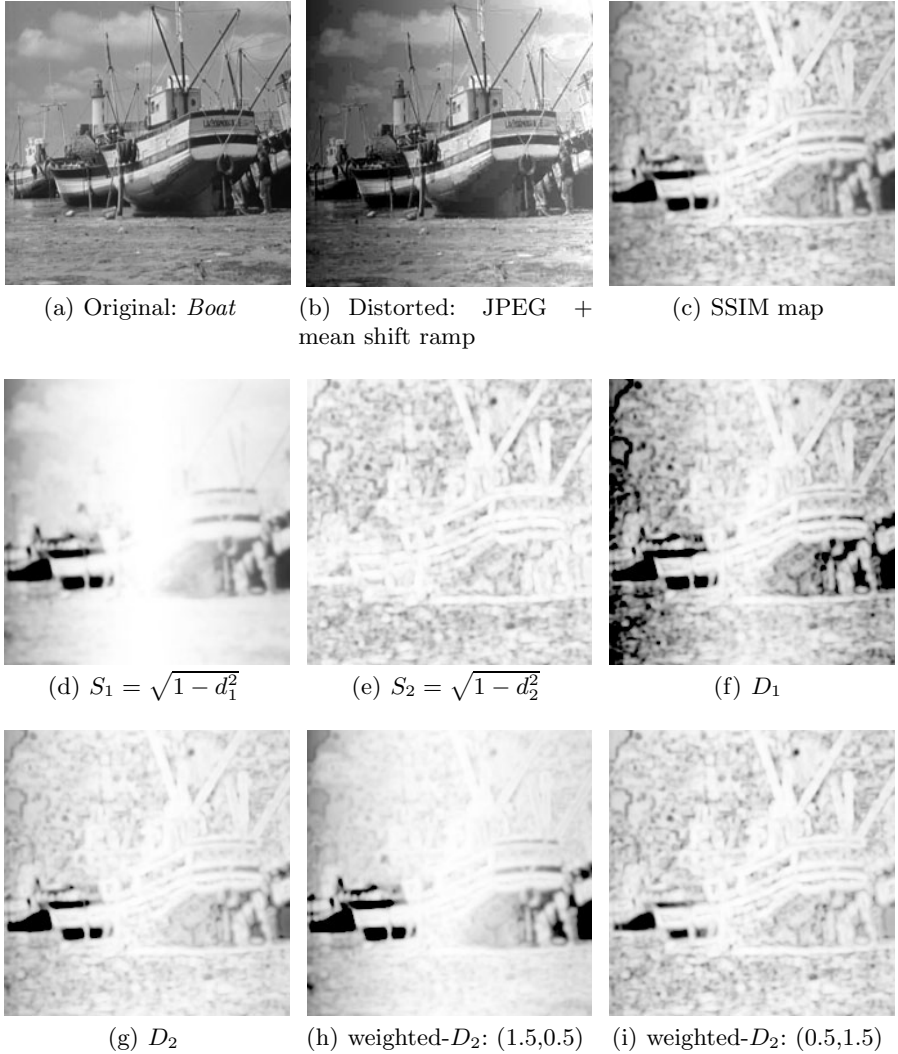


Fig. 2. (a) Original *Boat* image. (b) JPEG-compressed *Boat* image (quality factor 10/100) + horizontal mean shift ramp (from $-100/255$ to $+100/255$) (c) SSIM quality index map: Black=0, White=1. (d)-(e) S_1 and S_2 , the two components of the SSIM quality map (mean distortion and structural distortion) (f)-(i) SSIM-based metrics computed from different norms of $(d_1, d_2) = (\sqrt{1 - S_1}, \sqrt{1 - S_2})$: (f) D_1 , the L^1 -norm. (g) D_2 , the L^2 -norm. (h) Weighted L^2 -norm with $w_1 = 1.5$ and $w_2 = 0.5$. (i) Weighted L^2 -norm with $w_1 = 0.5$ and $w_2 = 1.5$. For all maps, a down-sampling was first performed and a sliding Gaussian window of STD = 1.5 pixels was used. The images (f)-(i) were rescaled with the formula $\sqrt{\max(0, 1 - D^2)}$ to look comparable to the SSIM map.

Note that in the special case $p = 2$ with unit weights, the above metric becomes the square root of the sum of the expressions in (6) and (7). Finally, the case $p = \infty$ may also be useful in some applications,

$$D_\infty(\mathbf{x}, \mathbf{y}) = \max\{\bar{d}(\bar{\mathbf{x}}, \bar{\mathbf{y}}), \bar{d}(\mathbf{x}_2, \mathbf{y}_2)\} . \quad (34)$$

By comparing the equation for D_2 with $\sqrt{1 - SSIM}$ we can understand their relationship:

$$\sqrt{1 - SSIM} = \sqrt{1 - (1 - d_1^2)(1 - d_2^2)} = \sqrt{d_1^2 + d_2^2 - d_1^2 d_2^2}. \quad (35)$$

D_2 may be viewed as a low order approximation of the SSIM index. In fact, most image distortions, e.g. JPEG and JPEG2000 compression, blur and zero-mean noise, preserve the mean. It implies that $d_1 = \bar{d}(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ will be close to zero. Thus, D_2 is a very good approximation of SSIM for most of the distortions encountered in image processing. In fact, when the means are exactly matched, $D_2(\mathbf{x}, \mathbf{y}) = \sqrt{1 - SSIM}(\mathbf{x}, \mathbf{y})$.

Example 1. To offer some comparison between the new class of metrics and SSIM – and to show some of their limitations – we present an example involving a distortion of both the local structure and the local mean value. In Fig. 2 are shown several quality maps which compare the test image *Boat* (top left) with a JPEG compressed version (quality factor 10/100) to which was added a horizontal mean shift ramp from $-100/255$ to $+100/255$ (top middle). We see that all the different metrics detect the same error than the SSIM map, but none of them give exactly the same weight than SSIM for luminance distortion and structural distortion.

Psychovisual experiments will need to be performed to find the best parameters p and $w_{k,p}$ associated with these metrics. One of these metrics could be then used in image processing applications as optimization objective.

Acknowledgements. We gratefully acknowledge the generous support of this research by the Natural Sciences and Engineering Research Council of Canada (NSERC) in the forms of Discovery Grants (ERV, ZW), a Strategic Grant (ZW), a collaborative research and development (CRD) grant (ERV, ZW) and a Post-graduate Scholarship (DB). ZW would also like to acknowledge partial support by the Province of Ontario Ministry of Research and Innovation in the form of an Early Researcher Award.

References

1. Wang, Z., Bovik, A.C.: Mean squared error: Love it or leave it? A new look at signal fidelity measures. *IEEE Signal Processing Magazine* 26(1), 98–117 (2009)
2. Wang, Z., Bovik, A.C.: *Modern Image Quality Assessment*. Morgan & Claypool Publishers (2006)
3. Li, X.: Collective sensing: a fixed-point approach in the metric space. *SPIE Conf. on VCIP* (July 2010)

4. Buades, A., Coll, B., Morel, J.M.: A review of image denoising algorithms, with a new one. *Multiscale Modelling and Simulation* 4, 490–530 (2005)
5. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. Image Processing* 16, 2080–2095 (2007)
6. Gilboa, G., Osher, S.: Nonlocal operators with applications to image processing. *Multiscale Modeling and Simulation* 7(3), 1005–1028 (2008)
7. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Processing* 13(4), 600–612 (2004)
8. Wainwright, M.J., Schwartz, O., Simoncelli, E.P.: Natural image statistics and divisive normalization: Modeling nonlinearity and adaptation in cortical neurons. In: Rao, R., et al. (eds.) *Probabilistic Models of the Brain: Perception and Neural Function*, pp. 203–222. MIT Press, Cambridge (2002)
9. Klamkin, M.S., Meir, A.: Ptolemy’s inequality, chordal metric, multiplicative metric. *Pacific J. Math.* 101(2), 389–392 (1982)
10. Brunet, D., Vrscay, E.R., Wang, Z.: On the mathematical properties of the structural similarity index (preprint). (March 2011), <http://www.math.uwaterloo.ca/~dbrunet/>