CrossMark

A New Class of Wavelet-Based Metrics for Image Similarity Assessment

Maria Grazia Albanesi[1] · Riccardo Amadeo[1] · Silvia Bertoluzza[2] ·
Giulia Maggi[3]

Abstract In this paper, we propose a new class of image similarity metrics based on a wavelet decomposition. By suitably combining weighted contributions of the different dyadic frequency bands, we define a class of similarity measures and we prove it is a metric. Moreover, we discuss the theoretical relationship between the novel class of metrics and the well-known structural similarity index (SSIM) and its multiscale versions (MSSSIM and CWSSIM). By using standard benchmark indexes over a reference database in the literature (the TID2013 database), we test the efficiency of the newly defined metrics in performing similarity assessment. We compare the performance of our metric with other well-known indexes in the literature, such as SSIM, FPH, MSSSIM, CWSSIM and PSNR, to demonstrate its improvement over the current state of the art, which becomes more evident when the query image is the one identified by the worst level of degradation which is perceived by the human visual system, as coded by the standard mean opinion score stored in the database.

The original version of this article was revised: The double vertical bars are inserted instead of single vertical bars in Equation 9.

✉ Silvia Bertoluzza
silvia.bertoluzza@imati.cnr.it

Maria Grazia Albanesi
mariagrazia.albanesi@unipv.it

1 Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Via Ferrata 1, 27100 Pavia, Italy

2 CNR IMATI Enrico Magenes, via Ferrata 1, 27100 Pavia, Italy

3 ICH Humanitas, Milan, Italy

1 Introduction

Numerous applications require tools allowing to compare two images and to decide whether they are "close" in some specific sense. To respond to this need, many different functionals (which are called, sometimes improperly, "metrics") have been introduced, which return a number measuring, in a way or another, either the "closeness" or the "diverseness" of the two images. Tools ranging from the simple *root-mean-square error* (RMSE) [29], to the more sophisticated *structural similarity index* (SSIM) [32] up to the extremely powerful but also quite expensive *mutual information* have been used in a wide variety of applications such as image quality assessment, image registration and image classification.

Obviously, the concept of "closeness" of two images is quite complex, as it depends on the one hand on different levels of description and analysis of the image content [from the simplest pixel-to-pixel comparison, to more sophisticated measures that can take into account the perceptual processing of the human visual system (HVS)], and on the other hand on the particular application. While for image quality assessment, it is, for instance, important to measure as precisely as possible the amount of noise that is contributing to the degradation of a given image, in applications such as image registration it is more important to measure the deformation between two images while filtering out the noise.

Depending on the application considered, it is then desirable that the tools to be used satisfy different requirements,

like, for instance, sensitiveness to noise, translation and rotation invariance, convexity, differentiability. One of the properties that is sometimes desirable is that the functional is (or that it is related to) a metric. This happens for SSIM, which is not a metric in itself but which generates a metric through some simple manipulations (see [4]).

In this paper, we introduce a class of multiscale versions of such a metric, of which we study the mathematical properties (local convexity, differentiability), and we point out the strong relation and the differences that these new metrics have with SSIM and its multiscale versions MSSSIM and CWSSIM [26,35], as well as with some classical functional analysis norm, namely the Besov norm.

In order to evaluate the efficacy of the new class of metrics proposed here, we have chosen a very challenging application field in computer vision: digital image similarity assessment. The first attempt to evaluate image similarity is the full reference image quality assessment, which can be considered a special case of image similarity evaluation. In fact, the two images to be compared are strictly related: the first image is the high (best) quality, *original* image, and the second image is an impaired version of the first one, i.e., the output of a process of quality degradation (noise addition, lossy compression, and so on). In this case, similarity is synonymous of quality, as the more the impaired image is similar to the original one, the higher is its quality.

It is not surprising that most of the measures of image similarity proposed in the literature are the same used in typical full reference image quality assessment. For example, the long-standing *mean squared error* (MSE), and its variant, the *peak signal-to-noise ratio* (PSNR) [10], are intensity-based measures that rely on a pixel-by-pixel comparison between the two images, evaluating each couple of pixels independently of their neighborhood. Obviously, PSNR is not a perceptual HVS-based index, although it has been used in many signal processing applications since the early beginning of computer vision, for its low computational complexity, and it is now used essentially for reference. Therefore, other measures have been proposed, based on a perceptual approach to image content analysis, such as SSIM and its multiscale evolution MSSSIM [35]. Nowadays, they are considered the best perceptual approaches for many different applications, not only for image quality assessment [31] but also for human perceived image similarity [36], image lossy compression [34], image fusion [21], video hashing [11], visual surveillance [28], infrared imaging [20], medical imaging [13] and target recognition [5].

The main advantage of SSIM is its generic definition, suitable to any kind of digital data, that guarantees its extendibility to applications beyond image processing, because of its independence from a specific image or visual model. On the other hand, SSIM turns out to be sensitive to all those geometric transformations that are non-structural [33], such as relative translations, scalings and rotations. Therefore, when referring to image similarity assessment tasks, SSIM is better suited to identify similarity (or dissimilarity) due to structural modifications on the original image, such as filtering or noise additions, rather than geometrical modifications, where rotations, translations, copy–move schemes are adopted.

The *Complex Wavelet SSIM* (CWSSIM, [26]) is an evolution of SSIM that tries to overcome these limitations. It exploits the relative phase patterns of local image frequencies (as sampled by complex wavelet coefficients) to effectively understand the structural similarity between two images, in the presence of geometrical transformations. CWSSIM, unlike the original SSIM, is an index designed precisely for image similarity assessment. Finally, while the original SSIM definition does not use color information, several attempts have been proposed to adapt the original SSIM computation algorithm to color management [14,18].

Another index that adopts the perceptual HVS point of view to discriminate similar images is the one proposed by the *Feature Point Hash* (FPH) methodology [16]. It defines perceptual image hash functions, which are able to map two images that are identical (or very similar), according the HVS, to the same hash values [16]. The study [16] presents a framework for perceptual image hashing using feature points, choosing and extracting them with a wavelet-based feature detection algorithm. This algorithm exploits known characteristics of the HVS [9,12].

In the experimental validation, we have chosen to compare our class of metrics to the SSIM, CWSSIM and FPH measures, as they are the state of the art for digital image similarity assessment, and the reader can find in the literature plenty of benchmarks as reference.

This paper is structured as follows: In Sect. 2, we define our class of metrics and discuss their mathematical properties. The results presented in this Section rely on the tools introduced in [4], which we recall in Sect. 2.2. In Sect. 3, we present the experimental setup we are using to validate our proposal. Finally, Sects. 4 and 5 report the commented results of the experiments and the conclusions, respectively.

2 The New Class of Metrics

Before going into the details of the definition of the new class of metrics, we would like to point out once again that most of the "metrics" that we mentioned in Sect. 1 are not actually metrics in the strict mathematical sense of the term. We recall that a metric on a set X is a function $d : X \times X \to \mathbb{R}^+$ satisfying the following three properties for any $x, y, z \in X$:

1. $d(x, y) = d(y, x)$ (symmetry);
2. $d(x, y) = 0$ if and only if $x = y$;
3. $d(x, y) \leq d(x, z) + d(z, y)$ (triangular inequality).

2.1 The Wavelet Transform and Some of its Properties

The class of metrics that we are going to introduce is based on the discrete wavelet transform (DWT). It is beyond the scope of this paper to give a full review of this tool (for which we refer to [15]), and we will limit ourselves to recall its definition and some related notations and properties that we will need in the following. We let $L^2(\mathbb{R})$ denote the space of finite-energy real-valued functions endowed with the norm and scalar product

$$\|f\|^2_{L^2(\mathbb{R})} = \int_{\mathbb{R}} |f(x)|^2 \,\mathrm{d}x, \quad \langle f, g \rangle = \int_{\mathbb{R}} f(x)g(x)\,\mathrm{d}x.$$

For a signal $\alpha = \{\alpha_n, \ n \in \mathbb{Z}\}$, we set

$$\|\alpha\|^2_2 = \sum_n |\alpha_n|^2.$$

We recall that $\ell^2(\mathbb{Z})$ denotes the space of infinite length signals with finite energy ($\|\cdot\|_2 < +\infty$).

We let ϕ and ψ denote a couple of compactly supported *scaling* and *wavelet* functions, and we let $h = \{h_0, \dots, h_L\}$ and $g = \{g_{-L+1}, \dots, g_1\}$ denote the related low-pass and band-pass filters, *i.e.,* the following identities hold:

$$\phi(x) = \sum_{n=0}^{L} h_n \sqrt{2}\phi(2x - n), \tag{1}$$

$$\psi(x) = \sum_{n=1-L}^{1} g_n \sqrt{2}\phi(2x - n), \quad g_n = (-1)^n h_{1-n}. \tag{2}$$

The functions ϕ and ψ are sometimes called, respectively, *father* and *mother* wavelets. We put ourselves in the orthonormal setting, that is we assume that

$$\int_{\mathbb{R}} \phi(x)\phi(x-n)\,\mathrm{d}x = 0, \ \forall n \neq 0, \quad \int_{\mathbb{R}} |\phi(x)|^2 \,\mathrm{d}x = 1, \tag{3}$$

$$\int_{\mathbb{R}} \psi(x)\psi(x-n)\,\mathrm{d}x = 0 \ \forall n \neq 0, \quad \int_{\mathbb{R}} |\psi(x)|^2 \,\mathrm{d}x = 1, \tag{4}$$

$$\int_{\mathbb{R}} \phi(x)\psi(x-n)\,\mathrm{d}x = 0, \ \forall n. \tag{5}$$

It is straightforward to check (and it is at the basis of the definition of the wavelet transform) that, setting

$$\phi_{jk} = 2^{j/2}\phi(2^j x - k), \quad \psi_{jk} = 2^{j/2}\psi(2^j x - k), \tag{6}$$

and letting $V_j \subset L^2(\mathbb{R})$ and $W_j \subset L^2(\mathbb{R})$ be defined by

$$V_j = \mathrm{span}\langle \phi_{jk}, \ k \in \mathbb{Z}\rangle, \quad W_j = \mathrm{span}\langle \psi_{jk}, \ k \in \mathbb{Z}\rangle, \tag{7}$$

the sets $\{\phi_{jk}, \ k \in \mathbb{Z}\}$ and $\{\psi_{jk}, \ k \in \mathbb{Z}\}$ form, respectively, orthonormal bases for V_j and W_j. Moreover, it holds that

$V_j \subset V_{j+1}$ and that $P_j : L^2(\mathbb{R}) \to V_j$ and $Q_j : L^2(\mathbb{R}) \to W_j$, defined as

$$P_j f = \sum_k \langle f, \phi_{jk}\rangle \phi_{jk}, \quad Q_j f = \sum_k \langle f, \psi_{jk}\rangle \psi_{jk},$$

are, respectively, the $L^2(\mathbb{R})$ orthogonal projections onto V_j and W_j. It is also possible to show that W_j is the orthogonal complement of V_j in V_{j+1} and that $Q_j = P_{j+1} - P_j$.

Just to fix the ideas one can consider, as an example, the Haar wavelet. For this case, we have

$$\phi = \chi_{[0,1[}, \quad \psi = \chi_{[0,.5[} - \chi_{[.5,1[}$$

with, for $I \subset \mathbb{R}$,

$$\chi_I(x) = \begin{cases} 1 & x \in I, \\ 0 & \text{otherwise.} \end{cases}$$

For this case, we have

$$h = \{h_0, h_1\} = \left\{ \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right\},$$

$$g = \{g_0, g_1\} = \left\{ \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right\}.$$

The corresponding space V_j is the space of piecewise constants on the dyadic grid with nodes that are integer multiples of 2^{-j}. This simple example is one in a long list of possible orthonormal wavelet bases that we can use as a starting point for the definition of our metric. Within such a list, there are different classes of bases, among which the most significant are Daubechies wavelets, Coiflets and Symlets.

The basis of the one-dimensional wavelet transform of a signal $s = \{s_k, k \in \mathbb{Z}\}$ is to interpret each sample s_k as the k-th coefficient of a function $s^J \in V_J$ for $J > 0$ given:

$$s^J(x) = \sum_k s_k \phi_{Jk}(x).$$

Given that, since $s^J \in V_J$, we have $s^J = P_J s^J$, and that $P_j = P_{j-1} + Q_{j-1}$ for all j, we can write

$$s^J = P_{J-1}s^J + Q_{J-1}s^J = P_{J-2}s^J + \sum_{j=J-2}^{J-1} Q_j s^J$$

$$= \cdots = P_0 s^J + \sum_{j=0}^{J-1} Q_j s^J.$$

Thanks to the orthonormality relations (3), (4) and (5) and to the *two scale relations* (1) and (2), if we let $s^j = \{s^j_n, \ n \in \mathbb{Z}\}$ and $d^j = \{d^j_n, \ n \in \mathbb{Z}\}$ denote the sequence of coefficients of $s^j = P_j s^J = P_j s^{j+1}$ and $d^j = Q_j s^J = Q_j s^{j+1}$ [with a slight abuse of notation, we denote by s^j (resp. d^j) both the

function in V_j—resp. W_j—and the sequence of its coefficient in the basis of expression (7)], a straightforward computation yields the following relations

$$s_n^j = \sum_{k=1}^{L} h_k s_{2n+k}^{j+1}, \qquad d_n^j = \sum_{k=L-1}^{1} g_k s_{2n+k}^{j+1}.$$

In other words, the signals s^j and d^j are obtained from the signal s^{j+1} by a convolution with, respectively, the low-pass filter h and the band-pass filter g, followed by undersampling. By recursively applying such a relation, one obtains the so-called discrete wavelet transform of the signal $s = s^J$

$$s^J \to (s^0, d^0, \ldots, d^{J-1}).$$

This can be seen as a discrete representation of the unitary change of basis in the space V_J

$$s^J = \sum_n s_n^J \phi_{Jn} = \sum_n s_n^0 \phi_{0n} + \sum_{j=0}^{J-1} \sum_{n=1}^{2^j} d_n^j \psi_{jn}.$$

Thanks to the orthonormality of the basis functions, one has the relations

$$\|s^J\|_{L^2(\mathbb{R})}^2 = \|s^J\|_2^2 = \|s^0\|_2^2 + \sum_{j=0}^{J-1} \|d^j\|_2^2.$$

For signals with 2^J nonzero entries, a fast implementation of this procedure is possible, yielding the so-called *fast wavelet transform*.

Remark 1 Like the continuous Fourier transform, the wavelet transform is defined on signals of infinite length. In order to compute the wavelet transform of a finite signal $s = \{s_k, \ k = 0, \ldots, N\}$, one needs to somehow extend it to an infinite signal $s = \{s_k, \ k \in \mathbb{Z}\}$, by suitably defining the values of s_k for $k \notin \{0, \ldots, N\}$. The same thing is done for the Fourier transform, by extending the signal periodically. However, unlike the Fourier transform, the wavelet transform has a local character, so that, if an infinite length signal s^J has a finite number (say $N \sim 2^J$) of nonzero entries, for each j the infinite length signals d^j have also a finite number $\sim 2^j$ of nonzero entries. Besides the periodic extension, this allows other "boundary treatments", like, for instance, the so-called "zero padding", where the signal is extended by zero.

In two dimensions, the wavelet transform is obtained by applying the one-dimensional wavelet transform row-wise and then column-wise. From the functional point of view, this corresponds to considering the tensor product space

$$\mathbb{V}_j = V_j \otimes V_j$$

split as

$$\mathbb{V}_j = \mathbb{V}_{j-1} \oplus \mathbb{W}_{j-1},$$

with

$$\begin{aligned} \mathbb{W}_{j-1} &= \mathbb{W}_{j-1}^{LH} \oplus \mathbb{W}_{j-1}^{HL} \oplus \mathbb{W}_{j-1}^{HH} \\ &= (V_{j-1} \otimes W_{j-1}) \oplus (W_{j-1} \otimes V_{j-1}) \oplus \\ &\quad (W_{j-1} \otimes W_{j-1}), \end{aligned}$$

where the detail spaces

$$\begin{aligned} \mathbb{W}_j^{LH} &= \text{span}\langle \Psi_{j\mathbf{k}}^{LH}, \ \mathbf{k} = (k_1, k_2) \in \mathbb{Z}^2 \rangle, \\ \mathbb{W}_j^{HL} &= \text{span}\langle \Psi_{j\mathbf{k}}^{HL}, \ \mathbf{k} = (k_1, k_2) \in \mathbb{Z}^2 \rangle, \\ \mathbb{W}_j^{HH} &= \text{span}\langle \Psi_{j\mathbf{k}}^{HH}, \ \mathbf{k} = (k_1, k_2) \in \mathbb{Z}^2 \rangle, \end{aligned}$$

are generated, respectively, by tensorized basis functions of the form

$$\begin{aligned} \Psi_{j\mathbf{k}}^{LH}(x, y) &= 2\phi(2^j x - k_1)\psi(2^j x - k_2), \\ \Psi_{j\mathbf{k}}^{HL}(x, y) &= 2\psi(2^j x - k_1)\phi(2^j x - k_2), \\ \Psi_{j\mathbf{k}}^{HH}(x, y) &= 2\psi(2^j x - k_1)\psi(2^j x - k_2). \end{aligned}$$

The two-dimensional wavelet transform is a discrete representation of the change of basis

$$s^J = \sum_{\mathbf{k} \in \mathbb{Z}^2} s_{\mathbf{k}}^J \Phi_{J\mathbf{k}} = s^0 + \sum_{j=0}^{J-1} d^j,$$

with

$$s^0 = \sum_{\mathbf{k} \in \mathbb{Z}^2} s_{\mathbf{k}}^0 \Phi_{0\mathbf{k}}$$

and

$$d^j = \sum_{\mathbf{k} \in \mathbb{Z}^2} d_{j\mathbf{k}}^{LH} \Psi_{j\mathbf{k}}^{LH} + \sum_{\mathbf{k} \in \mathbb{Z}^2} d_{j\mathbf{k}}^{HL} \Psi_{j\mathbf{k}}^{HL} + \sum_{\mathbf{k} \in \mathbb{Z}^2} d_{j\mathbf{k}}^{HH} \Psi_{j\mathbf{k}}^{HH}.$$

Setting

$$\mathbf{d}^j = \left(d_j^{LH}, d_j^{HL}, d_j^{HH} \right),$$

where d_j^{LH}, d_j^{HL} and d_j^{HH} are themselves bivariate signals, the bivariate signal $s^J = \{s_{\mathbf{k}}^J, \mathbf{k} = (k_1, k_2) \in \mathbb{Z}^2\}$ is transformed as

$$s^J \to \text{DWT}(s^J; \phi) = (s^0, \mathbf{d}^0, \ldots, \mathbf{d}^{J-1}). \qquad (8)$$

Remark that, unlike what is usually done, we explicitly indicate in the notation for the discrete wavelet transform, the initial choice of the father wavelet ϕ. This will be convenient

in the following, since different choices of such function will eventually result in different metrics.

Once again, the orthonormality properties of the chosen wavelet basis imply that

$$\|s^J\|_2^2 = \|s^{J-1}\|_2^2 + \|\mathbf{d}^{J-1}\|_2^2 = \|s^0\|_2^2 + \sum_{j=0}^{J-1} \|\mathbf{d}^j\|_2^2, \quad (9)$$

where for $j = 0, \ldots, J$ we set

$$\|s^j\|_2^2 = \sum_{\mathbf{k} \in \mathbb{Z}^2} |s_{\mathbf{k}}^J|^2,$$
$$\|\mathbf{d}^j\|_2^2 = \sum_{\mathbf{k} \in \mathbb{Z}^2} \left(|d_{j\mathbf{k}}^{LH}|^2 + |d_{j\mathbf{k}}^{HL}|^2 + |d_{j\mathbf{k}}^{HH}|^2 \right).$$

2.2 Normalized Metrics

Let f and g be two finite energy sequences in $\ell^2(\mathbb{Z})$. The *normalized root-mean-square error* with stability constant $c \geq 0$ is defined as

$$\text{NRMSE}(f, g; c) = \frac{\|f - g\|_2}{\sqrt{\|f\|_2^2 + \|g\|_2^2 + c}},$$

(if $c = 0$ and both f and g are null, then $\text{NRMSE}(f, g)$ is defined to be 0).

The NRMSE has been proposed in [4,8] and its mathematical properties extensively studied in [4], in the context of finite length signals ($f, g \in \mathbb{R}^N$). The proof of most of the results therein carries over, either unchanged or with minor changes, to the case of signals in $\ell^2(\mathbb{Z})$, that we are considering here. In this section, we recall some of such properties, which we will need in the following, and give some details on how to adapt to our framework the proofs for the finite-dimensional case.

We start by the following theorem.

Theorem 1 *For all $c \geq 0$, the NRMSE is a metric on $\ell^2(\mathbb{Z})$.*

Theorem 1 is a direct consequence of the following Lemma [8]

Lemma 1 *Let $(X, \|\cdot\|)$ be a Ptolemaic normed space and let $M(x, y)$ be a mildly increasing function with respect to x and to y. Then*

$$d_M(f, g) = \frac{\|f - g\|}{M(\|f\|, \|g\|)}$$

is a metric on X if and only if $|x - y|/M(|x|, |y|)$ is a metric on $\mathbb{R}$.

We recall that a function $F : \mathbb{R}^+ \to \mathbb{R}^+$ is said to be mildly increasing if

1. F is increasing
2. $F(x)/x$ is decreasing for $x > 0$.

We also recall that a normed space $(X, \|\cdot\|)$ is said to be ptolemaic if Ptolemy inequality [2]

$$\|x - y\|\|w - z\| + \|y - w\|\|x - z\| \geq \|x - w\|\|y - z\|$$

holds for any four elements $x, y, w, z \in X$.

On the other hand, it is proven in [4] that for

$$M(x, y) = \sqrt{|x|^2 + |y|^2 + c}$$

$d_M(x, y)$ is indeed a metric in $\mathbb{R}$. Since we easily see that M is, in this case, mildly increasing, in order to prove Theorem 1, it remains only to check that the normed space $\ell^2(\mathbb{Z})$ is ptolemaic. This is indeed the case, since a normed space is ptolemaic if and only if it is an inner product space (see [8], Lemma 2.1, and [1], 6.14).

Let now $\{X_m\}_{m=1}^M$ be an orthogonal decomposition of $\ell^2(\mathbb{Z})$ and let $P_m : \ell^2(\mathbb{Z}) \to X_m$ denote the orthogonal projection operator. For $m = 1, \ldots, M$ let δ_m be a metric on X_m and define $\Delta : \ell^2(\mathbb{Z}) \times \ell^2(\mathbb{Z}) \to \mathbb{R}_+^M$ as

$$\Delta(f, g) = (\delta_1(P_1 f, P_1 g), \ldots, \delta_m(P_m f, P_m g)).$$

The following theorem holds, with the same proof as in the finite-dimensional case.

Theorem 2 *Let $\|\cdot\|$ be a norm for $\mathbb{R}^M$ such that for any $a \in \mathbb{R}_+^M$ and $b \in \mathbb{R}_+^M$ it holds that*

$$\|a + b\| \geq \|a\|.$$

Then

$$\delta(f, g) = \|\Delta(f, g)\|$$

is a metric in $\ell^2(\mathbb{Z})$.

The third property that we need to extend to our framework is the existence and the explicit characterization of a convexity region for the square of NRMSE.

Theorem 3 *For g in $\ell^2(\mathbb{Z})$ given, the functional $J : \ell^2(\mathbb{Z}) \to \mathbb{R}$ defined as*

$$J(f) = (NRMSE(f, g; 0))^2$$

is convex on the region

$$X(g) = \{\alpha y_g + \beta u, \ (\alpha, \beta) \in \mathcal{X}, \ u \perp y_g \ unitary\}, \quad (10)$$

where $y_g = g/\|g\|_2$ and where $\mathscr{X} \in \mathbb{R} \times \mathbb{R}^+$ is the set of couples α and β satisfying

$$0 \leq \alpha \leq \sqrt{3}\|g\|_2,$$
$$\beta^2 \leq \frac{16(\alpha^2\|g\|_2^2 + \|g\|_2^4)^{1/2} - 2\alpha^2 - \|g\|_2^2}{2}.$$

The proof for the finite-dimensional case is easily adapted to our case, and we present it for the sake of completeness. It relies on the following technical Lemma (for which we provide a slightly simpler proof).

Lemma 2 *Let $y, \bar{u} \in \ell^2(\mathbb{Z})$ with $\|y\|_2 = \|\bar{u}\|_2 = 1$ and $y \perp \bar{u}$ and let $\bar{X}$ be a convex subset of the two-dimensional subspace spanned by y and $\bar{u}$. Let $F : \ell^2(\mathbb{Z}) \to \mathbb{R}$ be a functional verifying the following assumptions:*

1. *$F|_{\bar{X}}$ is convex;*
2. *F has radial symmetry around the direction of y in the following sense: if*

   $$(x - x') \perp y \text{ and } \|x\|_2 = \|x'\|_2,$$

   then $F(x) = F(x')$;
3. *(Monotonicity) F is not decreasing along the direction $\bar{u}$: if $\beta > \beta' \geq 0$, then $F(\alpha y + \beta \bar{u}) \geq F(\alpha y + \beta'\bar{u})$.*

Then F is convex on the set $X \subset \ell^2(\mathbb{Z})$

$$X = \{x : \exists x' \in \bar{X} \text{ such that } x - x' \perp y \text{ and } \|x\|_2 = \|x'\|_2\}.$$

obtained by rotating $\bar{X}$ around the direction of y.

Proof Let us start by observing that all $x \in \ell^2(\mathbb{Z})$ can be split in a unique way as

$$x = \alpha y + \beta u \text{ with } \beta \geq 0, \ \|u\|_2 = 1, \ u \perp y.$$

Using such a splitting, radial symmetry implies that

$$F(x) = F(\alpha y + \beta u) = F(\alpha y + \beta \bar{u}).$$

Let then $x_1, x_2 \in X$ with

$$x_1 = \alpha_1 y + \beta_1 u_1, \ x_2 = \alpha_2 y + \beta_2 u_2,$$

with $u_1, u_2 \perp y$ unitary, and let

$$z = \lambda x_1 + (1 - \lambda)x_2.$$

We set $\bar{x}_1 = \alpha_1 y + \beta_1 \bar{u}, \bar{x}_2 = \alpha_2 y + \beta_2 \bar{u}, \bar{z} = \lambda \bar{x}_1 + (1 - \lambda)\bar{x}_2 = \bar{\alpha} y + \bar{\beta} \bar{u}$, and we observe that by the convexity of $F|_{\bar{X}}$ we have

$$F(\bar{z}) \leq \lambda F(\bar{x}_1) + (1 - \lambda)F(\bar{x}_2). \tag{11}$$

On the other hand, we have

$$z = (\lambda \alpha_1 + (1 - \lambda)\alpha_2)y + \beta' u' = \bar{\alpha} y + \beta' u'$$

with

$$\beta' = \|\lambda \beta_1 u_1 + (1 - \lambda)\beta_2 u_2\|_2,$$
$$u' = (\lambda \beta_1 u_1 + (1 - \lambda)\beta_2 u_2)/\beta'.$$

We observe that, since $|\langle u_1, u_2 \rangle| \leq 1$,

$$\beta' = \sqrt{\lambda^2 \beta_1^2 + (1 - \lambda)^2 \beta_2^2 + 2\lambda(1 - \lambda)\beta_1\beta_2 \langle u_1, u_2 \rangle}$$
$$\leq \sqrt{\lambda^2 \beta_1^2 + (1 - \lambda)^2 \beta_2^2 + 2\lambda(1 - \lambda)\beta_1\beta_2} = \bar{\beta}.$$

Using radial symmetry and monotonicity, as well as (11), we can then write

$$F(z) = F(\bar{\alpha} y + \beta' u') = F(\bar{\alpha} y + \beta' \bar{u}) \leq F(\bar{\alpha} y + \bar{\beta}\bar{u})$$
$$\leq \lambda F(\bar{x}_1) + (1 - \lambda)F(\bar{x}_2) = \lambda F(x_1) + (1 - \lambda)F(x_2),$$

which proves the thesis. □

In view of Lemma 2, Theorem 3 is easily proven. Let f be split as $f = \alpha y_g + \beta u$ with $\alpha = \langle f, y_g \rangle, \beta = \|f - \alpha y_g\|_2$ and $u = (f - \alpha y_g)/\beta$. Observe that u is a unitary vector orthogonal to g. It is not difficult to check that J verifies

$$J(f) = \frac{(\alpha - \|g\|)^2 + \beta^2}{\alpha^2 + \beta^2 + \|g\|^2}.$$

Radial symmetry and monotonicity (in the sense of the assumptions of Lemma 2) are easily verified. Let us then fix an unitary vector $\bar{u}$ and verify that, setting

$$\bar{X} = \{f = \alpha y_g + \beta \bar{u} : (\alpha, \beta) \in \mathscr{X}\},$$

$J|_{\bar{X}}$ is convex. In order to do so, we observe that y_g and $\bar{u}$ form an orthonormal basis for their span, with α and β being the corresponding orthogonal coordinates. We can then directly compute the eigenvalues of the Hessian of $J|_{\bar{X}}$:

$$\lambda_+ = \frac{8\alpha\|g\|_2^3 + 4\|g\|_2\sqrt{\Delta}}{(\alpha^2 + \beta^2 + \|g\|_2^2)^3}, \quad \lambda_- = \frac{8\alpha\|g\|_2^3 - 4\|g\|_2\sqrt{\Delta}}{(\alpha^2 + \beta^2 + \|g\|_2^2)^3},$$

with

$$\Delta = \alpha^6 + 3\alpha^4\beta^2 - 2\alpha^4\|g\|_2^2 + 3\alpha^2\beta^4 + \alpha^2\|g\|_2^4$$
$$+ \beta^6 + 2\beta^4\|g\|_2^2 + \beta^2\|g\|_2^4,$$

and verify that they are both real and nonnegative if and only if $0 \leq \alpha \leq \sqrt{3}\|g\|_2$, and

$$\beta^2 \le \frac{16(\alpha^2\|g\|_2^2 + \|g\|_2^4)^{1/2} - 2\alpha^2 - \|g\|_2^2}{2}.$$

Lemma 2 then yields the thesis.

2.3 The SSIM and the SSIM Related Metrics

Let us recall the definition of the *structural similarity index* (SSIM). Letting f and g denote throughout this secion two $N \times M$ images to be compared, we let $\bar{f}$ and $\bar{g}$, respectively, denote the mean value of f and g and σ_f^2 and σ_g^2 the respective variances, while $\sigma_{f,g}$ denotes the covariance of the two images. The structural similarity index, introduced in [32], (locally) compares the two images by suitably combining independent measures of the similarity in luminance, contrast and structure:

$$l(f, g) = \frac{2\bar{f}\bar{g} + c_1}{\bar{f}^2 + \bar{g}^2 + c_1}, \quad c(f, g) = \frac{2\sigma_f\sigma_g + c_2}{\sigma_f^2 + \sigma_g^2 + c_2},$$
$$s(f, g) = \frac{\sigma_{f,g} + c_3}{\sigma_f\sigma_g + c_3},$$

where c_1, c_2 and c_3 are small constants whose purpose is on the one hand to take into account the saturation effects of the visual system at low luminance and contrast, and, on the other hand, to ensure stability when the denominators approach zero. The original definition of SSIM is

$$\text{SSIM}(f, g) = l(f, g)c(f, g)s(f, g).$$

If we choose $c_3 = c_2/2$, then such an expression reduces to

$$\text{SSIM}(f, g) = S_1(f, g)S_2(f, g)$$

with

$$S_1(f, g) = l(f, g) = \frac{2\bar{f}\bar{g} + c_1}{\bar{f}^2 + \bar{g}^2 + c_1},$$

and

$$S_2(f, g) = c(f, g) \cdot s(f, g) = \frac{2\sigma_{f,g} + c_2}{\sigma_f^2 + \sigma_g^2 + c_2}.$$

In [4], the authors show that

$$\text{NRMSE}(\bar{f}, \bar{g}; c_1) = \sqrt{1 - S_1(f, g)}. \tag{12}$$

and that (for $c_2' = (NM - 1)c_2/(NM)$)

$$\text{NRMSE}(f - \bar{f}, g - \bar{g}; c_2') = \sqrt{1 - S_2(f, g)}. \tag{13}$$

Thanks to Theorem 2, it is then easy to deduce that, for all $\alpha > 0, \omega > 0, 1 \le q < \infty$

$$d(f, g; q) := \sqrt[q]{\alpha(1 - S_1(f, g))^{q/2} + \omega(1 - S_2(f, g))^{q/2}}$$

and

$$d(f, g; \infty) := \max\left\{\alpha\sqrt{1 - S_1(f, g)}, \omega\sqrt{1 - S_2(f, g)}\right\}$$

are metrics.

Remark 2 While the new class of metrics that we are going to introduce in the following section will be defined on bi-infinite signals and will be computed on finite images by applying one of the extension techniques mentioned in Remark 1, the SSIM is directly defined for finite-dimensional $N \times M$ signals. It is worth remarking that the ratio c_2'/c_2 between the constants appearing in relation (13) tends to 1 as NM tends to $+\infty$.

2.4 Definition of the New Metrics

We are now ready to introduce the new class of metrics. Let $f \in \ell^2(\mathbb{Z})$ and $g \in \ell^2(\mathbb{Z})$ denote the two (bi-infinite) images to be compared. Choosing a father wavelet ϕ, we let

$$DWT(f; \phi) = (f^0, \mathbf{d}_f^0, \dots, \mathbf{d}_f^{J-1}),$$
$$DWT(g; \phi) = (g^0, \mathbf{d}_g^0, \dots, \mathbf{d}_g^{J-1})$$

denote the respective discrete wavelet transforms, defined according to Sect. 2.1. Given two small constants $c_1 \ge 0$ and $c_2 \ge 0$ we can define nonnegative functions ρ and δ^j by

$$|\rho(f, g; \phi)|^2 = \frac{\|f^0 - g^0\|_2^2}{\|f^0\|_2^2 + \|g^0\|_2^2 + c_1}, \tag{14}$$

$$|\delta^j(f, g; \phi)|^2 = \frac{\|\mathbf{d}_f^j - \mathbf{d}_g^j\|_2^2}{\|\mathbf{d}_f^j\|_2^2 + \|\mathbf{d}_g^j\|_2^2 + c_2}. \tag{15}$$

Given constants $\alpha > 0$ and $\omega_j > 0$, and given $1 \le q < \infty$, we let the nonnegative function $\delta(\cdot, \cdot; \phi, q)$ be defined by

$$|\delta(f, g; \phi, q)|^q = \alpha|\rho(f, g; \phi)|^q + \sum_{j=0}^{J-1} \omega_j|\delta^j(f, g; \phi)|^q, \tag{16}$$

and

$$\delta(f, g; \phi, \infty) = \max\{\alpha\rho(f, g; \phi),$$
$$\omega_j\delta^j(f, g; \phi), 0 \le j \le J - 1\}. \tag{17}$$

We have the following theorem.

Theorem 4 *For all $c_1 \geq 0$, $c_2 \geq 0$, $1 \leq q \leq \infty$ and for all choices of the orthonormal father wavelet ϕ, $\delta(\cdot, \cdot; \phi, q)$ is a metric on $\ell^2(\mathbb{Z}^2)$.*

The proof of Theorem 4 is straightforward, since the weighted ℓ^q norm on $\mathbb{R}^{J+1}$ involved in the definition of δ satisfies the assumptions of Theorem 2. It is enough to combine such a theorem with Theorem 1.

Remark 3 Theorem 4 is stated for infinite-dimensional signals, and holds for any choice of the number J of decomposition levels. Of course, in practice we will be dealing with finite-dimensional $N \times M$ images and it makes sense to choose $J \sim \log(\min\{N, M\}$. Moreover, the boundary treatment method (see Remark 1) should be chosen in such a way that the wavelet transform allows for perfect reconstruction.

Clearly depending on the choice of the father wavelet ϕ, of the constants c_1 and c_2, of the coefficients α and ω_j and of the parameter $q \in [0, +\infty]$, we have a whole set of different metrics. To fix the ideas in the following, we fix $c_1 = c_2 = 0$, $\alpha = \omega_j = 1$ and $q = 2$ and we will consider the class of metrics depending on ϕ defined as

$$|\delta(f, g; \phi)|^2 = |\rho(f, g; \phi)|^2 + \sum_{j=0}^{J-1} |\delta^j(f, g; \phi)|^2,$$

where omitting the reference to the parameter q in the notation for the metric implies the choice $q = 2$. We have the following corollary.

Corollary 1 *For all choices of the orthonormal father wavelet ϕ, $\delta(\cdot, \cdot; \phi)$ is a metric on $\ell^2(\mathbb{Z}^2)$.*

We will call the new metric *wavelet normalized root-mean-square error* (WNRMSE) and introduce the notation

$$\text{WNRMSE}(f, g; \phi) = \delta(f, g; \phi).$$

2.5 Differentiability and Convexity

Many image processing tasks can be performed via an optimization problem where an image is sought minimizing a given functional involving an image metric. In order to solve the optimization problem, it is then desirable that the metric involved is differentiable and that it has some convexity property. Let then g be a fixed image and consider the functionals $H_j : \ell^2(\mathbb{Z}^2) \to \mathbb{R}$, $K : \ell^2(\mathbb{Z}^2) \to \mathbb{R}$, and $J : \ell^2(\mathbb{Z}^2) \to \mathbb{R}$, defined as

$$H_j(f) = |\delta^j(f, g; \phi)|^2, \quad K(f) = |\rho(f, g; \phi)|^2,$$

$$J(f) = \delta(f, g; \phi, q) = \left(\alpha K(f)^{q/2} + \sum_{j=0}^{J-1} \omega_j H_j(f)^{q/2} \right)^{1/q}.$$

The following theorem holds

Theorem 5 *The functional $J : \ell^2(\mathbb{Z}^2) \to \mathbb{R}$ is Fréchet differentiable, and its Fréchet derivative $\mathscr{D}J$ is*

$$\mathscr{D}J(f) = \frac{1}{2} J(f)^{1/q-1} \Big(\alpha K(f)^{q/2-1} \mathscr{D}K(f) $$
$$+ \sum_{j=0}^{J-1} \omega_j H_j(f)^{q/2-1} \mathscr{D}H_j(f) \Big) \quad (18)$$

with

$$\mathscr{D}K(f) = 2 \frac{2\langle f_0, g_0 \rangle + c_1}{(\|f_0\|_2^2 + \|g_0\|_2^2 + c_1)^2} f_0 $$
$$- \frac{2g_0}{\|f_0\|_2^2 + \|g_0\|_2^2 + c_1}, \quad (19)$$

$$\mathscr{D}H_j(f) = 2 \frac{2\langle \mathbf{d}_f^j, \mathbf{d}_g^j \rangle + c_2}{(\|\mathbf{d}_f^j\|_2^2 + \|\mathbf{d}_g^j\|_2^2 + c_2)^2} \mathbf{d}_f^j $$
$$- \frac{2\mathbf{d}_j^g}{\|\mathbf{d}_f^j\|_2^2 + \|\mathbf{d}_g^j\|_2^2 + c_2}. \quad (20)$$

Proof We start by proving (20). Let

$$N_j(f) := \|\mathbf{d}_f^j - \mathbf{d}_g^j\|_2^2, \qquad D_j(f) := \|\mathbf{d}_f^j\|_2^2 + c,$$

where $c = \|\mathbf{d}_g^j\|_2^2 + c_2$ is a constant term which does not depend on f, denote the numerator and the denominator of the fraction at the right-hand side of (15). The Fréchet derivatives of N_j and D_j are simply

$$\mathscr{D}N_j = 2(\mathbf{d}_f^j - \mathbf{d}_g^j) \quad \text{and} \quad \mathscr{D}D_j = 2\mathbf{d}_f^j.$$

Applying the chain rule and the rule on the Fréchet derivative of the product of two scalar functions, we have

$$\mathscr{D}H_j(f) = \frac{\mathscr{D}N_j(f) D_j(f) - N_j(f) \mathscr{D}D_j(f)}{D_j^2(f)} $$
$$= \frac{2(\mathbf{d}_f^j - \mathbf{d}_g^j)(\|\mathbf{d}_f^j\|_2^2 + c) - \|\mathbf{d}_g^j - \mathbf{d}_f^j\|_2^2 2\mathbf{d}_f^j}{(\|\mathbf{d}_f^j\|_2^2 + c)^2}. \quad (21)$$

A straightforward calculation yields

$$2(\mathbf{d}_f^j - \mathbf{d}_g^j)(\|\mathbf{d}_f^j\|_2^2 + c) - 2\|\mathbf{d}_g^j - \mathbf{d}_f^j\|_2^2 \mathbf{d}_f^j $$
$$= 2[-\mathbf{d}_g^j D_j(f) + c_2 \mathbf{d}_f^j + 2\langle \mathbf{d}_f^j, \mathbf{d}_g^j \rangle \mathbf{d}_f^j].$$

Hence, we obtain (20). Equation (19) is proven analogously. Equation (18) follows by applying the chain rule. $\square$

We now restrict ourselves to the case $c_1 = c_2 = 0$ and $q = 2$, and we turn to the convexity of the square of the

functional J. A straightforward application of Theorem 3 yields the following Proposition.

Proposition 1 *Let g be such that $g^0 \neq 0$ and $\mathbf{d}_g^j \neq 0$ for all j. Then the functionals $K(f)$ and $H_j(f)$ are convex, respectively, in the regions $X^0(g)$ and $X_j(g)$*

$$X^0(g) = \{f : f^0 \in X(g^0)\}, \quad X_j(g) = \{f : \mathbf{d}_f^j \in X(\mathbf{d}_g^j)\}.$$

We recall that, for $y \in \ell^2(\mathbb{Z})$, $X(y)$ is defined by (10). By linearity, it is then easy to deduce the convexity of the functional $J(f)^2$ in the region

$$\mathbb{X}(g) = X^0(g) \cap_j X_j(g) = \{f : f^0 \in X^0(g), \ \mathbf{d}_f^j \in X_j(g)\}.$$

Remark 4 The condition that $g^0 \neq 0$ and $\mathbf{d}_g^j \neq 0$ for all j is not very restrictive, and it is in general satisfied. However, there might be cases in which $\mathbf{d}_g^j$ vanishes for some j. In such case, for $c_2 = 0$ we will have $J_j(f) = 1$, so that the discrepancy at level j will not be effectively measured by the metric. In such cases, it is crucial to choose $c_2 \neq 0$. The convexity region of the functional $J_j(f)^2$ will then depend on the constant c_2 (see [4]).

2.6 Comparison with the SSIM-Based Indexes and Metrics

The simplest choice of father wavelet ϕ is $\phi = \chi_{[0,1[}$, which yields the Haar transform. In this case, the one-dimensional transform takes a quite simple form:

$$s_k^j = \frac{1}{\sqrt{2}}\left(s_{2k}^{j+1} + s_{2k+1}^{j+1}\right), \quad d_k^j = \frac{1}{\sqrt{2}}\left(s_{2k}^{j+1} - s_{2k+1}^{j+1}\right).$$

If the signal is supported in $\left[0, 2^J - 1\right]$ (which is equivalent to the function $s = \sum_k s_k \phi_{Jk}$ being supported in $[0, 1]$), then

$$s^0(x) = s_0^0 \phi(x), \ s_0^0 = \int_0^1 s^J(x)\,\mathrm{d}x = 2^{-J/2}\sum_k s_k.$$

In other words, s^0 is identified by a single number, which, modulo a factor depending on J, is nothing else than the mean value of the original signal s^J. This observation allows us to draw a parallel between the new class of image metrics that we introduced and the metrics derived from SSIM in [4] and described in Sect. 2.3. By comparing the definition of $d(f, g; q)$ and of $\delta(f, g; \phi, q)$ we see that, at least when using the Haar transform in the definition of δ, the contribution of the mean component to the two metrics is the same. In the new metric, the contribution of the mean free component is further split into dyadic frequency bands, with the aid of the wavelet transform, and the contribution of each frequency band independently undergoes divisive renormalization.

Other multiscale versions of SSIM have been proposed in the literature. We recall, in particular, the Multiscale SSIM (MSSSIM, introduced in [35]) and the Complex Wavelet SSIM (CWSSIM, introduced in [26]). Of course, these, as well as SSIM, are all similarity indexes (assuming higher values as similarity increases) while WNRMSE is a distance (assuming lower values as similarity increases). This, however, is, in itself, only a minor difference: a similarity index and a distance, if suitably related to each other, might result in equivalent classification criteria. Let us then look more closely at the differences between WNRMSE and, respectively, MSSSIM and CWSSIM.

MSSSIM relies on calculating the SSIM on a sequence of approximations of the given images at different resolutions obtained by successive applications of a low-pass filter, which we can choose to be the one derived from the wavelet basis involved in the definition of WNRMSE. However, differently from WNRMSE, MSSSIM does not make direct use of the details which result from the application of a band-pass filter. It is of course possible to apply the reasoning of [4] to the MSSSIM, and this will lead to a metric obtained by building blocks of the form

$$|\vartheta_j(f, g; c)|^2 = \frac{\|(f^j - f^0) - (g^j - g^0)\|_2^2}{\|f^j - f^0\|_2^2 + \|g^j - g^0\|_2^2 + c}$$

To underline the difference of the two metrics remark that if f and g are such that $\mathbf{d}_f^j = \mathbf{d}_g^j$ but $f^{j-1} \neq g^{j-1}$, we have

$$\delta_j(f, g; \phi) = 0 \neq \vartheta_j(f, g; c),$$

while, if $\mathbf{d}_f^j \neq \mathbf{d}_g^j$ but $f^{j-1} = g^{j-1}$ we have

$$
\begin{aligned}
|\vartheta_j(f, g; c)|^2 &= \frac{\|(f^j - g^j) - (f^0 - g^0)\|_2^2}{\|f^j - f^0\|_2^2 + \|g^j - g^0\|_2^2 + c} \\
&= \frac{\|\mathbf{d}_f^j - \mathbf{d}_g^j\|_2^2}{2\|f^{j-1} - f^0\|_2^2 + \|\mathbf{d}_f^j\|_2^2 + \|\mathbf{d}_g^j\|_2^2 + c} \\
&= C(f)|\delta_j(f, g; \phi)|^2,
\end{aligned}
$$

where $C(f)$ goes to zero as $\|f^{j-1}\|_2$ goes to infinity. In other words, we can either have $\delta_j \ll \vartheta_j$ or $\vartheta_j \ll \delta_j$.

CWSSIM is also obtained by first decomposing the images using, this time, a complex version of a so-called shiftable wavelet transform [27], specifically designed to avoid aliasing in frequency bands. This results in a redundant representations of the images f and g as two sets of sub-band complex coefficient matrices $\mathbf{d}_f^{j,\eta}$, and $\mathbf{d}_g^{j,\eta}$, each indexed by level j and orientation η. If, for any two given complex signals $\mathbf{x} = (x_i)$ and $\mathbf{y} = (y_i)$, we define the complex similarity index as

$$S_c(\mathbf{x}, \mathbf{y}) = \frac{2|\sum x_i y_i^*| + c}{\sum_i |x_i|^2 + \sum_i |y_i|^2 + c},$$

(where for $x \in \mathbb{C}$, we denote by x^* its complex conjugate), the CWSSIM index is computed by first applying S_c to each sub-band and then averaging across sub-bands.

Letting aside the obvious difference deriving from using different classes of wavelets, it is natural to ask oneself if a relation of the form (13) holds for S_c, which would strongly relate CWSSIM and WNRMSE. This is not the case, due to the complex nature of the transform underlying CWSSIM. In fact, setting

$$\|\mathbf{x}\|_2^2 = \mathbf{x} \cdot \mathbf{x}^* \quad \text{where} \quad \mathbf{x} \cdot \mathbf{y} = \sum_i x_i y_i^*,$$

and letting $\mathfrak{Re}(x)$ and $\mathfrak{Im}(x)$ denote, respectively, the real and imaginary part of the complex number x, on the one hand we have

$$\frac{\|\mathbf{x} - \mathbf{y}\|_2^2}{\|\mathbf{x}\|_2^2 + \|\mathbf{y}\|_2^2 + c} = 1 - \frac{2\mathfrak{Re}(\mathbf{x} \cdot \mathbf{y}^*) + c}{\|\mathbf{x}\|_2^2 + \|\mathbf{y}\|_2^2 + c}$$

while, on the other hand

$$\begin{aligned} S_c(\mathbf{x}, \mathbf{y}) &= \frac{2\sqrt{\mathfrak{Re}(\mathbf{x} \cdot \mathbf{y}^*)^2 + \mathfrak{Im}(\mathbf{x} \cdot \mathbf{y}^*)^2} + c}{\|\mathbf{x}\|_2^2 + \|\mathbf{y}\|_2^2 + c} \\ &\geq \frac{2\mathfrak{Re}(\mathbf{x} \cdot \mathbf{y}^*) + c}{\|\mathbf{x}\|_2^2 + \|\mathbf{y}\|_2^2 + c}, \end{aligned}$$

yielding

$$\frac{\|\mathbf{x} - \mathbf{y}\|_2^2}{\|\mathbf{x}\|_2^2 + \|\mathbf{y}\|_2^2 + c} \geq 1 - S_c(\mathbf{x}, \mathbf{y}),$$

where the equality holds if and only if $\mathbf{x} \cdot \mathbf{y}^*$ is real and nonnegative. Then, if we were to carry out our construction starting from the same frequency decomposition used for building the CWSSIM index, the resulting metric would, in general, give us a different classification: the distance might be large, even for images that the index S_c classifies as similar. This is of course coherent with the reduced sensitivity of CWSSIM to translations and rotations, which is not shared by WNRMSE: two rotated and/or translated versions of the same image are classified as very close by CWSSIM ($S_c \sim 1$) but can still be seen as distant by WNRMSE.

Another significant difference between WNRMSE and CWSSIM lies in the computational cost, which is sensibly greater for CWSSIM, due to the high redundancy involved in the complex wavelet transform (in our case obtained by steerable pyramid algorithms [27], as proposed in [26]), as opposed to the non-redundant orthogonal wavelet transform at the basis of WNRMSE. To give a rough idea, while one

evaluation of WNRMSE and MSSSIM are, respectively, performed in .0505 and .0593 s,[1] one evaluation of CWSSIM requires .7322 s, that is, it is roughly 15 times more expensive. Of course, the use of a more expensive redundant transform gives CWSSIM more robustness with respect to translations, scaling and rotation, an highly desirable feature in many (though not in all) applications.

2.7 Relation with the Besov Distance

Another class of distances to which we can relate to the new metric is the one derived from the norms relative to Besov spaces. These are a class of function spaces introduced in the fifties by O.V. Besov with the aim of generalizing more simple function spaces (namely, Sobolev spaces) and effectively measuring smoothness. It is way beyond the scope of this paper to give the original definition of the related norms. The thing that is interesting for comparison with the WNRMSE metrics is the characterization of such spaces via wavelet coefficients. We recall that the infinite set

$$\{\phi_{0k}, \; k \in \mathbb{Z}\} \cup_{j \geq 0}^{\infty} \{\psi_{jk}, \; k \in \mathbb{Z}\}$$

forms an orthonormal basis for the space $L^2(\mathbb{R})$ of finite energy functions. Letting

$$s(x) = \sum_{k \in \mathbb{Z}} s_k^0 \phi_{0k}(x) + \sum_{j=0}^{\infty} \sum_{k \in \mathbb{Z}} d_k^j \psi_{jk}(x),$$

the Besov norm of s, which depends on three parameters $\beta \in \mathbb{R}$, $p, q \in [1, \infty]$, is equivalent (in the sense that it yields the same function space) to the following expression

$$\|s\|_{\beta,p,q}^q = \left(\sum_{k \in \mathbb{Z}} |s_k^0|^p \right)^{q/p} + \sum_{j=0}^{\infty} \omega_j \left(\sum_{k \in \mathbb{Z}} |d_k^j|^p \right)^{q/p},$$

(22)

with $\omega_j = 2^{qj(\beta+1/2-1/p)}$. For $p = 2$, this reduces to

$$\|s\|_{\beta,2,q}^q = \|s^0\|_2^q + \sum_{j=0}^{\infty} 2^{q\beta j} \|d^j\|_2^q.$$

The Besov distance of two functions f and g is then

$$\left(\|s_f^0 - s_g^0\|_2^q + \sum_{j=0}^{\infty} 2^{q\beta j} \|d_f^j - d_g^j\|_2^q \right)^{1/q}.$$

---

[1] Average computed over 100 evaluations on images of size 512×384 pixels.

If we were to apply divisive renormalization to each of the term in this sum, we would obtain something quite similar to WNRMSE. In a way, we can say that WNRMSE can be obtained by combining the Besov distance with frequency band-wise divisive renormalization.

The parallel between WNRMSE and the Besov norm also allows to give an interpretation to the different parameters in the definition (16) and to justify the choices we made for such parameters in our numerical tests. The first, quite natural, choice was implicitly made before defining δ, and it is to fix $p = 2$. This puts us in the framework of inner product spaces, allowing, for instance, to prove Theorem 4.

The coefficient ω_j in the definition (22) of the Besov norm is directly related to the smoothness of the function s, and, more precisely, to its degree of differentiability. As images are generally not smooth, we chose $\omega_j = 1$ (corresponding to $\beta = 0$). The role of the parameter q is more subtle. It allows for a kind of fine-tuning of the smoothness, beyond the action of the coefficients ω_j [30]. It is worth remarking—this observation is at the basis of our choice—that for $p = q = 2$, the Besov norm is equivalent to the norm of the Sobolev space of functions in $L^2(\mathbb{R})$ with (for β integer) β derivatives in $L^2(\mathbb{R})$, which is, for $\beta = 0$, a quite natural space for image processing.

Clearly, the expression on the right-hand side of (22) depends on the choice of the wavelet (that is of the functions ϕ and ψ). If ϕ and $\tilde{\phi}$ are two different orthonormal scaling functions and ψ and $\tilde{\psi}$ the relative wavelet functions, for each function $s \in L^2(\mathbb{R})$, we have two different wavelet decompositions

$$
\begin{aligned}
s(x) &= \sum_k s_k^0 \phi_{0,k}(x) + \sum_{j=0}^{\infty} \sum_k d_k^j \psi_{j,k}(x) \\
&= \sum_k \tilde{s}_k^0 \tilde{\phi}_{0,k}^2(x) + \sum_{j=0}^{\infty} \sum_k \tilde{d}_k^j \tilde{\psi}_{j,k}^2(x)
\end{aligned}
$$

and, consequently, two different norms, the first defined by (22) and the second by

$$
\|s\|_{\beta,p,q}^q = \left(\sum_{k \in \mathbb{Z}} |\tilde{s}_k^0|^p \right)^{q/p} + \sum_{j=0}^{\infty} \omega_j \left(\sum_{k \in \mathbb{Z}} |\tilde{d}_k^j|^p \right)^{q/p}.
\tag{23}
$$

It is possible to prove that, provided ϕ and $\tilde{\phi}$ are sufficiently regular, the two norms are equivalent, in the sense that there exists a positive constant $C = C(\phi, \tilde{\phi}, \beta, p, q)$ such that for all s

$$
C^{-1}\|s\|_{\beta,p,q} \leq \|s\|_{\beta,p,q} \leq C\|s\|_{\beta,p,q}.
$$

It would be interesting to know whether such a property also holds for WNRMSE, whose definition does, of course, depend on the wavelet chosen to perform the frequency splitting. We were not able to prove a property of this type, but the first simple numerical tests seem to indicate that this might indeed be the case.

2.8 Extension to Color Images

WNRMSE is defined on monochromatic images and, in order to use it for assessing the similarity of color images it is, for many purposes, sufficient to apply it to the luminance component (see, for instance, the tests in Sect. 4). It is possible, however, to define a distance that takes into account the chrominance information by suitably combining the results of WNRMSE applied to the luminance component and to the two chrominance components. More precisely let f_Y, f_I, f_Q and g_Y, g_I, g_Q denote the three components of the two color images f and g in the YIQ color space. We can define the distance of f and g as

$$
\text{WNRMSE}_C(f, g; \phi) = \sqrt{\alpha_Y \delta_Y^2 + \alpha_I \delta_I^2 + \alpha_Q \delta_Q^2}
$$

with α_Y, α_I and α_Q weights measuring the relative importance of the three components, and with

$$
\begin{aligned}
\delta_Y &= \text{WNRMSE}(f_Y, g_Y; \phi), \\
\delta_I &= \text{WNRMSE}(f_I, g_I; \phi), \\
\delta_Q &= \text{WNRMSE}(f_Q, g_Q; \phi).
\end{aligned}
$$

Proving the following Proposition is straightforward.

Proposition 2 *If the triple* $(\alpha_Y, \alpha_I, \alpha_Q)$ *satisfies* $\alpha_Y > 0$, $\alpha_I > 0$, $\alpha_Q > 0$, *then WNRMSE$_C$ is a metric.*

Remark that, while the choice of the coefficients α_Y, α_I and α_Q gives us an easy way to take into account the relative weight of the luminance and chrominance components of the images—which, of course will depend on the target application, considering the more general distance defined in equation (16) would give us a much greater flexibility in treating the different components of the YIQ decomposition of the image. One could, for instance, take advantage of the weights ω_j in such an equation to take into account the known fact that the human eye is less sensitive to high frequencies in the chrominance components than to the same frequencies in the luminance components. One could also use the parameter q to fine-tune the distance to better mimic the HVS. Moreover, one could, of course, apply the same approach to other color spaces, such as HSI or CieLAB. Fully exploring these possibilities is, however, beyond the scope of this paper and will be the object of a future study.

Fig. 1 Twenty-five original images of the TID2013 database

3 The Experimental Validation

3.1 The Reference Database

The first problem in validating a criterion for image similarity assessment is the choice of the reference database. We test the WNRMSE metric by considering the TID2013 database [23,24]. The reasons are various: First, it contains images which differ a lot, in terms of both visual properties (color distribution, texture) and semantic content, i.e., they depict very different subjects, such as faces, objects of everyday life, outdoor scenes, animals, and synthetic patterns. The database has been populated starting from 25 original, unimpaired (i.e., of the best available quality) images. Let us denote the original images with Q_i, $i \in [1, 25]$. In Fig. 1, the snapshots of the 25 original images are shown: Their real size is 512×384 pixels. In order to populate the database, each image has been impaired by 24 different distortion types. The great variety of the distortions is the second reason for the choice of the TID2013 database. In fact, the distortion types include: Gaussian and spatially correlated noise, masking, high frequency, quantization, impulse noise, lossy JPEG and JPEG2000 compression, transmission errors, change of color saturation, contrast change, intensity shift degradation and chromatic aberrations. All these distortion types are applied at 5 different levels of increasing impairment, for a total of 3000 degraded images (identified with D_j, $j \in [1, 3000]$). Each Q_i has 120 impaired versions (120 images in the class C_{Q_i}). The large number of distortion types covers almost all the main fields of image processing, such as image acquisition, registration, compression, watermarking and transmission. This assures that the validation of the metric is valuable for a large variety of applications. From a perceptual point of view, the different impairments stress several features of the HVS, such as color, contrast, spatial and frequency sensitivity, robustness and adaptivity.

We choose TID2013 also because it includes a subjective quality evaluation for each impaired image D_j; each image is labeled by a mean opinion score (MOS) value, expressed over a set of more than 900 human observers, a very high number of testers, which makes the TID2013 one of the most interesting database in the literature. Let denote with MOS_j the MOS of the image D_j; it is a number on a scale from 0 to 9 (0—minimal, 9—maximal quality), which expresses the quality as perceived by the human tester when he/she watches the image on the monitor. We have used the MOS information to perform the third set of experiments (see Sect. 4 for details), in order to involve explicitly in our validation a perceptive feature, related to the human visual system.

Last, but not least, the TID2008 database [22,25], a subset of TID2013, has been already used in previous contributions to test performance of various metrics on different tasks [3, 17, 19].

3.2 The Methodology of the Experimental Validation

A well-established methodology consists in considering the similarity assessment as a a typical classification problem: We analyze at which extent each "metric" is able to classify a generic image D_j of the database, as belonging to the correct C_{Q_i} class, where *correct* means that D_j is one of the 120 impaired version of Q_i (the *query* image). In order to test the performance of each metric under analysis, we use the consolidated approaches in the literature of computing the area under curve (AUC) of a receiver operating characteristic (ROC) analysis [7], and the empirical probability of inter-class and intra-class membership. In the intra-class classification task, the query image used in the test is compared to all the images belonging to the same class C_{Q_i} of the test image, while in the inter-class test it is compared to all the other images, namely all the images in the database except the ones of its own class. Classification may fail because similarity is not well determined, namely images which are similar (i.e., impaired version of the same original image) are considered dissimilar, or vice versa. Considering a binary classifier depending on a threshold parameter, the ROC curve measures its ability to produce good scores using a 2D space: it plots the True Positive Rate (TPR, on the Y axis) of the classifier against its False Positive Rate (FPR, on the X axis) as the threshold varies. A perfect classification is possible whenever the ROC curve passes through the point $(0, 1)$, implying the existence of threshold parameters yielding 0 false positive and 100% of true positives; the corresponding AUC is 1. In Sect. 4, AUC and ROC analysis is presented for different measurements of similarity (or dissimilarity) mentioned in Sect. 1 (to which we will from now on improperly refer as *metrics*, though most of them are not metrics in the proper mathematical sense of the term), namely WNRMSE, SSIM, MSSSIM, CWSSIM, FPH and PSNR.

The experimental validation consists of three different tasks:

– The identification of which wavelet is best suited for WNRMSE, using as query image each of the 25 original images Q_i and varying the mother wavelet function.
– The comparison of the performance of WNRMSE to PSNR, SSIM, MSSSIM, CWSSIM and FPH, by using as query image each of the 25 original images Q_i.
– The comparison of the same metrics in a more challenging case: when for each i, the query image Q_i is replaced by the corresponding image $D_j \in C_{Q_i}$ which has the lowest MOS.

We consider the third task as one of the theoretical novelties of this research. In fact, we choose as query image the one which has the worst perceived quality. We call this validation a *perceptive approach*, as the goal is to study the robustness of each similarity (or dissimilarity) measurement, according to a subjective evaluation driven by the perceived quality of the image by the human observer. The goal is to individuate the best metric when similarity is evaluated in real applications, where images are very often impaired by different types of distortion. Moreover, distortion is not evaluated in an objective way, by considering its traditional formulation (usually, a component superimposed on the original signal) but in a subjective way, according the HVS of the observer. Another advantage of the perceptive approach is that in this worst-case analysis, we consider a specific kind of distortion for each image (the one which minimize the MOS) and not the same distortion for all the images of the database. This is important because the validation takes into account the fact that the effect of a given distortion is not the same for all the images, but it depends on the image content itself. For example, some images are more sensitive to high-frequency noise, other to impulse noise, and so on. By using the lower MOS we assure that, for each image, we perform really the worst-case analysis, by using the human observer's final judgment.

As most of the image similarity measures do not include color management [14], we perform the above tasks on a converted color-to-grayscale version of TID2013. One might object that in the third experimental task (worst quality query images), the choice of the query image is driven by the MOS expressed by the human observers on the color images, while the classification is carried out on the grayscale images. Obviously, we have no guarantees that MOS values remain unchanged when switching to gray scale. However, the approach of using correlation between MOS expressed on color images and intensity (luminance) component-based metrics is pretty common in the literature [24]. Moreover, in our perceptive approach we do not actually use the MOS for comparison with the results of the different metrics, but only as a criterion to choose

a "bad" query image, as opposed to choosing the "bad" query image randomly or using one of the metrics, which would possibly be disfavored in the subsequent tests. In other words, the use of the MOS (though for the color images) is one way of choosing a "bad" representative of a class, without introducing a bias toward one of the metrics.

Remark that using a grayscale version of the images to be classified is not necessarily a downside, since for the kind of classification task that we are considering here, the color information might be misleading, and taking it into account might hinder the effectiveness of the classifier. That said, in Sect. 4.4 we will nevertheless test the performance of the color version of WNRMSE on the ISA task.

In the following section, we present and comment the results for the three validation tasks.

4 Results and Comments

4.1 The Choice of the Wavelet for WNRMSE

The first validation task is the study of the performance of WNRMSE, when varying the choice of the mother wavelet. We consider each of the 25 original images as a query image, Q_i, $i \in [1, 25]$, and we study the ability of WNRMSE to evaluate the similarity among all the D_j, $j \in [1,3000]$ impaired images of the database. We consider 23 different wavelets [6], ϕ_k, $k \in [1, 23]$, namely (Table 1):

– the Daubechies (DB) wavelets, 1 through 10,
– the Symlets (Sym) wavelets, 2 through 8,
– the Coiflet (Coif) wavelets, 1 through 5,
– the Discrete Meyer (DMey) wavelet.

We compute

$$m(i, j, k) = \text{WNRMSE}(Q_i, D_j; \phi_k),$$
$$\forall i \in [1, 25]; \forall j \in [1, 3000]; \forall k \in [1, 23]$$

In Table 1, we report the AUC of the WNRMSE ROC curve for each wavelet. Each ROC curve has been determined by performing 75,000 comparisons (3000×25). A true positive is when the WNRMSE metric returns a *positive*, or *correct* identification of the impaired image, i.e., $m(i, j, k)$ is less than a parametric threshold T and D_j is one of the impaired versions of Q_i ($D_j \in C_{Q_i}$). A false positive is an under-threshold value of WNRMSE for the couple (Q_i, D_j) when D_j is not an impaired version of Q_i. According to the classical ROC analysis, by varying the threshold T we can plot the ROC curve (number of true positive vs. number of false positive), and we can compute the area under curve
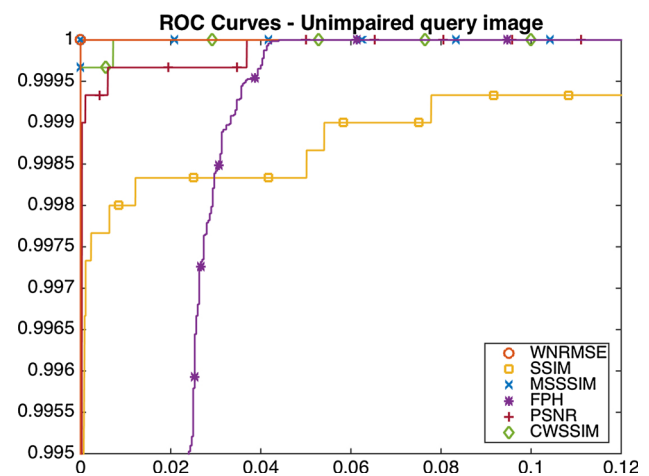
Table 1 WNRMSE: AUC for different wavelet types

Wavelet name	AUC	Wavelet name	AUC
DB1/Haar	1.00000000	Sym4	0.99998578
DB2	0.99999695	Sym5	0.99998921
DB3	0.99999570	Sym6	0.99998930
DB4	0.99999958	Sym7	0.99999861
DB5	0.99999668	Sym8	0.99999402
DB6	0.99999632	Coif1	0.99998519
DB7	0.99998580	Coif2	0.99999478
DB8	0.99999560	Coif3	0.99998657
DB9	0.99999707	Coif4	0.99999621
DB10	0.99999624	Coif5	0.99999014
Sym2	0.99999695	DMey	0.99999114
Sym3	0.99999570		

Fig. 2 ROC analysis on the TID2013 database for WNRMSE, SSIM, MSSSIM, FPH, PSNR and CWSSIM. Only the ROC for MSSSIM and WNRMSE passes through the point (0, 1) (perfect classification)

Table 2 AUC values—unimpaired query images

Metric	AUC	Metric	AUC
WNRMSE	1.00000000	FPH	0.98840902
SSIM	0.99984043	PSNR	0.99998361
MSSSIM	1.00000000	CWSSIM	0.99999759

Perfect classification is attained if and only if AUC = 1

(AUC) values. In the ideal case, the true positive rate is the 100% of the 75,000 tests, while the false positive rate is equal to zero. In this ideal case, the AUC value is equal to 1.

Table 1 shows that the WNRMSE performance is excellent: All the AUC values are very close to 1, with the notable case of the DB1 (Haar) wavelet, that actually achieves that value. This means that, on TID2013, WNRMSE is able to identify which images D_j are impaired copies of each Q_i with notable accuracy and quite independently of the wavelet ϕ_k. The experimental results prove in particular that, surprisingly, the wavelet of the lowest computational complexity, i.e., the Haar (or DB1) basis, is able to capture the structural similarity between any pair of images of the TID2013 database. Therefore, in the following of the experiment description, WNRMSE is always computed with $k = 1$.

4.2 Image Similarity for Unimpaired Query Images

The second set of experiments aims at comparing WNRMSE with the other previously selected metrics: SSIM, MSSSIM, CWSSIM, PSNR and the Feature Point Hash scheme (FPH). All tests are performed with Matlab© implementations for SSIM (of which we chose an optimized version with automatic downsampling),[2] MSSSIM,[3] CWSSIM,[4] FPH[5] and PSNR. The results are well described by analyzing the ROC curves (see Fig. 2) for each metric: WNRMSE returns a perfect classification, with the ideal 0–1 step curve. The same result holds also for MSSSIM, whose curve superposes exactly to the ideal plot of WNRMSE. The other metrics also perform quite well. In particular, both CWSSIM and PSNR

---

[2] http://www.cns.nyu.edu/~lcv/ssim/.

[3] https://ece.uwaterloo.ca/~70wang/research/iwssim.

[4] https://it.mathworks.com/matlabcentral/fileexchange.

[5] http://users.ece.utexas.edu/bevans/projects/hashing/software.html.

attain a rate of 100% TP against less than 5% FP. This good behavior is reflected by the AUC, as presented in Table 2, which are all very close to the ideal value 1.

These results are also confirmed by computing the histograms of the values of the distances expressed by each metrics, both for inter-class and for intra-class membership comparison (Fig. 3). The distributions allow to assess the degree of efficacy of the metric in evaluating the similarity of images of the same class, versus the efficacy of evaluating the dissimilarity among images of different classes. If the two distributions do not overlap, the metric is able to correctly classify—i.e., to assess similarity—the 100% of the images in the database. By referring to Fig. 3, we can see once again that the only two metrics that do not present overlapping histograms are WNRMSE and MSSSIM.

Remark that while SSIM, MSSSIM, PSNR and CWSSIM measure the similarity between two images, WNRMSE and FPH rather measure their dissimilarity. This explains why the intra-class and inter-class histogram (full curve, resp. dotted curve) are shifted toward, respectively, low and high values in the last two cases and the other way round in the former four. Remark also that, with the exception of PSNR, the similarity indexes are designed in such a way that they assume values in [0, 1], 1 meaning perfect similarity. No such constraint is imposed in the design of the two distances, so that the range of the corresponding histograms is larger. This does not pose
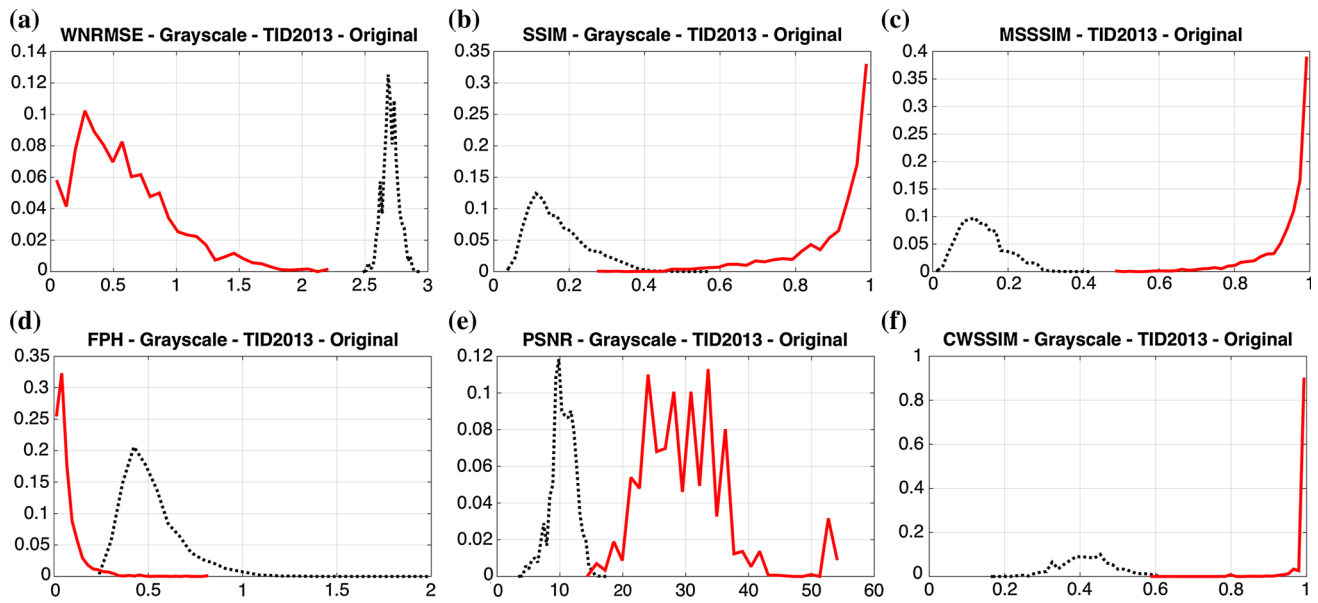
Fig. 3 Distribution of the values of each metric computed on the whole TID2013 database (*Y* axes), for intra-class (*solid line*) and inter-class (*dotted line*), for unimpaired (original) query images: WNRMSE (**a**), SSIM (**b**), MSSSIM (**c**), FPH (**d**), PSNR (**e**), CWSSIM (**f**). Remark that only for WNRMSE and CWSSIM the supports of the two curves do not overlap, and so a threshold exists allowing perfect classification

a problem for image classification since the values returned by different metrics/indexes are not directly compared with each other. The comparison between different metrics rather depends on the existence and on the width of a range of thresholds (or on the lack thereof) separating the inter-class and the intra-class values of the same metric computed on all the images, thus allowing perfect classification of an image to be carried out by simply comparing its distance from the query image with the threshold.

4.3 Image Similarity for Worst Quality Query Images

The third validation task aims at comparing the ability of the different metrics to replicate a peculiar characteristic of the HVS in similarity assessment. In fact, the human visual system is able to find similarity among images even if the query image is corrupted by a consistent distortion, probably because the similarity assessment implies also some high-level semantic recognition task. For example, if we consider the worst distorted images of Fig. 4, (*worst* according to the subjective MOS analysis), the human observer has no difficulty in deciding whether two images belong to the same class C_{Q_i} by comparing them directly, without resorting to intermediate comparisons with the *original, unimpaired* image Q_i. However, the same capacity is not assured in a metric which can capture only structural and local content in the visual stimuli. We exploit the peculiarity of TID2013, which stores the MOS for all the impaired images D_j. We can then simulate the most difficult task for the metric, by choosing a different set of query images, $\bar{Q}_i$. The 25 elements of this
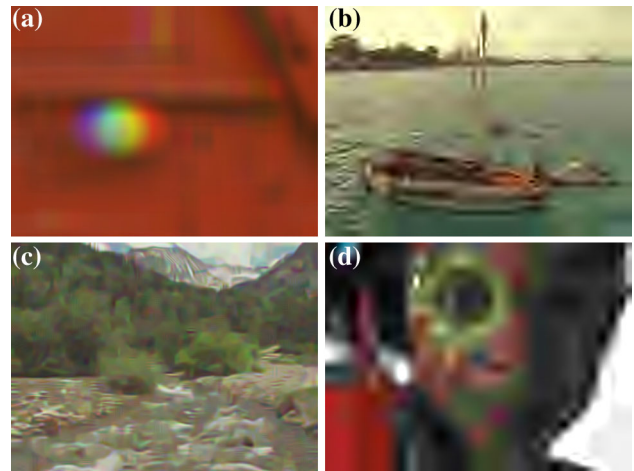

Fig. 4 Four samples of worst quality images in the TID2013 database

new set are the worst quality images in each class: for all $i \in [1, 25]$

$$\bar{Q}_i = D_j \in C_{Q_i} \Leftrightarrow \mathrm{MOS}_j = \min(\mathrm{MOS}_{C_{Q_i}}).$$

The performance for all the metrics is reported in Fig. 5, which shows the histograms for intra-class and inter-class analysis. In this more challenging task, the only metric that does not show overlapping between the two distributions is the WNRMSE metric, while also the MSSSIM, that has an optimal behavior in the second validation test, now presents a non-negligible range of overlapping. To further demonstrate this behavior, in Table 3 we report, for each metric,
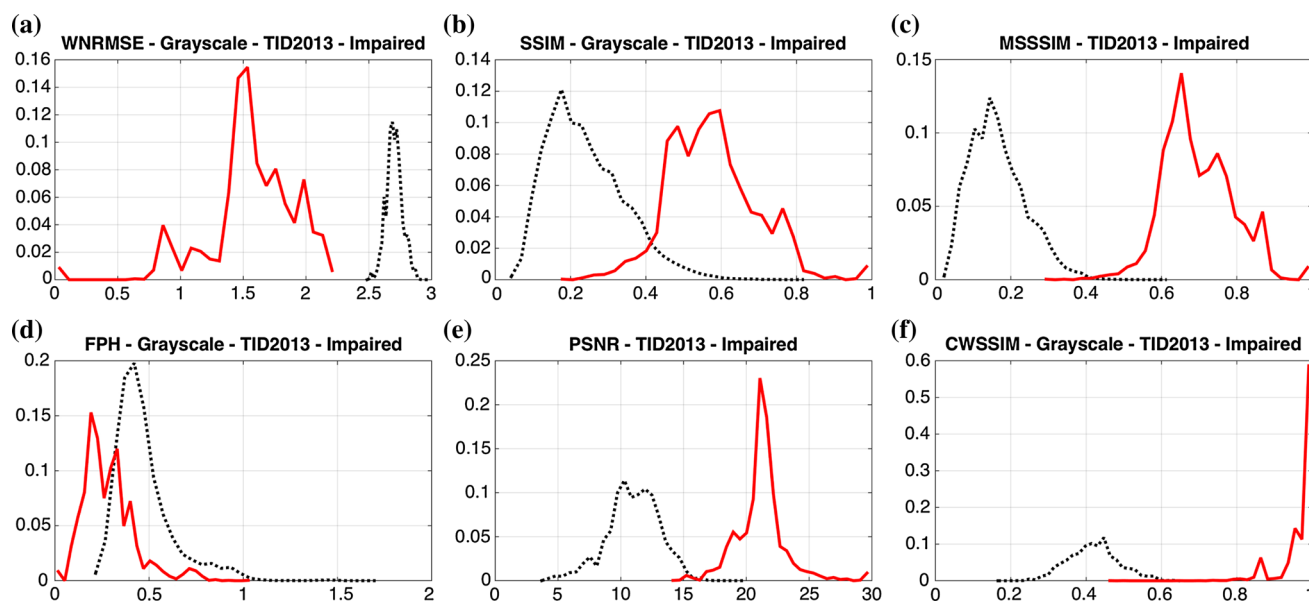
Fig. 5 Distribution of the values of each metric computed on the whole TID2013 (*Y* axes) for intra-class (*solid line*) and inter-class (*dotted line*), for the case of worst quality query images: on the *X* axes the values of WNRMSE (**a**), SSIM (**b**), MSSSIM (**c**), FPH (**d**), PSNR (**e**), CWSSIM (**f**). See Table 3 for the analytic values. Remark that a threshold allowing for perfect classification exists only for WNRMSE, the only metric for which the supports of the two curves do not overlap

Table 3 Range of intra-class and inter-class distribution for the "worst quality query images" experiment

Metrics	Inter-Class	Intra-Class	Overlap width
WNRMSE	[2.478,2.981]	[0,2.249]	0
SSIM	[0.029, 0.829]	[0.164, 1]	0.665
MSSSIM	[0.006, 0.620]	[0.282, 1]	0.338
FPH	[0.190 1.728]	[0. 1.044]	0.857
PSNR	[3.433, 19.890]	[13.893, ∞]	5.997
CWSSIM	[0.158, 0.654]	[0.454, 1]	0.200

the inferior and superior extrema of support of the inter-class and intra-class distributions, and the width of the overlapping interval. In the case of MSSSIM, for example, the values for inter-class comparisons are between 0.006 and 0.620, while the values for intra-class comparisons span from 0.282 up to 1. This means that metric values in the interval [0.282, 0.620] cannot directly identify the membership to a class represented by its worst quality query image, as the same value may correspond both to an inter-class and to an intra-class membership. Analogous considerations also hold for the other metrics in our analysis. This result is confirmed in Fig. 6, which shows the ROC curves for the six metrics considered, and by Table 4, that reports the corresponding AUC values.

Interestingly, Fig. 5 shows that CWSSIM, rather than MSSSIM, gives the best performance among the state-of-the-art metrics we tested, showing the smallest overlap between the inter-class and intra-class values. It is worth remarking
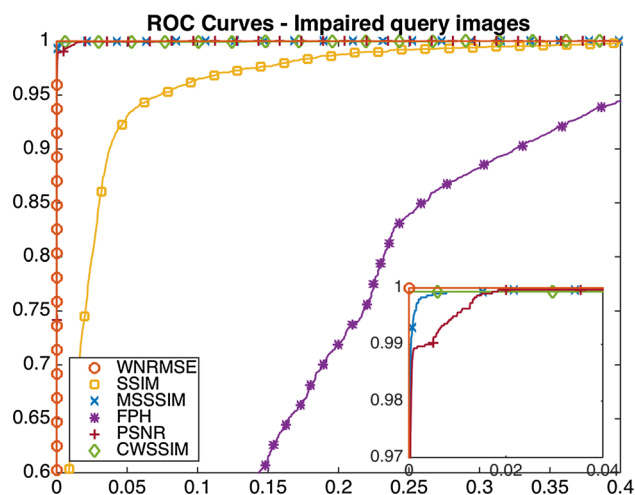

Fig. 6 ROC analysis for the six metrics when the query images set is $\bar{Q}_i$. In the *box* in the *low right corner*, a zoom in on (0,1), showing that CWSSIM and MSSSIM perform very well and confirm that WNRSMSE alone passes through the point (0, 1)

Table 4 AUC values—worst quality query images

Metric	AUC	Metric	AUC
WNRMSE	1.00000000	FPH	0.84540606
SSIM	0.98121397	PSNR	0.99985990
MSSSIM	0.99993921	CWSSIM	0.99986043

Perfect classification is attained if and only if AUC = 1

(a)

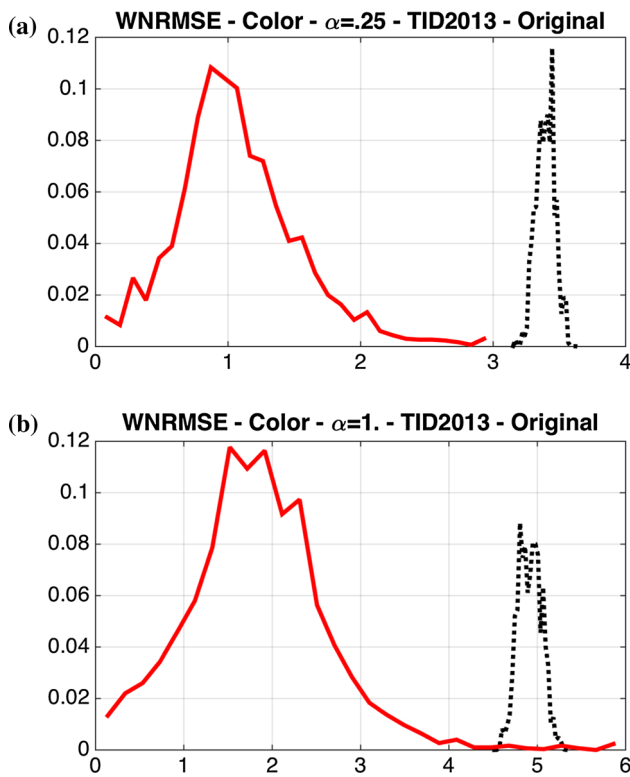**(a)**

(b)

(b)

Fig. 7 Distribution of the values of the *color* version of WNRMSE computed on the whole TID2013 (Y axes), with unimpaired query image for $\alpha_Y = 1$, $\alpha_I = \alpha_Q = .25$ (**a**) and for $\alpha_Y = \alpha_I = \alpha_Q = 1$ (**b**). For $\alpha = .25$ the intra-class (*solid*) curve and the inter-class (*dotted*) curve do not overlap (perfect classification)

Fig. 8 Distribution of the values of the *color* version of WNRMSE computed on the whole TID2013 (Y axes), with worst query image for $\alpha_Y = 1$, $\alpha_I = \alpha_Q = .25$ (**a**) and for $\alpha_Y = \alpha_I = \alpha_Q = 1$ (**b**). For $\alpha = .25$ the intra-class (*solid*) curve and the Inter-Class (*dotted*) curve do not overlap (perfect classification)

that, in this experiment, the three metrics involving some kind of multilevel decomposition, namely MSSSIM, CWS-SIM and WNRMSE, outperform the other metrics, all but WNRMSE, however, with suboptimal performance.

4.4 Image Similarity for Color Images

While most image similarity assessment tasks can be satisfactorily carried out using only the luminance information, for the sake of completeness we repeat both set of tests using the color version of WNRMSE, as proposed in Sect. 2.8. As for the grayscale case, we set $q = 2$ and $\omega_j = 1$ for all j. As far as the weights α_Y, α_I and α_Q are concerned we set $\alpha_Y = 1$ and $\alpha_I = \alpha_Q = \alpha$. We test two choices of the parameter α, namely

- the trivial choice $\alpha = 1$;
- $\alpha = 0.25$, which roughly corresponds to giving the color information half of the importance we give to the luminance.

In Fig. 7 (resp. Fig. 8), we display the histograms of the intra-class and inter-class distribution, when the query image is the
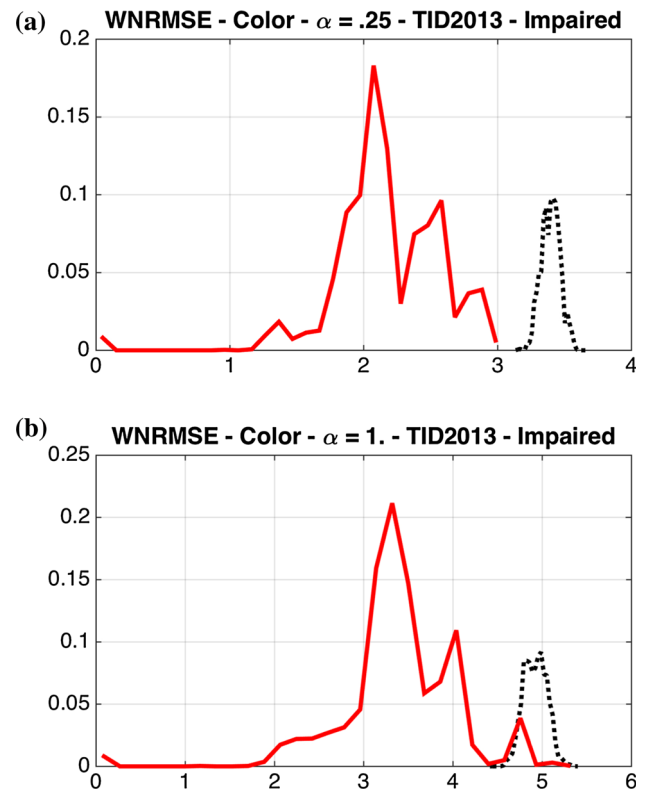
original unimpaired version (resp. the worst impaired version), of the reference image. Observe that the trivial choice $\alpha_Y = \alpha_I = \alpha_Q = 1$ gives too much weight to the color information, and this results in incorrect classification, which occurs when the distortion heavily involves the color. For the second choice, our metric reaches perfect classification with correct separation of inter- and intra-class distribution also in the challenging case of worst query image.

5 Conclusion

We presented a new class of image similarity metrics, namely the *Wavelet Normalized RMSE*, with the aim to replicate the HVS ability to identify and classify two images as copies, independently of their level of structural impairment.

We proved that WNRMSE is a metric, and we described some of its properties, such as Fréchet differentiability and local convexity. To validate the new metric, we performed several experimental activities on a standard database of human-evaluated images, TID2013. WNRMSE has been compared to different metrics in the literature: SSIM, Multiscale SSIM, PSNR, FPH and CWSSIM. The results of our

experimental activity proved that, when the query image is heavily impaired and therefore more challenging for the algorithms, WNRMSE outperforms all the other metrics, in terms of effectiveness (higher degree of separation between intra- and inter-class distribution).

Future development of this study will include the evaluation of the performance of WNRMSE against geometrical distortions, a further study of the influence of the choice of the wavelet on the performance of the metric— and, possibly its generalization to a metric based on non-tensor wavelets, such as shearlets and curvelets—as well as the study of a more refined metric to deal with color images.

References

1. Amir, D.: Characterizations of Inner Product Spaces. Birkhauser Verlag, Basel (1986)
2. Apostol, T.: Ptolemy inequality and the chordal metric. Math. Mag. **40**, 233–235 (1967)
3. Brunet, D., Vrscay, E.R., Wang, Z.: Structural similarity-based approximation of signals and images using orthogonal bases. In: Kamel, M., Campilho, A. (eds.) Image Analysis and Recognition, pp. 11–22. Springer, Berlin (2010)
4. Brunet, D., Vrscay, E.R., Wang, Z.: On the mathematical properties of the structural similarity index. IEEE Trans. Image Process. **21**(4), 1488–1499 (2012). doi:10.1109/TIP.2011.2173206
5. Chang, H., Zhang, J.: New metrics for clutter affecting human target acquisition. IEEE Trans. Aerosp. Electron. Syst. **1**(42), 361–368 (2006). doi:10.1109/TAES.2006.1603429
6. Daubechies, I., et al.: Ten Lectures on Wavelets, vol. 61. SIAM, Philadelphia (1992)
7. Fawcett, T.: An introduction to ROC analysis. Pattern Recognit. Lett. **27**(8), 861–874 (2006)
8. Hästö, P.A.: A new weighted metric: the relative metric I. Math. Anal. Appl. **274**, 38–58 (2002)
9. Hendee, W.R., Wells, P.N. (eds.): The Perception of Visual Information. Springer, Berlin (1997)
10. Horé, A., Ziou, D.: Image quality metrics: PSNR vs. SSIM. In: Proceedings of IEEE International Conference on Pattern Recognition, pp. 2366–2369. IEEE, Istanbul, Turkey (2010)
11. Hsu, C.Y., Lu, C.S.: Geometric distortion-resilient image hashing system and its application scalability. In: Proceedings of the 2004 Workshop on Multimedia and Security, pp. 81–92. ACM, New York, NY, USA (2004). doi:10.1145/1022431.1022448
12. Hubel, D.H., Wiesel, T.N.: Receptive fields and functional architecture in two non striate visual areas of the cat. J. Neurophysiol. **28**, 229–289 (1965)
13. Leach, M.: A complete distortion correction for MR images: Ii. rectification of static-field inhomogeneities by similarity-based profile mapping. PHYSICS IN MEDICINE AND BIOLOGY **50**(11), 2651–2661 (2005). http://publications.icr.ac.uk/2089/
14. Lissner, I., Preiss, J., Urban, P., Lichtenauer, M.S., Zolliker, P.: Image-difference prediction: from grayscale to color. IEEE Trans. Image Pocess. **22**(2), 435–446 (2013)
15. Mallat, S.: A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way, 3rd edn. Academic Press, Cambridge (2008)
16. Monga, V., Evans, B.L.: Perceptual image hashing via feature points: performance evaluation and tradeoffs. IEEE Trans. Image Pocess. **15**(11), 3452–3465 (2006)
17. Nikvand, N., Wang, Z.: Image distortion analysis based on normalized perceptual information distance. Signal Image Video Process. **7**(3), 403–410 (2013)
18. Okarma, K.: Colour image quality assessment using structural similarity index and singular value decomposition. In: Computer Vision and Graphics, pp. 55–65. Springer (2009)
19. Okarma, K.: Combined image similarity index. Opt. Rev. **19**(5), 349–354 (2012)
20. Pezoa, J., Torres, S., Cdova, J., Reeves, R.: An enhancement to the constant range method for nonuniformity correction of infrared image sequences. In: A. Sanfeliu, J. Martnez Trinidad, J. Carrasco Ochoa (eds.) Progress in Pattern Recognition, Image Analysis and Applications, Lecture Notes in Computer Science, vol. 3287, pp. 525–532. Springer Berlin Heidelberg (2004)
21. Piella, G., Heijmans, H.: A new quality metric for image fusion. In: 2003 IEEE International Conference on Image Processing, vol. 2, pp. 173–176. IEEE (2003). doi:10.1109/ICIP.2003.1247209
22. Ponomarenko, N., Battisti, F., Egiazarian, K., Astola, J., Lukin, V.: Metrics performance comparison for color image database. In: Fourth international workshop on video processing and quality metrics for consumer electronics, vol. 27 (2009)
23. Ponomarenko, N., Ieremeiev, O., Lukin, V., Egiazarian, K., Jin, L., Astola, J., Vozel, B., Chehdi, K., Carli, M., Battisti, F., Jay Kuo, C.C.: Color image database TID2013: peculiarities and preliminary results. In: Proceedings of the 4th European Workshop on Visual Information Processing EUPIV2013
24. Ponomarenko, N., Jin, L., Ieremeiev, O., Lukin, V., Egiazarian, K., Astola, J., Vozel, B., Chehdi, K., Carli, M., Battisti, F., Jay Kuo, C.C.: Image database TID2013: peculiarities, results and perspectives. Signal Process. Image Commun. **30**, 57–77 (2015)
25. Ponomarenko, N., Lukin, V., Zelensky, A., Egiazarian, K., Carli, M., Battisti, F.: TID2008-a database for evaluation of full-reference visual quality assessment metrics. Adv. Mod. Radioelectron. **10**(4), 30–45 (2009)
26. Sampat, M.P., Wang, Z., Gupta, S., Bovik, A.C., Markey, M.K.: Complex wavelet structural similarity: a new image similarity index. IEEE Trans. Image Pocess. **18**(11), 2385–2401 (2009)
27. Simoncelli, E.P., Freeman, W.T., Adelson, E.H., Heeger, D.J.: Shiftable multiscale transforms. IEEE Trans. Inform. Theory **38**(2), 587–607 (1992)
28. Snidaro, L., Foresti, G.L.: A multi-camera approach to sensor evaluation in video surveillance. In: 2005 IEEE International Conference on Image Processing, pp. 1101–1104. IEEE (2005)
29. Solomon, D.: A Guide to Data Compression Methods. Springer, Berlin (2002)
30. Torres, L.H.: Modern Sampling Theory. Birkhäuser, Basel (2001)
31. Wang, Z., Bovik, A.C.: Mean squared error: Love it or leave it? A new look at signal fidelity measures. Signal Process. Mag. **26**(1), 98–117 (2009)
32. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Pocess. **13**(4), 600–612 (2004). doi:10.1109/tip.2003.819861
33. Wang, Z., Bovik, A.C., Simoncelli, E.P.: Structural approaches to image quality assessment. In: Bovik, A. (ed.) Handbook of Image and Video Processing, chap. 8.3, 2nd edn, pp. 961–974. Academic Press, Cambridge (2005)
34. Wang, Z., Shang, X.: Spatial pooling strategies for perceptual image quality assessment. In: 2006 IEEE International Conference on Image Processing, pp. 2945–2948. IEEE (2006). doi:10.1109/ICIP.2006.313136
35. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multiscale structural similarity for image quality assessment. In: Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers, 2004, vol. 2, pp. 1398–1402. IEEE (2003)

36. Zujovic, J., Pappas, T.N., Neuhoff, D.L.: Structural texture similarity metrics for image analysis and retrieval. IEEE Trans. Image Pocess. **22**(7), 2545–2558 (2013)

Maria Grazia Albanesi after the Master Degree cum laude in Electronic Engineering in 1986, got the Ph.D. in Electronic Engineering and Computer Science at the University of Pavia, Italy. Actually, she is Associate Professor at the same University. She works in the field of multimedia and image processing applications. The main fields are image quality assessment and visual data analysis, applied to sustainability and environment preservation. Since 2014, she coordinates an interdisciplinary research unit in the field of the Computational Sustainability, for the application of image processing techniques to biodiversity estimation.

Riccardo Amadeo obtained a Master Degree in Computer Engineering and a Ph.D. in Electronic Engineering and Computer Science at the University of Pavia, Italy. He pursues his research interest developing and testing algorithms for image processing and Multimedia quality evaluation at the Digital Content Retrieval Lab at the University.

Silvia Bertoluzza graduated in mathematics at the University of Pavia, Italy. Since 1988, she holds a position at the CNR Institute for Applied Mathematics and Information Technologies in Pavia, where she is currently a Research Director. Her research activity is mainly focused on the study of numerical methods for the solution of partial differential equations. At the same time, she carries out research on wavelet methods for image processing. She has been coordinator of the two EC Projects "Wavelets for Numerical Simulation" (1998–2002) and "Breaking Complexity" (2002–2006), which gathered researchers from France, Germany, Great Britain, Portugal, Spain, Switzerland and the USA. She is the author of more than 60 scientific papers.

Giulia Maggi was born in Pavia, Italy, on December 29, 1983. She received her M.S. (summa cum laude) and Ph.D. degrees in 2008 and 2012, respectively, from University of Pavia, Italy. Her research interests included the application of image processing techniques to biomedical problems. She is currently a Data Manager from IRCCS Humanitas Research Hospital, Milan, which she joined in 2013.