Divisive Normalization Image Quality Metric Revisited

Valero Laparra, Jordi Muñoz-Marí and Jesús Malo[1]

Image Processing Laboratory, Universitat de València.
Catedrático A. Escardino, 46980 Paterna, València, (Spain).
Valero.Laparra@uv.es, Jordi@uv.es, Jesus.Malo@uv.es
http://www.uv.es/vista/vistavalencia

Structural similarity metrics and *information theory* based metrics have been proposed as a completely different alternative to the traditional metrics based on *error visibility* and human vision models. Three basic criticisms were raised against the traditional error visibility approach: (1) it is based on near threshold performance, (2) its geometric meaning may be limited, and (3) stationary pooling strategies may not be statistically justified. These criticisms and the good performance of structural and information theory based metrics have popularized the idea of their superiority over the error visibility approach.

In this work we experimentally or analytically show that the above criticisms do not apply to error visibility metrics that use a general enough Divisive Normalization masking model. According to this, the traditional Divisive Normalization metric [1] is not intrinsically inferior to the newer approaches. In fact, experiments on a number of databases including a wide range of distortions show that Divisive Normalization is fairly competitive with the newer approaches, robust, and easy to interpret in linear terms.

These results suggest that, despite the criticisms to the traditional error visibility approach, Divisive Normalization masking models should be considered in the image quality discussion.

1. Introduction

Reproducing subjective opinion of image distortion has two broad applications: in *engineering*, image quality metrics may replace the (time consuming) human evaluation to assess the

[1]This work was partially supported by projects CICYT-FEDER TEC2009-13696, AYA2008-05965-C04-03, and CSD2007-00018. Valero Laparra acknowledges the support of the Ph.D grant BES-2007-16125

results of the algorithms, and in *vision science*, image quality results may provide insight on the way the brain processes visual information.

Nowadays there is a fruitful debate about the right approach in the image quality assessment problem. The image quality metrics have been classified according to this broad taxonomy [2]: (1) error visibility techniques based on human visual system (HVS) models, (2) structural similarity techniques, and (3) information theoretic techniques.

The classical *error visibility* approach to simulate human judgement naturally tried to include empirical aspects of the HVS in the image metric [3, 4]. Basic features taken into account include decomposition in orientation and scale channels [5], contrast sensitivity [3, 6–8], and contrast masking non-linearities through simple point-wise models [9–11], the more general Divisive Normalization [1, 12, 13], or equivalently, the non-uniform nature of just noticeable differences (JND) [14]. The final distance measure is typically obtained from a certain summation norm of the difference vector in the internal image representation domain (Minkowski pooling) [3, 15].

Recently, alternatives to the above empirical approach have been proposed: *structural similarity* methods [16–18] and *information theoretic* methods [19, 20]. The common ground of these new techniques rely on the relation between image statistics and the behavior of the visual system [21, 22]: since the organization and non-linearities of visual sensors seem to emerge from image statistics [23, 24], it is sensible to assess the image distortion by measuring the departure of the corrupted image from the average behavior of natural images.

The *structural similarity* approach quantifies visual quality by comparing three statistical measures in the original and distorted images: mean (related to luminance), variance (related to contrast), and cross-correlation (related to structure). The aim of using a characterization of the structure is achieving invariance under small changes in the image [25]. This general concept has been applied both in the spatial domain (SSIM) [17] and in multi-scale image representations (MSSIM) [16] and (CW-SSIM) [18].

The *information theoretic* approach (VIF [20]) quantifies the similarity by comparing the information that could ideally be extracted by the brain from the distorted and the original image respectively. The authors assume a certain image source model and characterize the HVS as a simple channel that introduces additive noise in the wavelet domain. The amount of information that can be extracted about the original signal from the perceived images is modeled by the mutual information between the output of the above simplified HVS model and the original image.

Despite some of the new approaches claim to be a new philosophy [17], a number of qualitative relations have been pointed out among the newer approaches and Divisive Normalization masking models [19, 20, 26]. However, no explicit comparison has been done with metrics based on updated versions of the Divisive Normalization error visibility. Moreover,

the new approaches criticize the classical error visibility approach in many ways:

- Suprathreshold problem. Since the empirical HVS models are based on near threshold measurements using too simple (academic) stimuli, it is argued that there is no guarantee that these models are applicable to suprathreshold distortions on complex natural images [17, 20].

- Geometric limitations of error visibility techniques. In [17], the authors criticize linear and point-wise non-linear HVS models because they give rise to too rigid discrimination regions, while stressing the flexibility of structural measures. Nevertheless, the authors (qualitatively) recognize that general Divisive Normalization models (including inter-coefficient masking) may induce a richer geometric behavior.

- Minkowski pooling assumes statistical independence among error coefficients. It has been argued that this is not an appropriate summation strategy in linear domains where there are statistical relations among coefficients [17]. This criticism is certainly appropriate for linear (CSF-based) HVS models. Again, this wouldn't be the case for image representations with reduced relations among coefficients.

These non-addressed criticisms, and the fact that the new approaches are easy to use in a number of engineering applications [2], have popularized the idea of their superiority over the error visibility approach.

The aim of this work is to provide new results in favor of the classical error visibility approach by showing that the above criticisms do not apply to the Divisive Normalization masking models, and by showing that what will be referred to as Divisive Normalization metric (originally proposed as image quality measure in [1]) can be easily adapted to be competitive with the new approaches. This is an additional evidence to confirm the link among the different strategies, and suggests that, despite the criticisms, Divisive Normalization masking models should still be considered in the image quality discussion.

The paper is organized as follows: In Section 2 we review and generalize the Divisive Normalization masking model, and we show that the resulting metric successfully addresses the criticisms against the error visibility approach. Finally we show the relation of the proposed metric to other error visibility techniques. In Section 3 we compare the performance of the proposed metric to structural similarity techniques (SSIM [17] and MSSIM [16]), and information theoretic techniques (VIF [20]). An extensive comparison is made according to standard procedures in a number of recently available subjectively rated databases including a total of 2173 distorted images and 25 kinds of distortion. Finally in Section 4 we draw the conclusions of the work and discuss additional issues that may improve the Divisive Normalization performance.

2. The Divisive Normalization metric

In this section we first review the Divisive Normalization model and the associated error visibility metric as implemented here. In subsection 2.A we describe the procedure to set the parameters of the model. Afterwards (in subsections 2.B to 2.D), we address the criticisms made against error visibility techniques: we show (1) the ability to simultaneously reproduce high level and low level distortion data, (2) the geometric richness of the model, and (3) the statistical independence effect that justifies uniform Minkowski pooling. Finally in subsection 2.E we show how the proposed metric relates to other error visibility metrics.

The Divisive Normalization metric originally proposed by Teo and Heeger [1] is based on the standard psychophysical and physiological model that describes the early visual processing up to the V1 cortex [15, 27–29]. In this model, the input image, $\mathbf{x} = (x_1, \cdots, x_n)$, is first analyzed by a set of wavelet-like linear sensors, $\mathbf{T}_{ij}$, that provide a scale and orientation decomposition of the image [15]. The linear sensors have a frequency dependent linear gain according to the Contrast Sensitivity Function (CSF), $\mathbf{S}_j$, [27, 28]. The weighted response of these sensors is non-linearly transformed according to the Divisive Normalization, $\mathbf{R}$ [15, 29], in which they are rectified and normalized by a pooling of the responses of the neighboring sensors in scale, orientation and spatial position:

$$\mathbf{x} \xrightarrow{\mathbf{T}} \mathbf{w} \xrightarrow{\mathbf{S}} \mathbf{w}' \xrightarrow{\mathbf{R}} \mathbf{r} \tag{1}$$

In this scheme, the rows of the matrix $\mathbf{T}$ contain the linear receptive fields of V1 neurons. In this work the V1 linear stage is simulated by an orthogonal 4-scales QMF wavelet transform [30]. $\mathbf{S}$ is a diagonal matrix containing the linear gains to model the contrast sensitivity. Here, the diagonal in $\mathbf{S}$, is described by a function that depends on the scale, $e = 1, 2, 3, 4$, (e ranges from fine to coarse), may depend on the orientation, $o = 1, 2, 3$, (the o values stand for horizontal, diagonal and vertical), but it is constant for every spatial position, $\mathbf{p}$:

$$S_i = S_{(e,o,\mathbf{p})} = A_o \cdot \exp\left(-\frac{(4-e)^\theta}{s_o^\theta}\right) \tag{2}$$

where A_o is the maximum gain for the considered orientation, s_o controls the bandwidth of the frequency response, and θ determines the sharpness of the decay with spatial frequency. Finally, $\mathbf{R}$ is the Divisive Normalization response:

$$\mathbf{R}(\mathbf{w}')_i = r_i = \text{sign}(w_i') \frac{|S_i \cdot w_i|^\gamma}{\beta_i^\gamma + \sum_{k=1}^n H_{ik} |S_k \cdot w_k|^\gamma} \tag{3}$$

where H is a kernel matrix that controls how the responses of neighboring linear sensors, k, affect the non-linear response of sensor i [15].

Even though in the original use of Divisive Normalization for image quality purposes [1] the interaction kernel weights every sensor in a certain neighborhood in the same way, here we

use the Gaussian interaction kernel proposed by Watson and Solomon [15], which has been successfully used in block-frequency domains [13, 31, 32], and in steerable wavelet domains [33]. In the orthogonal wavelet domain this reduces to:

$$H_{ik} = H_{(e,o,\mathbf{p}),(e',o',\mathbf{p}')} = K \cdot \exp\left(-\left(\frac{(e-e')^2}{\sigma_e^2} + \frac{(o-o')^2}{\sigma_o^2} + \frac{(\mathbf{p}-\mathbf{p}')^2}{\sigma_p^2}\right)\right) \tag{4}$$

where $(e, o, \mathbf{p})$ and $(e', o', \mathbf{p}')$ refer to the scale, orientation and spatial position meaning of the wavelet coefficients i and k respectively, and K is a normalization factor to ensure $\sum_k H_{ik} = 1$.

In our implementation of the model we set the profile of the regularizing constants β_i according to the standard deviation of each subband of the wavelet coefficients of natural images in the selected wavelet representation. This is consistent with the interpretation of the values β_i as priors of the amplitude of the coefficients [22]. This profile (computed from 100 images of a calibrated image data base [34]) is further multiplied by a constant b to be set in the optimization process.

Given an input image, $\mathbf{x}$, and its distorted version, $\mathbf{x}' = \mathbf{x} + \Delta\mathbf{x}$, the above model provides two response vectors, $\mathbf{r}$, and $\mathbf{r}' = \mathbf{r} + \Delta\mathbf{r}$. The perceived distortion can be obtained through the appropriate pooling of the one dimensional deviations in the vector $\Delta\mathbf{r}$. Non-quadratic pooling norms have been reported [3, 15, 35]. Moreover, different summation exponents, for the pooling across spatial position, q_p, and frequency, q_f, may be used:

$$d_{pf}(\mathbf{x}, \mathbf{x}') = \frac{1}{n}\left[\sum_{\mathbf{f}}\left[\left[\sum_{\mathbf{p}} \Delta r_{\mathbf{fp}}^{q_p}\right]^{\frac{1}{q_p}}\right]^{q_f}\right]^{\frac{1}{q_f}} \tag{5}$$

$$d_{fp}(\mathbf{x}, \mathbf{x}') = \frac{1}{n}\left[\sum_{\mathbf{p}}\left[\left[\sum_{\mathbf{f}} \Delta r_{\mathbf{fp}}^{q_f}\right]^{\frac{1}{q_f}}\right]^{q_p}\right]^{\frac{1}{q_p}} \tag{6}$$

where $\mathbf{f} \equiv \{e, o\}$. In this general case, the order in which dimensions are pooled matters. Pooling across space and frequency is not commutative unless both pooling exponents are the same. In particular, Teo and Heeger proposed to compute the perceived distortion as the Euclidean norm of the difference vector (quadratic Minkowski pooling exponent $q_p = q_f = 2$).

The color version of the V1 response model involves the same kind of spatial transforms described above applied on the image channels in an opponent color space [36]. Here we use the standard YUV (luminance, yellow-blue, red-green) representation [37]. According to the well known differences in frequency sensitivity in the achromatic and chromatic channels [38], we will allow for different matrices $\mathbf{S}$ in the YUV channels. In particular, we will allow for different gains (A_{oY}, $A_{oU} = A_{oV}$) and different bandwidths (s_{oY}, $s_{oU} = s_{oV}$). We will assume the same behavior for the other spatial transforms since the non-linear behavior of the chromatic channels is similar to the achromatic non-linearities [36].

2.A. Setting the model parameters

In the original work introducing the metric based on Divisive Normalization [1] and in the sequels [12, 13] the parameters were inspired in psychophysical facts. In general there are three basic strategies to obtain the parameters of the model:

- The *direct empirical approach* implies fitting the parameters to reproduce direct low-level perception data such as physiological recordings on V1 neurons (as in [29]), or psychophysical measurements of contrast incremental thresholds (as in [15]). Since the realization of direct experiments is beyond the scope of this paper, this low-level empirical approach is not straightforward because the physiological and psychophysical literature is often interested in a subset of the parameters, and a variety of experimental settings is used in these restricted experiments (e.g. different selected stimuli, different contrast units...). As a result, it is not easy to unify the wide range of experimental results into a common computational framework.

- The *indirect empirical approach* implies fitting the parameters of the model to reproduce higher level visual tasks such as image quality assessment: for instance, in [35] the authors fitted the parameters of the Standard Spatial Observer to the VQEG subjectively rated data.

- The *statistically-based approach* assumes that the goal of the different signal transforms is to increase the independence among the coefficients of the image representation [21, 23, 24]. In this case, the parameters of the model may be optimized to maximize some statistical independence measure as in [22].

In this work we take the second approach: we fitted the parameters of the Divisive Normalization metric to maximize the Pearson correlation with the subjective ratings of a subset of the LIVE Quality Assessment Database [39]. In order to point out the generalization ability of the proposed metric, we optimized the Divisive Normalization model just for 3 of the 27 images in the database (*house, sailing2* and *womanhat*) that represents about 10% of the available data. In Section 3 we not only test the behavior of the model in the whole dataset but also in other databases not including LIVE distortions (TID [40], IVC [41], and Cornell [14]). By using this testing strategy, we address one of the criticisms to the error visibility techniques: the model is applicable to a variety of new supra-threshold distortions, while still reproducing the low-level psychophysical results (as will be shown in Section 2.B).

Assuming the same behavior in the horizontal and vertical directions ($o = 1, 3$), and assuming that the oblique effect in the frequency sensitivity [42] is described by a single attenuation of the gain in the diagonal direction (i.e. $A_2 = d \cdot A_1$ in every chromatic channel),

the model described so far has 13 free parameters:

$$\Omega \equiv \{ A_{1Y}, d, A_{1UV}, s_Y, s_{UV}, \theta, \gamma, b, \sigma_e, \sigma_o, \sigma_{\mathbf{p}}, q_s, q_f \}. \tag{7}$$

In order to simplify the optimization process, we didn't explore all the dimensions of the parameter space at the same time, but optimized the parameters using a three stages procedure obtaining local optima in restricted subspaces. We first obtained the basic parameters of the model by neglecting the chromatic channels, the oblique effect and the non-quadratic summation, i.e. using $A_{1UV} = 0$, $d = 1$, and $q_s = q_f = 2$, thus reducing the dimensions of the parameter space to 8, $\Omega_1 \equiv \{ A_Y, s_Y, \theta, \gamma, b, \sigma_e, \sigma_o, \sigma_{\mathbf{p}} \}$. Afterwards, we checked the eventual improvements obtained from the previous (local) optimal configuration by considering the chromatic channels and allowing different values for the sensitivity in the diagonal direction, $\Omega_2 \equiv \{ A_{UV}, s_{UV}, d \}$. Finally, different summation exponents for the spatial and frequency pooling (in both possible orders) were considered $\Omega_3 \equiv \{ q_s, q_f \}$.

The only computational inconvenience of the proposed metric is the size of the kernel H. In order to circumvent this problem, two approximations were necessary:

- Kernel thresholding and quantization. The Gaussian interaction matrices were converted to sparse matrices by eliminating those elements below a given threshold, that in our experiments was set to 1/500 of the maximum in each interaction neighborhood. Once the best Gaussian kernel was obtained, their size was further reduced by quantizing it using 6 bits. No appreciable reduction of the performance was introduced by this quantization, while extremely reducing the storage requirements.

- Limitation of the image size. The LIVE database include images of size 512×768. This size implies a huge kernel. Since the computation and storage of a number of non-quantized kernels is necessary for the optimization process, we decided to restrict ourselves to work with cropped versions of the images in the database. The cropped versions of the images were obtained by selecting the 256×256 area around the most salient point of each (original) image for 10 observers. The most salient point was estimated as the average of the points selected by the observers. This approximation is relevant just in the optimization process. Actually, the resulting Divisive Normalization is used for images of any size by applying it first to each 256×256 block of the image and then by merging the result of each block into a single pyramid.

The parameter ranges were set starting from an initial guess obtained from the low-level psychophysical behavior [15] and previous use of similar models in image processing applications [12, 13, 31, 32]. The explored ranges for the parameters and the optimal values found are shown in Table 1. The optimal pooling strategy found in our experiment was Eq. 6: first sum over subbands and then over spatial positions. Figure 1 shows the shape of the linear

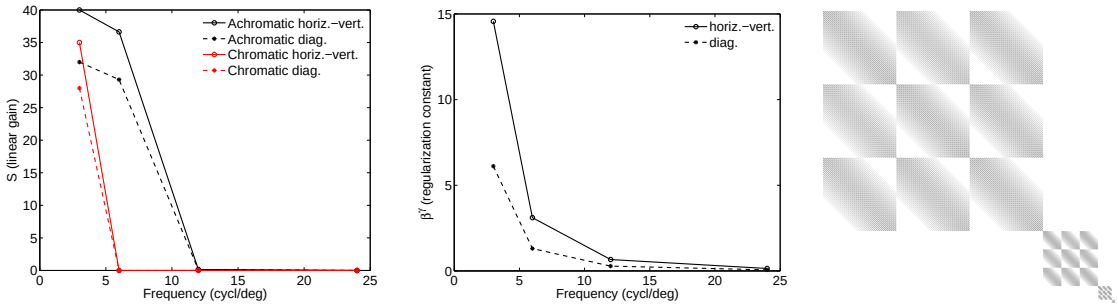

Fig. 1. Linear gains S (left), regularization constants β^γ (center), and interaction kernel H (right).

gains $\mathbf{S}$, the regularization constants β^γ and the interaction kernel H when using the optimal parameters. The structure of the interaction kernel comes from the particular arrangement of wavelet coefficients used in the transform [30].

Parameter	Range	Optimal	Correlation
A_Y	$30, \ldots, 60$	**40**	
s_Y	$0.25, \ldots, 3$	**1.5**	
θ	$2, \ldots, 8$	**6**	
γ	$0.5, \ldots, 3$	**1.7**	
b	$0.5, \ldots, 8$	**2**	
σ_e	$0.15, \ldots, 3$	**0.25**	
σ_o	$0.15, \ldots, 3$	**3**	
$\sigma_{\mathbf{p}}$	$0.03, \ldots, 0.4$	**0.25** (in deg)	$\rho_p = 0.916$
A_{UV}	$30, \ldots, 40$	**35**	
s_{UV}	$0.25, \ldots, 1.5$	**0.5**	
d	$0.6, \ldots, 1.4$	**0.8**	$\rho_p = 0.922$
q_p	$0.5, \ldots, 6$	**2.2**	
q_f	$0.5, \ldots, 6$	**4.5**	$\rho_p = 0.931$

Table 1. Parameter space, optimal values found, and improvement of the Pearson correlation in the progressive stages of the optimization.

2.B. Consistency with low level psychophysical data

In this section we show that the model optimized to reproduce (high-level) image quality results also reproduces the basic (low-level) trends on frequency sensitivity and contrast masking.

Here, the response of the model to a given incremental pattern (target), $\Delta\mathbf{x}$, seen on top of a background, $\mathbf{x}$, is computed as the perceptual distance $d(\mathbf{x}, \mathbf{x} + \Delta\mathbf{x})$.

The contrast sensitivity can be simulated by computing the above distances between sinusoids with fixed contrast, but different frequencies and orientations, and a uniform gray background. Figure 2 compares the result of this simulation for achromatic sinusoids in a wide range of spatial frequencies with the corresponding achromatic CSF of the Standard Spatial Observer [42]. Note that the model approximately reproduces the band pass behavior and the oblique effect.

In order to simulate contrast masking results the contrast of a Gabor patch is increased on top of different backgrounds (sinusoids with different contrasts and orientations). Figures 3 and 4 show examples of this kind of stimuli when test and background have the same and different orientation. Note that the visibility of the target increases quickly for low contrast targets, while remains more stable for higher contrast targets, thus revealing a non-linear response. Moreover, the visibility of the target is reduced as the contrast of the background is increased. This effect is bigger in figure 3 than in 4 because the background has the same orientation as the target.

Figure 5 shows the responses of the model to the targets for the different background sets.

The model response to the target is a saturating non-linearity when the target is shown on top of no background (auto-masking). The model predicts the reduction of the response

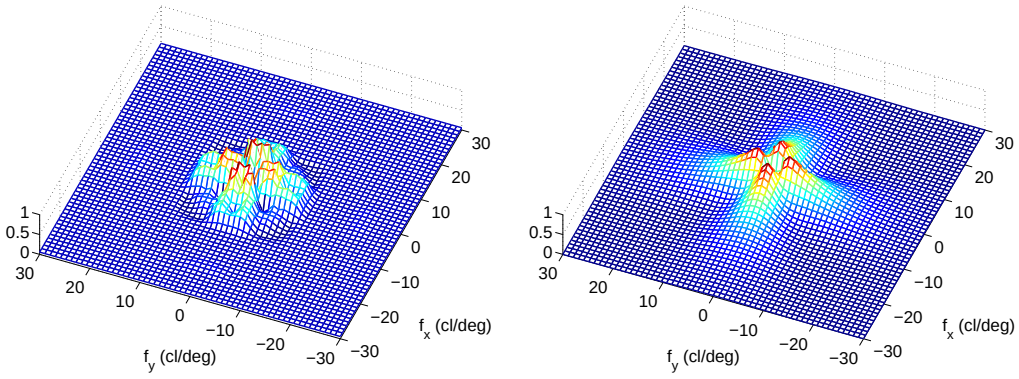

Fig. 2. Frequency sensitivity prediction for achromatic sinusoids (left) and the corresponding CSF of the Standard Spatial Observer (right).

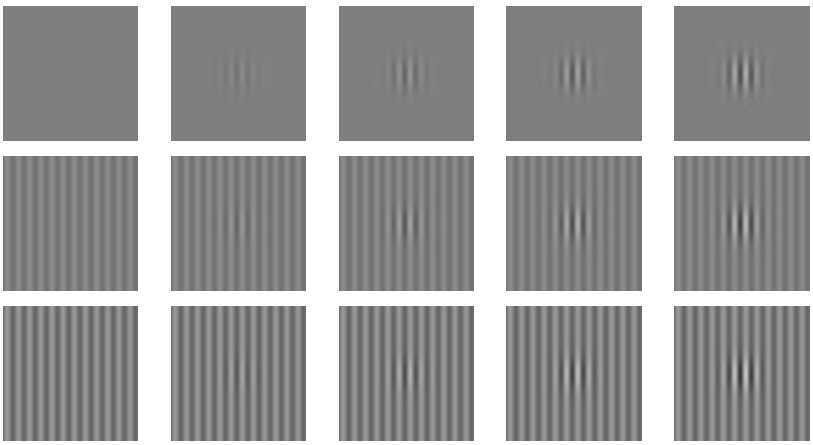

Fig. 3. Gabor patch of 6 cycl/deg (target) on top of sinusoids of the same frequency and orientation (background). In each row the contrast of the Gabor patch is increased from 0 to 0.6. The contrast of the background is 0 (top row) 0.1 (middle row) and 0.2 (bottom row).

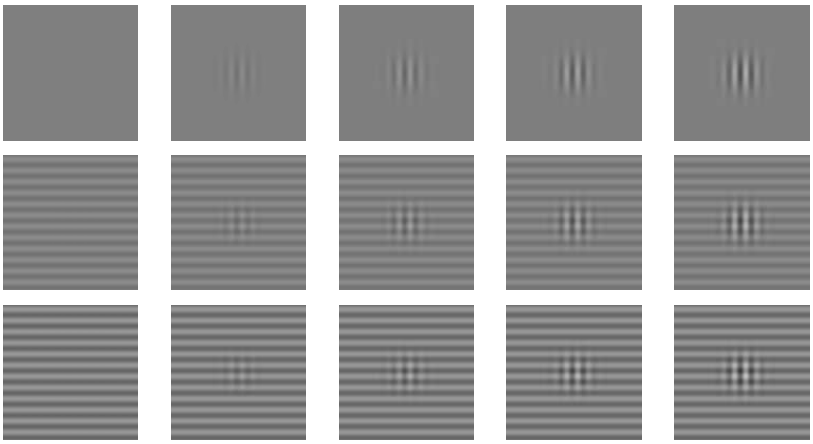

Fig. 4. Gabor patch of 6 cycl/deg (target) on top of sinusoids of the same frequency and different orientation (background). The visibility of the target on top of non-zero backgrounds is reduced but not as much as in the previous example.

when the target is shown on top of a background (cross-masking). The reduction increases with the contrast of the mask. Moreover, note that the reduction in visibility is bigger for backgrounds of the same nature. Therefore, the behavior of the model with the proposed

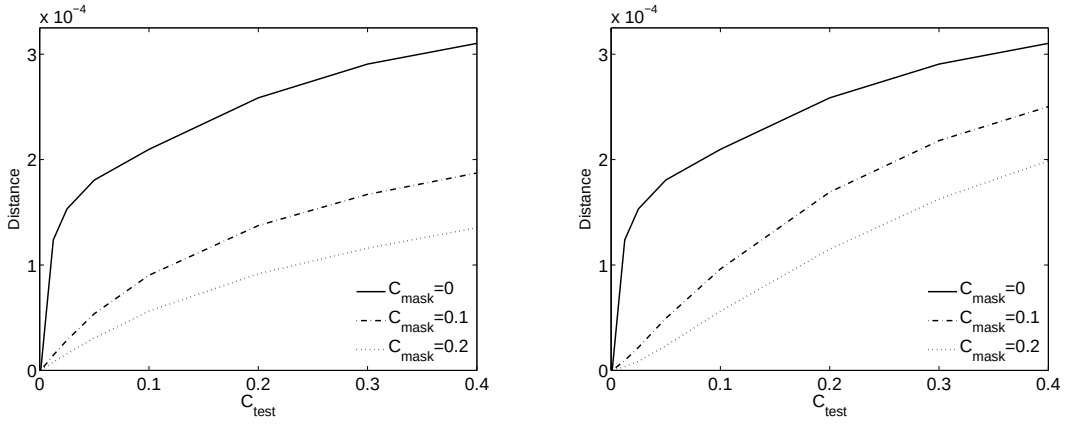

Fig. 5. Response predictions for masks of the same orientation (left) and orthogonal orientation (right). Different lines represent different background (mask) contrast. The three lines of each plot correspond to the visibility of the three rows in figures 3 and 4.

parameters is compatible with the low-level behavior of human observers reported in [15].

The results in this section show that the Divisive Normalization model optimized for a restricted set of high level distortions (such as those in the LIVE database) can reproduce the basic features of low-level psychophysics, while dealing with different suprathreshold distortions (as shown in Section 3). This shows that the suprathreshold criticism does not apply to the Divisive Normalization metric.

2.C. Geometry of the Divisive Normalized domain

Assuming a quadratic pooling in the distance computation, a number of analytical results can be obtained that show the appealing geometric behavior of the proposed metric. This behavior still holds for non quadratic schemes.

In the quadratic summation case, the Euclidean metric, I, in the Divisive Normalization domain may be interpreted as using non-Euclidean (Riemannian) metrics, M, in other image representation domains [12, 13]. The metric matrix, M, is a quadratic form that determines the size and shape (orientation) of the ellipsoidal discrimination regions in the corresponding image representation domain. The diagonal or non-diagonal nature of the metric determines whether the discrimination regions are oriented along the axes of the representation. The magnitude of the metric elements determines the size of the discrimination regions.

In particular, in the spatial, the wavelet, and the normalized representations, we have:

$$\begin{aligned} d(\mathbf{x}, \mathbf{x}+\Delta\mathbf{x})^2 &= \Delta\mathbf{x}^T \cdot M(\mathbf{x}) \cdot \Delta\mathbf{x} = \\ &= \Delta\mathbf{w}^T \cdot M(\mathbf{w}) \cdot \Delta\mathbf{w} = \Delta\mathbf{r}^T \cdot I \cdot \Delta\mathbf{r} \end{aligned} \tag{8}$$

Since the sequence of transforms in Eq. 1 are differentiable, a small distortion $\Delta\mathbf{r}$ may be written as:

$$\Delta\mathbf{r} = \nabla\mathbf{R}(\mathbf{w}') \cdot \mathbf{S} \cdot \mathbf{T} \cdot \Delta\mathbf{x} \tag{9}$$

Therefore (from Eqs. 8 and 9), the expression of the metrics in the spatial and the wavelet domain are:

$$M(\mathbf{x}) = \mathbf{T}^T \cdot \mathbf{S} \cdot \nabla\mathbf{R}(\mathbf{w}')^T \cdot \nabla\mathbf{R}(\mathbf{w}') \cdot \mathbf{S} \cdot \mathbf{T} \tag{10}$$

$$M(\mathbf{w}) = \mathbf{S} \cdot \nabla\mathbf{R}(\mathbf{w}')^T \cdot \nabla\mathbf{R}(\mathbf{w}') \cdot \mathbf{S} \tag{11}$$

According to the above expressions, the metric in the spatial and wavelet domains critically depends on the Jacobian of the Divisive Normalization, which is:

$$\begin{aligned} \nabla\mathbf{R}(\mathbf{w}')_{ij} &= \frac{\partial \mathbf{R}_i}{\partial w'_j} = \\ &= \gamma \left(\frac{|w'_i|^{\gamma-1}}{\beta_i + \sum_k H_{ik}|w'_k|^\gamma} \cdot \delta_{ij} - \frac{|w'_i|^\gamma |w'_j|^{\gamma-1}}{(\beta_i + \sum_k H_{ik}|w'_k|^\gamma)^2} \cdot H_{ij} \right) \end{aligned} \tag{12}$$

A number of interesting geometric conclusions can be obtained from the above expressions:

- Linear image spaces are not perceptually Euclidean since the distortion metric is image dependent. As one could expect from contrast masking, the non-linear nature of the Divisive Normalization transform implies that the visibility of a given distortion $\Delta\mathbf{x}$ depends on the background image $\mathbf{x}$.

- Discrimination regions increase with the contrast of the image. Note that the elements of the Jacobian $\nabla\mathbf{R}$ (Eq. 12) decrease as the magnitude of the wavelet coefficients (or contrast of the image components) increases. The reduction of the sensitivity is bigger in high activity regions where a number of linear sensors $|w'_k|$ have non-zero values in the denominators of Eq. 12.

- Discrimination regions are not aligned with the axes of the wavelet representation. Note that the Jacobian has a positive diagonal contribution (proportional to δ_{ij}) and a negative non-diagonal contribution due to the kernel, H_{ij}, and depending on w'_i and w'_j with $i \neq j$. This coupling implies that the discrimination ellipsoids are not oriented along the axes of the wavelet representation. Since the Jacobian is input dependent, it can not be strictly diagonalized in any linear representation.

The above considerations on the metric, $M(\mathbf{w})$, analytically demonstrate that the appealing geometric behavior of structural similarity techniques (as in Fig. 4 in [17]) can be shared by error visibility techniques when considering non-linearities including relations among wavelet coefficients (e.g. Divisive Normalization).

Note also that the above considerations (that show that the geometric criticism does not apply to the Divisive Normalization metric) still hold even though non-quadratic schemes are considered. In that general case the shape of the discrimination regions will not be ellipsoidal, but still its size and orientation will be determined by $\nabla\mathbf{R}(\mathbf{w}') \cdot \mathbf{S} \cdot \mathbf{T}$ or $\nabla\mathbf{R}(\mathbf{w}') \cdot \mathbf{S}$.

2.D. Statistical effect of the Divisive Normalization

Euclidean metrics and Minkowski pooling in the response domain implicitly assume statistical independence among the coefficients of the representation since the distortions in every coefficient are individually considered. Existence of relations among the coefficients would imply the consideration of couplings among pairs (or bigger groups) of coefficients. Therefore, from the statistical point of view Minkowski pooling is fully justified in domains in which the relations among coefficients are negligible.

In order to assess the independence among coefficients in the proposed Divisive Normalization domain we use mutual information (MI) measures. In order to gather the appropriate amount of data for MI estimation, we took 8000 patches of size 72×72 from the McGill image data base [34] and computed their wavelet transform and their Divisive Normalization transform. 120000 pairs of coefficients were used in each MI estimation. Two kinds of MI estimators were used: (1) direct computation of MI, which involves 2D histogram estimation [43], and (2) estimation of MI by PCA-based Gaussianization (GPCA) [44], which only involves univariate histogram estimations. Table 2 shows the MI results (in bits) for pairs of coefficients in the wavelet and the divisive normalized domains. The spatial (intra-band) and the frequency (inter-scale and inter-orientation) relations were explored. Just for reference, the MI among luminance values in the spatial domain is 1.79 bits.

These results are consistent with previous redundancy reduction results of Divisive Normalization transform in other domains (and with different parameters) [13, 45], thus suggesting that this particular vision model (optimized for image quality purposes) also reduces dramatically the dependence among coefficients. The statistical effect of the proposed Divisive Normalization has been analyzed in detail in [46]. This fact statistically justifies the use of simple Minkowski pooling in the considered case and addresses the corresponding criticism.

Table 2. MI measures in bits. GPCA MI estimations are shown in parenthesis.

	Wavelet	Div. Norm.
Intraband (scale = 2)	0.29 (0.27)	0.16 (0.15)
Intraband (scale = 3)	0.24 (0.22)	0.09 (0.09)
Inter-scale, scales = (1,2)	0.17 (0.17)	0.08 (0.08)
Inter-scale, scales = (2,3)	0.17 (0.15)	0.04 (0.04)
Inter-scale, scales = (3,4)	0.09 (0.07)	0.01 (0.01)
Inter-orientation (H-V), scale = 2	0.10 (0.08)	0.01 (0.01)
Inter-orientation (H-V), scale = 3	0.08 (0.06)	0.01 (0.01)
Inter-orientation (H-D), scale = 2	0.16 (0.15)	0.03 (0.03)
Inter-orientation (H-D), scale = 3	0.15 (0.14)	0.02 (0.02)

2.E. Relations to other error visibility *metrics*

The proposed model can reproduce Just Noticeable Differences (JNDs), which is a key factor in other recent error visibility metrics [14]. JNDs of a certain target can be computed from the inverse of the slope of the corresponding non-linear response.

On the other hand, if the proposed model is simplified to be completely linear by setting, $\nabla \mathbf{R} = I$, the proposed metric reduces to $M(\mathbf{w}) = \mathbf{S}^2$. In this case, the distortion is just the sum of differences in the transform domain weighted by the contrast sensitivity values (as in [7]): $d(\mathbf{x}, \mathbf{x}') = (\sum_i S_i^2 \Delta w_i^2)^{\frac{1}{2}}$.

If the proposed model is simplified to be point-wise non-linear by neglecting the non-diagonal elements in $\nabla \mathbf{R}$, a contrast dependent behavior (smaller sensitivity for higher contrasts) is achieved as in [9–11].

3. Metric results

In this section we compare the performance of the proposed Divisive Normalization metric[2] with state of the art structural similarity metrics (SSIM [17] and MSSIM [16]), and information theoretic measures (VIF [20]) on a number recently available subjectively rated databases (LIVE [39, 47], TID [40], IVC [41], Cornell (on-line supplement to [14])[3]). Note that more recent structural measures on wavelet domains (such as CW-SSIM [18]) are designed to take into account phase distortions (translations and rotations). For registered images, as is the case in the available databases, the results of CW-SSIM basically reduce to

[2]Available at: `http://www.uv.es/vista/vistavalencia/div_norm_metric/`

[3]Available at: `http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html`

the results of previously reported structural measures[4]. On-line available implementations from the authors were used in each case (SSIM[5], VIF and MSSIM[6]). In the SSIM case, the two available implementations were used: the standard one (`ssim_index.m`), and a posterior recommended modification (`ssim.m`) that subsamples the images to look for the best scale to apply SSIM. This will be referred as SSIM_{sub} in the experiments. SSIM results will not be shown since they are always worse than those obtained with SSIM_{sub}. In every case, we used the RGB to Luminance conversion recommended by the authors. In the experiments we also include the Euclidean measure RMSE for illustration purposes.

The experiments will be analyzed in two parts: (1) LIVE database, and (2) additional databases with different distortions.

This distinction comes from the fact that even though a small subset of images of the LIVE database was used to derive the parameters of the Divisive Normalization model, all the five distortions in the LIVE database were used. One could argue that using LIVE to check the performance of the model is not fair since it learnt the distortions. According to this, we will show the results on the whole LIVE database for illustrative purposes, but more interestingly, we will check the generalization ability of the model using data of other subjectively rated databases corresponding to distortions *not included* in the LIVE database.

The good performance of the proposed metric on the new data can not come from over fitting a particular database, but from the fact that it accurately models human perception. A different indication of this accuracy is that even though the model was set using suprathreshold data, it also reproduces the basic trends of threshold psychophysics (frequency sensitivity and contrast masking, as shown in Figs. 2 and 5).

3.A. Accuracy of a metric: correlations and calibration functions

Representing the ground truth subjective distortions (referred to as DMOS) as a function of the distances, d, computed by some metric leads to a scatter plot. Ideally, the data in this scatter plot should follow a straight line thus showing a perfect correlation among the computed distances and the subjective ratings. In real situations the data depart from this ideal behavior.

From the *engineering* point of view, any monotonic (not necessarily linear) relation between d and DMOS is good enough provided that the calibration function, $\text{DMOS} = f(d)$, is known by the metric user. According to this, non-parametric rank order correlation measures (such as the Spearman correlation) or prediction errors using standard non-linear calibration functions have been used to measure the accuracy of the distortion metrics [47, 48]. Rank order correlations have been criticized for a number of reasons [47]: they do not take into

[4]Personal communication by Z. Wang.

[5]Available at: `http://www.ece.uwaterloo.ca/~z70wang/research/ssim/`

[6]Available at: `http://live.ece.utexas.edu/research/quality/`

account the magnitude of the departure from the predicted behavior and, as a result, it is difficult to obtain useful confidence intervals to discriminate between metrics. Therefore, even though the Spearman correlation is usually given for illustrative purposes, F-test on the quotient of the sum of squared prediction errors using standard non-linear calibration functions is usually preferred [47], and has been extensively used [14, 20, 35, 47, 48].

However, from the *vision science* point of view, systematic deviations from the linear prediction suggest a failure (or limitation) of the underlying model: residual non-linearities should be avoided by including the appropriate (perceptually meaningful) correction in the model, instead of using an *ad-hoc* calibration afterwards. Besides, since distortion metrics are commonly used without reference to such calibration functions[7], the unexperienced user may (erroneously) interpret the metric results in a linear way.

In the experiments below we analyze the results of the considered metrics by using the standard F-test [47] along with the intuitive linear calibration and the previously reported (standard) non-linear calibration functions [14, 20, 48]. Even though we feel that linear calibration is the most intuitive scale for the final user and the most challenging situation for a model intended to reproduce human perception, we will see that the basic message (the proposed error visibility metric is competitive with the newer techniques) is independent from the calibration measure. This is good since F-test may be criticized as well because it depends on an arguable choice of the calibration function. For illustration purposes we will also include the (linear) Pearson correlation and the Spearman correlation. Note that the Pearson correlation on the raw data as done here conveys the same kind of information as the F-test when using a linear calibration function. The difference is that the F-test is useful to establish confidence levels in the results so that it is easy to assess when the differences in prediction errors (or Pearson correlation) are statistically significant.

3.B. Performance of the metrics

In this section we show the scatter plots, the correlations, the fitted calibration functions and the F-test results for (1) the LIVE database, and (2) additional databases (TID, IVC, and Cornell) excluding LIVE-like distortions. Note that distortions in Cornell database are different since it consists of achromatic images.

As stated above, we used the linear calibration and three additional non-linear calibration functions used in the literature: a 4 parameter logistic [14], a 5 parameter logistic [20, 47], and a 4th order polynomial [48]. In every case, the calibration functions were fitted using the Nelder-Mead simplex search method [49] with equivalent initial guesses (according to the corresponding ranges of the distances).

Provided that the prediction errors of the metrics m_i and m_j are independent and Gaus-

[7] The software implementations [14, 16, 17, 20] do not come with this non-linearity.

sian, the F-test gives the probability that the sum of squared errors of metric i, ε_i^2, is smaller than the corresponding value of metric j, ε_j^2. This probability, $P(\varepsilon_i^2 < \varepsilon_j^2)$, can be used to assess if metric i is better than metric j. The F-test has been applied to compare among the previously reported metrics [14, 47]. Here we apply the same standard procedure. In the case of the proposed Divisive Normalization metric the correlation between its residuals and the residuals of the other metrics is similar or smaller than the equivalent results among the other (previously compared) metrics. Therefore, the independence condition holds as accurately as in previously reported comparisons. Unless explicitly stated, residuals can be taken as Gaussian according to previously used kurtosis-based criteria [14, 47]. Therefore, the Gaussianity condition holds as accurately as in previously reported comparisons.

None of the available image quality databases used an experimental procedure similar to [50] (that gives rise to subjective ratings in meaningful JND units). The differences in the experimental procedures implies that the available results are not ready to be merged into a single database. Nevertheless, the different DMOS data were linearly scaled to fall within the range of the LIVE database for visualization purposes[8]. This convenient linear DMOS scaling is not a problem since (1) a separate analysis for each database is done, and (2) it does not modify the correlation results (either Pearson or Spearman), nor the F-test results (since the scaling is taken into account in fitting the corresponding calibration functions and it cancels out in the quotient of squared errors).

Figures 6-9 show the scatter plots and the fitted functions for the considered metrics in the considered situations (1) LIVE, Fig. 6; and (2) TID, Fig. 7, IVC, Fig. 8, Cornell, Fig. 9. Each distortion is identified by a different symbol/color combination. The details on these distortions can be found in the corresponding references. In every case increasing functions were obtained by linearly turning similarity measures, s, into distortion measures, d (as indicated in the plots).

Note that non-linear fitting functions may be unreliable: too flexible fitting functions (such as the 4th order polynomial and the 5 parameter sigmoid) may give rise to non-monotonic behavior. The behavior of these functions strongly depends on the considered data, thus suggesting that it may not account for more general data.

In tables 3-6 we show the results of the F-test for the quotients of the sum of residuals of the considered metrics in the two considered situations: (1) LIVE, table 3; and (2) TID, table 4, IVC, table 5, Cornell, table 6. In these tables, highlighted cells in a row mean that the model in the row is better than the model in the column at 90% confidence level.

[8]Note that two sets of DMOS scores are available in the LIVE database. In this work we used the set that comes with the on-line file `databaserelease2.zip`, as used in [20].

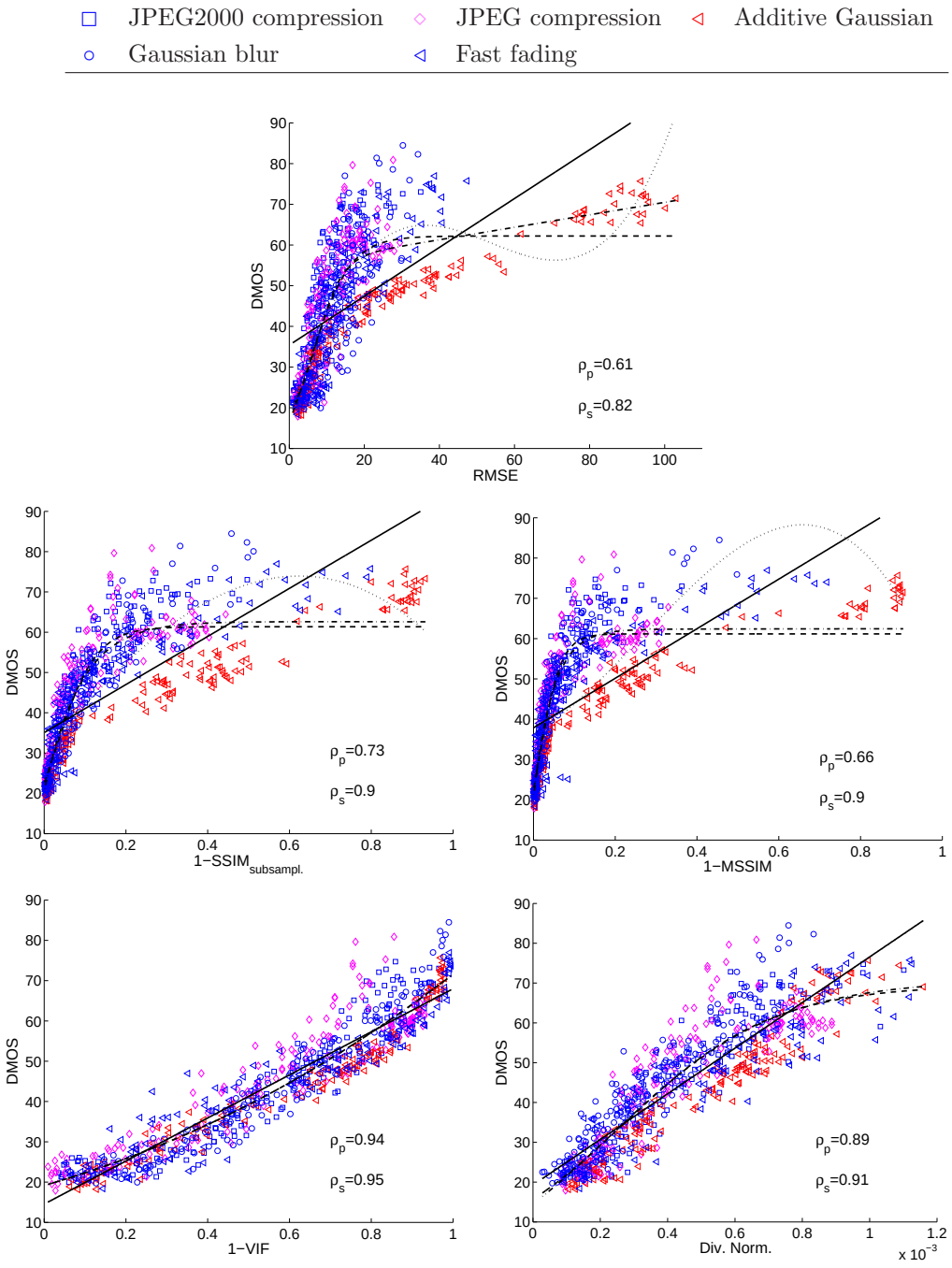

Fig. 6. Scatter plots, fitted functions and correlation coefficients for the considered metrics on the LIVE database. The legend shows the symbols representing each distortion in the LIVE database. The solid line represents the linear fitting. The dashed line represents the 4 parameter sigmoid function used in [14], the dash-dot line represents the 5 parameter sigmoid used in [20, 47]. The dotted line stands for the 4th order polynomial used in [48].

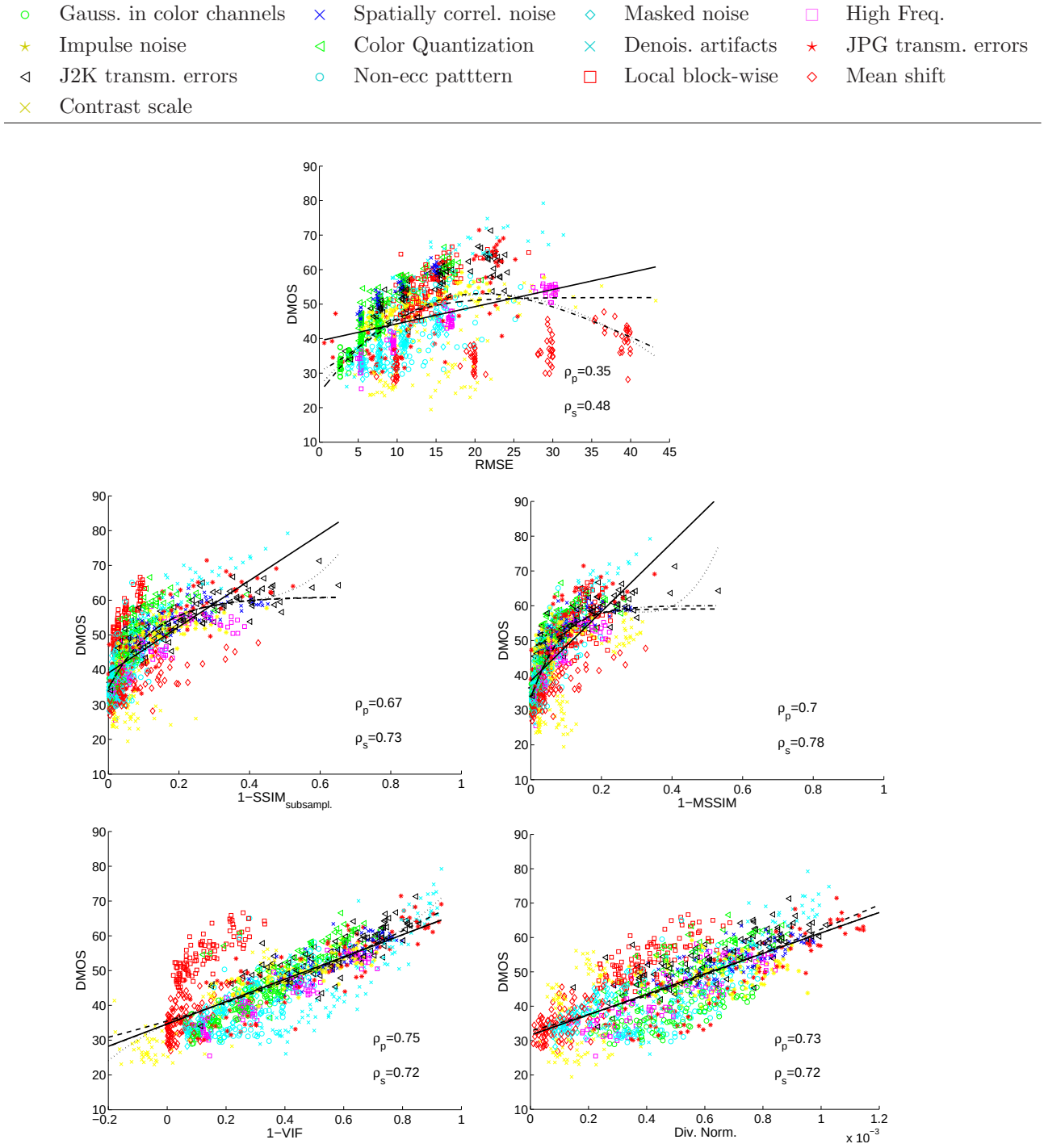

Fig. 7. Scatter plots, fitted functions and correlation coefficients for the considered metrics on the TID database (excluding LIVE-like distortions). The legend represents the symbols corresponding to the distortions which are not present in the LIVE database. Line styles for the calibration functions have the same meaning as in figure 6.

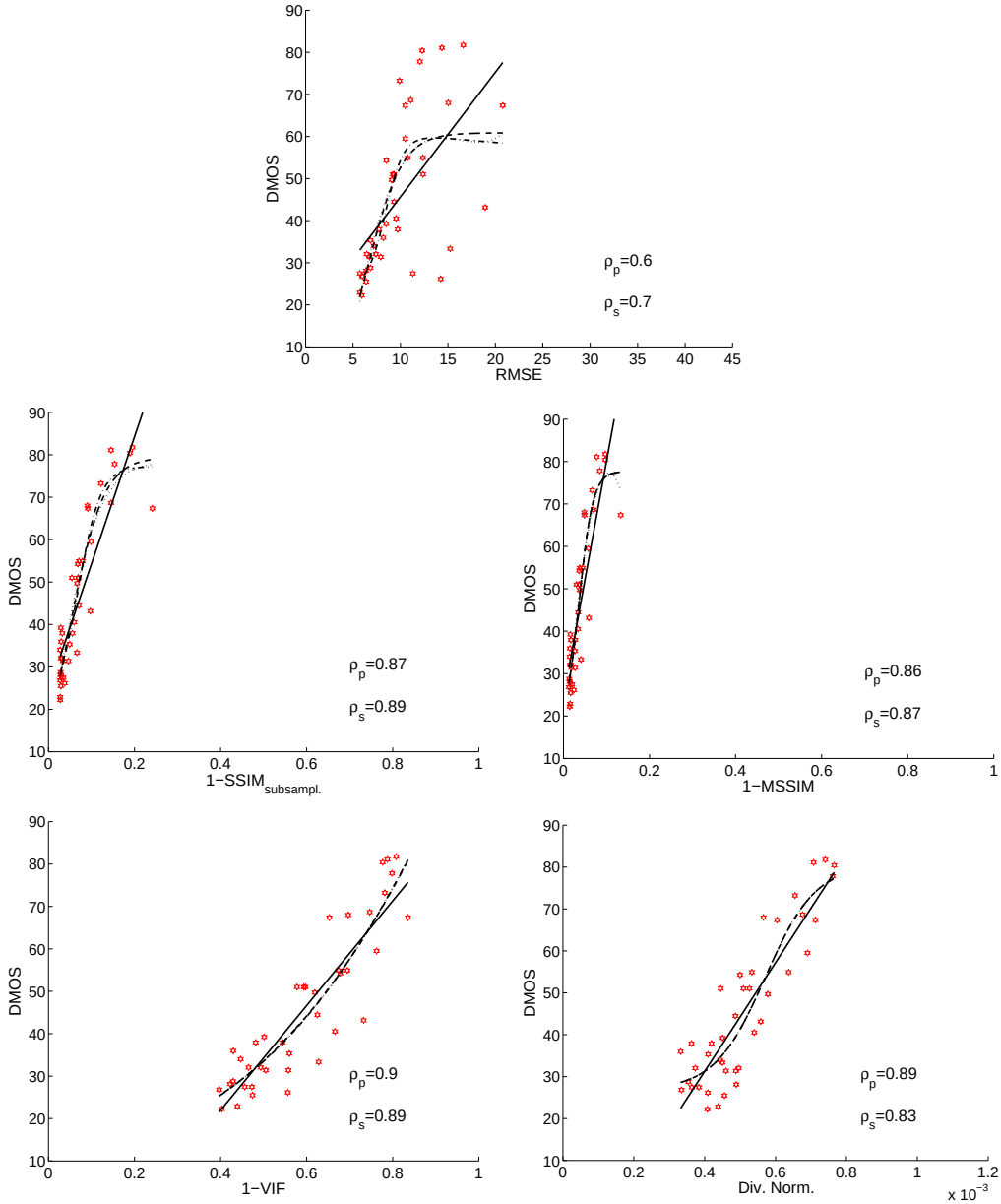

Fig. 8. Scatter plots, fitted functions and correlation coefficients for the considered metrics on the IVC database (excluding LIVE-like distortions). The only non-LIVE distortion in the IVC database is what they call LAR distortion (see [41] for details), depicted here in red stars. Line styles for the calibration functions have the same meaning as in figure 6.

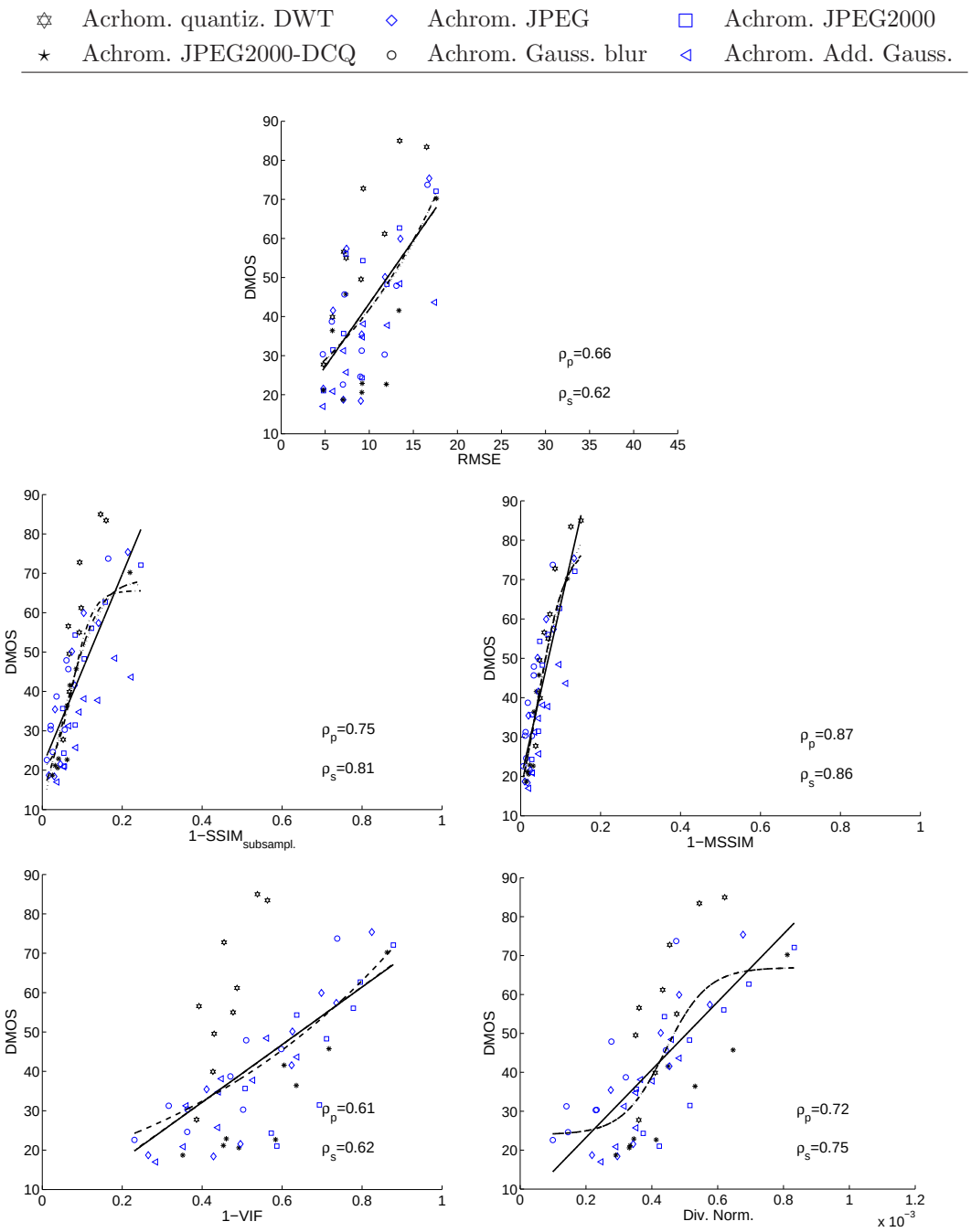

Fig. 9. Scatter plots, fitted functions and correlation coefficients for the considered metrics on the Cornell database. The legend represents the symbols corresponding to the distortions which are not present in the LIVE database (no Cornell distortion is present in LIVE since Cornell is an achromatic database). Line styles for the calibration functions have the same meaning as in figure 6.

Table 3. Quality of metrics on the LIVE database (F-test): probability that the model in the row is better than the model in the column for the linear and several non-linear fits. Highlighted cells mean that model in the row is better than the model in the column at 90% confidence level. The models highlighted with * have non-Gaussian residuals, so the result is not strictly correct.

$P(\varepsilon^2_{row} < \varepsilon^2_{col})$ Linear Fit

	RMSE	$\text{SSIM}_{sub.}$	MSSIM	VIF	DN
RMSE:	-	0.00	0.08	0.00	0.00
$\text{SSIM}_{sub.}$:	1.00	-	1.00	0.00	0.00
MSSIM:	0.92	0.00	-	0.00	0.00
VIF:	1.00	1.00	1.00	-	1.00
DN:	1.00	1.00	1.00	0.00	-

$P(\varepsilon^2_{row} < \varepsilon^2_{col})$ 4 parameter Sigmoid Fit

RMSE:	-	0.00	0.00	0.00	0.00
$\text{SSIM}_{sub.}$:	1.00	-	0.64	0.00	0.09
MSSIM:	1.00	0.36	-	0.00	0.04
VIF*:	1.00	1.00	1.00	-	1.00
DN:	1.00	0.91	0.96	0.00	-

$P(\varepsilon^2_{row} < \varepsilon^2_{col})$ 5 parameter Sigmoid Fit

RMSE:	-	0.00	0.00	0.00	0.00
$\text{SSIM}_{sub.}$:	1.00	-	0.58	0.00	0.14
MSSIM:	1.00	0.42	-	0.00	0.10
VIF*:	1.00	1.00	1.00	-	1.00
DN:	1.00	0.86	0.90	0.00	-

$P(\varepsilon^2_{row} < \varepsilon^2_{col})$ 4th order polynomial Fit

RMSE:	-	0.12	1.00	0.00	0.00
$\text{SSIM}_{sub.}$:	0.88	-	1.00	0.00	0.00
MSSIM:	0.00	0.00	-	0.00	0.00
VIF:	1.00	1.00	1.00	-	1.00
DN:	1.00	1.00	1.00	0.00	-

Table 4. Quality of metrics on the TID database (excluding LIVE-like distortions). See caption of table 3 for details.

$P(\varepsilon^2_{row} < \varepsilon^2_{col})$ Linear Fit

	RMSE	$SSIM_{sub.}$	MSSIM	VIF*	DN
RMSE:	-	0.00	0.00	0.00	0.00
$SSIM_{sub.}$:	1.00	-	0.12	0.00	0.00
MSSIM:	1.00	0.88	-	0.01	0.03
VIF*:	1.00	1.00	0.99	-	0.75
DN:	1.00	1.00	0.97	0.25	-

$P(\varepsilon^2_{row} < \varepsilon^2_{col})$ 4 parameter Sigmoid Fit

RMSE:	-	0.00	0.00	0.00	0.00
$SSIM_{sub.}$:	1.00	-	0.01	0.10	0.29
MSSIM*:	1.00	0.99	-	0.86	0.97
VIF*:	1.00	0.90	0.14	-	0.77
DN:	1.00	0.71	0.03	0.23	-

$P(\varepsilon^2_{row} < \varepsilon^2_{col})$ 5 parameter Sigmoid Fit

RMSE:	-	0.00	0.00	0.00	0.00
$SSIM_{sub.}$:	1.00	-	0.01	0.13	0.32
MSSIM*:	1.00	0.99	-	0.91	0.98
VIF*:	1.00	0.87	0.09	-	0.75
DN:	1.00	0.68	0.02	0.25	-

$P(\varepsilon^2_{row} < \varepsilon^2_{col})$ 4th order polynomial Fit

RMSE:	-	0.00	0.00	0.00	0.00
$SSIM_{sub.}$:	1.00	-	0.01	0.05	0.31
MSSIM*:	1.00	0.99	-	0.79	0.97
VIF:	1.00	0.95	0.21	-	0.86
DN:	1.00	0.69	0.03	0.14	-

Table 5. Quality of metrics on the IVC database (excluding LIVE-like distortions). See caption of table 3 for details.

$P(\varepsilon^2_{row} < \varepsilon^2_{col})$ Linear Fit

	RMSE	$SSIM^{\star}_{sub.}$	MSSIM*	VIF	DN
RMSE:	-	0.00	0.00	0.00	0.00
$SSIM^{\star}_{sub.}$:	1.00	-	0.62	0.28	0.34
MSSIM*:	1.00	0.38	-	0.19	0.23
VIF:	1.00	0.72	0.81	-	0.56
DN:	1.00	0.66	0.77	0.44	-

$P(\varepsilon^2_{row} < \varepsilon^2_{col})$ 4 parameter Sigmoid Fit

RMSE:	-	0.00	0.00	0.00	0.00
$SSIM_{sub.}$:	1.00	-	0.72	0.86	0.86
MSSIM*:	1.00	0.28	-	0.69	0.68
VIF:	1.00	0.14	0.31	-	0.50
DN:	1.00	0.14	0.32	0.50	-

$P(\varepsilon^2_{row} < \varepsilon^2_{col})$ 5 parameter Sigmoid Fit

RMSE:	-	0.00	0.00	0.00	0.00
$SSIM_{sub.}$:	1.00	-	0.75	0.87	0.87
MSSIM*:	1.00	0.25	-	0.68	0.68
VIF:	1.00	0.13	0.32	-	0.50
DN:	1.00	0.13	0.32	0.50	-

$P(\varepsilon^2_{row} < \varepsilon^2_{col})$ 4th order polynomial Fit

RMSE:	-	0.00	0.00	0.00	0.01
$SSIM_{sub.}$:	1.00	-	0.64	0.81	0.89
MSSIM*:	1.00	0.36	-	0.70	0.82
VIF:	1.00	0.19	0.30	-	0.65
DN:	0.99	0.10	0.18	0.35	-

Table 6. Quality of metrics on the (achromatic) Cornell database. See caption of table 3 for details.

$P(\varepsilon^2_{row} < \varepsilon^2_{col})$ Linear Fit

	RMSE	$\text{SSIM}_{sub.}$	MSSIM	VIF*	DN
RMSE:	-	0.17	0.00	0.63	0.27
$\text{SSIM}_{sub.}$:	0.83	-	0.03	0.90	0.63
MSSIM:	1.00	0.97	-	1.00	0.99
VIF*:	0.37	0.10	0.00	-	0.17
DN:	0.73	0.37	0.01	0.83	-

$P(\varepsilon^2_{row} < \varepsilon^2_{col})$ 4 parameter Sigmoid Fit

RMSE:	-	0.07	0.00	0.64	0.18
$\text{SSIM}_{sub.}$:	0.93	-	0.06	0.97	0.72
MSSIM:	1.00	0.94	-	1.00	0.98
VIF*:	0.36	0.03	0.00	-	0.10
DN:	0.82	0.28	0.02	0.90	-

$P(\varepsilon^2_{row} < \varepsilon^2_{col})$ 5 parameter Sigmoid Fit

RMSE:	-	0.05	0.00	0.63	0.17
$\text{SSIM}_{sub.}$:	0.95	-	0.08	0.97	0.74
MSSIM:	1.00	0.92	-	1.00	0.98
VIF*:	0.37	0.03	0.00	-	0.10
DN:	0.83	0.26	0.02	0.90	-

$P(\varepsilon^2_{row} < \varepsilon^2_{col})$ 4th order polynomial Fit

RMSE:	-	0.08	0.00	0.64	0.28
$\text{SSIM}_{sub.}$:	0.92	-	0.06	0.96	0.80
MSSIM:	1.00	0.94	-	1.00	0.99
VIF*:	0.36	0.04	0.00	-	0.17
DN:	0.72	0.20	0.01	0.83	-

3.C. Discussion

In the LIVE case, VIF is the best performing metric. The proposed Divisive Normalization metric is the second best and shows a significantly better performance than structural methods. This reveals that the proposed model can adequately account for the whole database even though its parameters were set by using the 10% of the data (and cropped images). This good performance is independent from the fitting function.

When considering a wider range of distortions (TID and IVC), and using the most challenging linear fit, no algorithm outperforms the proposed Divisive Normalization metric. In the (small) Cornell database, MSSIM is the only metric that significantly outperforms the proposed metric. However, note that the proposed metric significantly outperforms MSSIM in the (bigger) LIVE case no matter the calibration function.

To summarize, the proposed metric performs quite well in the LIVE database (5 distortions) and successfully generalizes to a wide range of distortions (e.g. 20 new distortions in the TID, IVC and Cornell databases). This suggests that the parameters found are perceptually meaningful thus giving rise to a robust metric. In most of the cases the proposed metric is statistically indistinguishable from structural and information theoretic methods. In some particular cases, it is outperformed by VIF (as in LIVE) or by MSSIM (as in Cornell), but it is important to note that, conversely, it significantly outperforms MSSIM in LIVE, and works better than VIF in Cornell (at 80% confidence level). The above is true for all the considered calibration functions.

As a result, the proposed error visibility metric based on Divisive Normalization seems to be competitive with structural and information theoretic metrics. It is quite robust and easy to interpret in linear terms. This is consistent with the fact that the criticisms made to the error visibility techniques do not apply to the Divisive Normalization metric as shown in sections 2.B, 2.C and 2.D.

4. Conclusions and further work

In this work, the classical Divisive Normalization metric [1] was revisited to address the criticisms raised against error visibility techniques. The model was generalized to include weighted relations among coefficients (as in [15]) and extended to work with color images (as in [36]). It was straightforwardly fitted by using a small subset of the subjectively rated LIVE database, and proved to generalize quite well for the whole database as well as for more general databases including distortions of different nature (e.g. TID, IVC, Cornell) .

We showed that the three basic criticisms made against error visibility techniques do not apply to the Divisive Normalization metric: (1) even though the Divisive Normalization is inspired in low-level (threshold) psychophysical and physiological data, it can account for higher-level (suprathreshold) distortions while approximately reproducing the frequency sen-

sitivity and masking results. (2) It was analytically shown that the Divisive Normalization has a rich geometric behavior, so it is not a singular feature of structural similarity metrics. (3) It was shown that the Divisive Normalization representation reduces the statistical relations among the image coefficients, thus justifying the use of uniform Minkowski summation strategies in the normalized domain.

The experiments show that the proposed metric is competitive with structural and information theoretic metrics, it performs consistently when facing a wide range of distortions and it is easy to interpret in linear terms. These results suggest that the classical error visibility approach based on gain control models should still be considered in the image quality discussion.

In fact, the proposed Divisive Normalization framework can still be improved in many ways. The linear chromatic and spatial transforms can be improved by (1) using non-linear color representations to account for the chromatic adaptation ability of human observers [51], and (2) better wavelet transforms may be used for a better simulation of V1 receptive fields (e.g. steerable pyramids [52]). Useful Divisive Normalization transforms for image enhancement have already been proposed on steerable pyramids [53]. Different wavelet basis (as in CW-SSIM [18]) could be used to introduce translation and rotation invariance. Better (non-linear) color representations can be useful to assess changes in average luminance or in the spectral radiance (i.e. including color constancy). Linear models may overestimate the effect of such distortions. The proposed non-linear transform can also be generalized since masking interactions among sensors of different chromatic channels may occur [54], but they were not considered here in order to keep the interaction kernel small. Summation over the color dimension can be generalized as well by including different summation exponents on the opponent channels. Another issue to be explored is the role of the low-frequency residual which was neglected in this work. Weber-law like non-linearities should be used in this case (in agreement with non-linear color appearance models) together with and appropriate relative weight between the low-pass and the higher frequency subbands. From a more general point of view, the proposed model may be complemented by bottom-up techniques for saliency prediction based on the V1 image representation [55]. Finally, better optimization techniques instead of the reported exhaustive search of the parameter subspaces may be used in order to obtain a more accurate estimation of the optimal parameters with a reduced computational burden.

References

1. P. Teo and D. Heeger, "Perceptual image distortion," *Proceedings of the SPIE*, vol. 2179, pp. 127–141, 1994.
2. Z. Wang and A. Bovik, "Mean squared error: Love it or leave it?" *IEEE Signal Processing Magazine*, pp. 98–117, Jan. 2009.
3. A. Ahumada, "Computational image quality metrics: A review," in *Intl. Symp. Dig. of Tech. Papers*, ser. Proceedings of the SID, J. Morreale, Ed., vol. 24, 1993, pp. 305–308.
4. A. B. Watson, *Digital Images and Human Vision*. Massachusetts: MIT Press, 1993.
5. J. Lubin, "*The Use of Psychophysical Data and Models in the Analysis of Display System Performance*," in *Digital Images and Human Vision*, A. Watson, Ed. Massachusetts: MIT Press, 1993, pp. 163–178.
6. L. Saghri, P. Cheatheam, and A. Habibi, "Image quality measure based on a human visual system model," *Optical Engineering*, vol. 28, no. 7, pp. 813–819, 1989.
7. N. Nill, "A visual model weighted cosine transform for image compression and quality assessment," *IEEE Transactions on Communications*, vol. 33, pp. 551–557, 1985.
8. X. Zhang and B. A. Wandell, "A spatial extension to CIELAB for digital color image reproduction," *Society for Information Display Symposium Technical Digest*, vol. 27, pp. 731–734, 1996.
9. S. Daly, "Application of a noise-adaptive Contrast Sensitivity Function to image data compression," *Optical Engineering*, vol. 29, no. 8, pp. 977–987, 1990.
10. P. Barten, "Evaluation of subjective image quality with the square root integral method," *Journal of the Optical Society of America A*, vol. 7, no. 10, pp. 2024–2031, 1990.
11. J. Malo, A. Pons, and J. Artigas, "Subjective image fidelity metric based on bit allocation of the human visual system in the DCT domain," *Image & Vision Computing*, vol. 15, no. 7, pp. 535–548, 1997.
12. I. Epifanio, J. Gutiérrez, and J.Malo, "Linear transform for simultaneous diagonalization of covariance and perceptual metric matrix in image coding," *Pattern Recognition*, vol. 36, pp. 1799–1811, 2003.
13. J. Malo, I. Epifanio, R. Navarro, and E. Simoncelli, "Non-linear image representation for efficient perceptual coding," *IEEE Transactions on Image Processing*, vol. 15, no. 1, pp. 68–80, 2006.
14. D. Chandler and S. Hemami, "VSNR: A wavelet based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Proc.*, vol. 16, no. 9, pp. 2284–2298, 2007.
15. A. B. Watson and J. A. Solomon, "A model of visual contrast gain control and pattern masking," *JOSA A*, vol. 14, no. 9, pp. 2379–2391, 1997.
16. Z. Wang, E. Simoncelli, and A. Bovik, "Multi-scale structural similarity for image quality

assessment," in *IEEE Asilomar Conf. on Signals, Systems and Computers*, vol. 37, 2003.

17. Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Proc.*, vol. 13, no. 4, pp. 600–612, April 2004.
18. Z. Wang and E. Simoncelli, "Translation insensitive image similarity in complex wavelet domain," in *Proc. IEEE Int. Conf. Ac. Speech, Sig. Proc.*, 2005, pp. 573–576.
19. H. Sheikh, A. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2117–2128, Dec 2005.
20. H. Sheikh and A. Bovik, "Image information and visual quality," *IEEE Trans. Image Proc.*, vol. 15, no. 2, pp. 430–444, Feb 2006.
21. H. B. Barlow, "Possible principles underlying the transformation of sensory messages," in *Sensory Communication*, W. Rosenblith, Ed. Cambridge, MA: MIT Press, 1961, pp. 217–234.
22. O. Schwartz and E. P. Simoncelli, "Natural signal statistics and sensory gain control," *Nature Neuroscience*, vol. 4, no. 8, pp. 819–825, 2001.
23. B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, pp. 607–609, 1996.
24. J. Malo and J. Gutiérrez, "V1 non-linear properties emerge from local-to-global non-linear ICA," *Network: Computation in Neural Systems*, vol. 17, pp. 85–102, 2006.
25. Z. Wang and E. P. Simoncelli, "An adaptive linear system framework for image distortion analysis," in *Proc 12th IEEE Int'l Conf on Image Proc*, vol. III. Genoa, Italy: IEEE Computer Society, 11-14 Sep 2005, pp. 1160–1163.
26. K. Seshadrinathan and A. Bovik, "Unifying analysis of full reference image quality assessment," in *IEEE Intl. Conf. Image Processing*, 2008, pp. 1200–1203.
27. F. W. Campbell and J. G. Robson, "Application of Fourier analysis to the visibility of gratings," *Journal of Physiology*, vol. 197, no. 3, pp. 551–566, August 1968.
28. J. Malo, A. M. Pons, A. Felipe, and J. Artigas, "Characterization of human visual system threshold performance by a weighting function in the Gabor domain," *Journal of Modern Optics*, vol. 44, no. 1, pp. 127–148, 1997.
29. D. J. Heeger, "Normalization of cell responses in cat striate cortex," *Vis. Neurosci.*, vol. 9, pp. 181–198, 1992.
30. E. Simoncelli and E. Adelson, *Subband Image Coding*. Norwell, MA: Kluwer Academic Publishers, 1990, ch. Subband Transforms, pp. 143–192.
31. J. Gutiérrez, F. J. Ferri, and J. Malo, "Regularization operators for natural images based on nonlinear perception models," *IEEE Transactions on Image Processing*, vol. 15, no. 1, pp. 189–2000, January 2006.

32. G. Camps, G. Gómez, J. Gutiérrez, and J. Malo, "On the suitable domain for SVM training in image coding," *J. Mach. Learn. Res.*, vol. 9, pp. 49–66, 2008.
33. V. Laparra, J. Gutiérrez, G. Camps, and J. Malo, "Image denoising with kernels based on natural image relations," *Accepted in: J. Mach. Learn. Res.*, 2010.
34. A. Olmos and F. A. A. Kingdom, "Mcgill calibrated colour image database," *http://tabby.vision.mcgill.ca.*, 2004.
35. A. B. Watson and J. Malo, "Video quality measures based on the Standard Spatial Observer," in *Proc. of the IEEE Intl. Conf. Image Processing*, vol. 3, 2002, pp. 41–44.
36. E. Martinez-Uriegas, "Color detection and color contrast discrimination thresholds," in *Proceedings of the OSA Annual Meeting ILS–XIII*, Los Angeles, 1997, p. 81.
37. W. Pratt, *Digital Image Processing.* New York: John Wiley & Sons, 1991, ch. 3: *Photometry and Colorimetry.*
38. K. T. Mullen, "The contrast sensitivity of human colour vision to red-green and yellow-blue chromatic gratings," *Journal of Physiology*, vol. 359, pp. 381–400, 1985.
39. H. Sheikh, Z.Wang, L. Cormack, and A. Bovik, "LIVE image quality assessment database," 2006. [Online]. Available: http://live.ece.utexas.edu/research/quality
40. N. Ponomarenko, M. Carli, V. Lukin, K. Egiazarian, J. Astola, and F. Battisti, "Color image database for evaluation of image quality metrics," *Proc. Int. Workshop on Multimedia Signal Processing*, pp. 403–408, Oct. 2008.
41. P. Le Callet and F. Autrusseau, "Subjective quality assessment irccyn/ivc database," 2005, http://www.irccyn.ec-nantes.fr/ivcdb/.
42. A. Watson and C. Ramirez, "A Standard Observer for Spatial Vision," *Investig. Opht. and Vis. Sci.*, vol. 41, no. 4, p. S713, 2000.
43. T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York, USA: Wilson and Sons, 1991.
44. V. Laparra, G. Camps, and J. Malo, "PCA gaussianization for image processing," in *Proc. IEEE Int. Conf. Im. Proc.* IEEE, 2009.
45. J.Malo, R.Navarro, I.Epifanio, F.Ferri, and J.M.Artigas, "Non-linear invertible representation for joint statistical and perceptual feature representation," *Lect. Not. Comp. Sci.*, vol. 1876, pp. 658–667, 2000.
46. J. Malo and V. Laparra, "Psychophysically tuned divisive normalization approximately factorizes the PDF of natural images," *Submitted to: Neural Computation*, 2010.
47. H. Sheikh, M. Sabir, and A. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Proc.*, vol. 15, no. 11, pp. 3440–3451, November 2006.
48. VQEG, "Final report from the video quality experts group on the validation of objective models of multimedia quality assessment, phase I," Video Quality Experts

Group, Tech. Rep. 2.6, 2008. [Online]. Available: http://www.its.bldrdoc.gov/vqeg/projects/multimedia/

49. J. Lagarias, J. Reeds, M. Wright, and P. Wright, "Convergence properties of the nelder-mead simplex method in low dimensions," *SIAM Journal of Optimization*, vol. 9, no. 1, pp. 112–147, 1998.
50. A. Watson and L. Kreslake, "Measurement of visual impairment scales for digital video," in *Proc. SPIE Human Vision, Visual Processing, and Digital Display*, vol. 4299, 2001.
51. M. Fairchild, *Color Appearance Models*. New York: Addison-Wesley, 1997.
52. E. Simoncelli, W. Freeman, E. Adelson, and D. Heeger, "Shiftable multi-scale transforms," *IEEE Trans. Information Theory*, vol. 38, no. 2, pp. 587–607, 1992.
53. S. Lyu and E. P. Simoncelli, "Nonlinear image representation using divisive normalization," in *Proc. Computer Vision and Pattern Recognition*. IEEE Computer Society, Jun 23-28 2008, pp. 1–8.
54. K. Gegenfurtner and D. Kiper, "Contrast detection in luminance and chromatic noise," *J. Opt. Soc. Am. A*, vol. 9, no. 11, pp. 1880–1888, Nov. 1992.
55. L. Zhaoping, "Theoretical understanding of the early visual processes by data compression and data selection," *Network: Computation in neural systems*, vol. 17, no. 4, pp. 301–334, 2006.