Psychophysically Tuned Divisive Normalization Approximately Factorizes the PDF of Natural Images

Jesús Malo and Valero Laparra

Image Processing Laboratory, Universitat de València.
Dr. Moliner 50, 46100 Burjassot, València, (Spain).
Jesus.Malo@uv.es, Valero.Laparra@uv.es
http://www.uv.es/vista/vistavalencia

Abstract. The conventional approach in Computational Neuroscience in favor of the efficient encoding hypothesis goes *from image statistics to perception*. It has been argued that the behavior of the early stages of biological visual processing (e.g. spatial frequency analyzers and their non-linearities) may be obtained from image samples and the efficient encoding hypothesis using no psychophysical or physiological information.
In this work we address the same issue in the opposite direction, *from perception to image statistics*: we show that psychophysically fitted image representation in V1 has appealing statistical properties, e.g. approximate PDF factorization and substantial mutual information reduction, even though no statistical information is used to fit the V1 model. These results are a complementary evidence in favor of the efficient encoding hypothesis.

1 Introduction

Horace Barlow suggested that functional properties of biological vision sensors should be matched to the signal statistics faced by these sensors [Barlow, 1961]. The conventional approach to confirm the plausibility of such efficient encoding hypothesis goes *from image statistics to perception*.

Over the last decades a number of evidences in the above *conventional* direction have been reported. First, the shape of the linear receptive fields in V1 was derived using different network architectures and learning algorithms to optimize different statistical criteria such as energy minimization, enforcing decorrelation of the outputs or maximizing the mutual information between input and output: for instance, in [Linsker, 1986, Sanger, 1989, 1990] low-pass filtered random noise was used as a rough model for natural images to feed the networks, while [Foldiak, 1989] focused on information transmission. Then, more attention was devoted to statistical independence beyond decorrelation. When higher order moments are considered in natural images (using linear ICA), sets of localized and oriented edge detectors are found [Olshausen and Field, 1996, Bell and Sejnowski, 1997, van Hateren and van der Schaaf, 1998]. Another linear feature of perception explained from the spectrum of natural images and maximization of signal to noise ratio is the spatial frequency sensitivity [Van Hateren, 1992, 1993]. The works of Van Hateren also explain a global non-linear dependence on

the luminance and the strength of the stimulus in accordance with Weber's law [Van Hateren, 1992, 1993].

More recently, attention has shifted from the linear receptive fields to the specific non-linearities of V1 cells, namely surround effects and contrast adaptation or gain control. In this case, parametric models using divisive normalization [Schwartz and Simoncelli, 2001] or other specific non-linearities [Kayser et al., 2003] have been fitted using image statistics and efficient encoding arguments. Feedback and feedforward connections in hierarchical networks have been used to reproduce surround inhibition [Rao and Ballard, 1999]. Non-parametric approaches, such as non-linear ICA used in [Malo and Gutiérrez, 2006], exemplifies the *image statistics to perception* way of reasoning since the right non-linearities directly emerge from the data using no perceptually inspired functional form.

However, despite the above evidences, nowadays there is a productive debate about the generality of the efficient encoding hypothesis, or the strict applicability of redundancy reduction arguments [Barlow, 2001, Simoncelli, 2003].

In order to contribute to this debate, two complementary lines of research are possible:

- The conventional direction, *from image statistics to perception*, as described above. This approach derives computational models or architectures from statistical principles, and then simulates perceptual (physiological or psychophysical) measurements from the statistically derived model. The eventual match between simulated and experimental recordings suggests that the efficient encoding hypothesis is correct, since the experimental behavior emerges from image statistics (even though no perceptual information was used in deriving such behavior).
- The reverse direction, i.e. *from perception to image statistics*. This approach starts from the response of real neurons (or equivalently from the response of a perceptually derived model) at different stages along the visual pathway. When such a perception system is stimulated with natural images it is possible to obtain statistical measurements about the transmitted signal at different processing stages. The eventual good statistical behavior of the perceptual responses at a certain stage (e.g. independence) suggests that the efficient encoding hypothesis is correct, since the brain is reducing the redundancy in the signal along the visual pathway (even though no statistical information was used in computing these responses -direct recordings or perceptually transformed signals-).

In this work we take the second approach: we show that a psychophysically fitted version of the divisive normalization contrast masking model has appealing statistical properties (e.g. approximate factorization of the PDF of natural images) even though no statistical information is used to fit the model. Therefore, this work can be seen as the *reverse approach* version of [Schwartz and Simoncelli, 2001, Malo and Gutiérrez, 2006], which are examples of the *direct approach* applied to the non-linear behavior of V1 cells.

The structure of the paper is as follows. In section 2 we review the the standard non-linear model of the V1 visual cortex and propose a new (indirect) psychophysical procedure to set its parameters. In our case, the model parameters are obtained to predict perceived distortions on a large subjectively rated

database. Appendices show that the perception model works as well as state-of-the-art image quality metrics using natural stimuli while qualitatively reproducing traditional psychophysics (frequency sensitivity and masking) that uses more simple stimuli. Section 3 analytically shows how the proposed perception model may factorize a plausible PDF for natural images (which captures local image dependencies), provided perception is matched to the statistics. Section 4 empirically shows the good statistical behavior of the perceptual model when exposed to natural images: the non-linear part of the V1 model strongly reduces the mutual information between coefficients of the previous linear stage and approximately achieves the predicted component independence. Finally, section 5 draws the conclusions of the work.

2 The Divisive Normalization V1 model

The perceptual image representation considered here is based on the standard psychophysical and physiological model that describes the early visual processing up to the V1 cortex. The linear part of the model describes the shape of the receptive fields as linear edge detectors tuned to different scales [Daugman, 1980, Watson, 1983, 1987], and accounts for the threshold contrast sensitivity [Campbell and Robson, 1968, Mullen, 1985, Malo, 1997]. The non-linear part of the model accounts for the non-linearities related to contrast masking [Heeger, 1992, Foley, 1994, Watson and Solomon, 1997, Carandini and Heeger, 1994, Carandini et al., 1997]. 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 [Daugman, 1980, Watson, 1983, 1987]. The linear sensors have a frequency dependent linear gain according to the Contrast Sensitivity Function (CSF), $\mathbf{S}_{ii}$, [Campbell and Robson, 1968, Mullen, 1985, Malo, 1997]. The weighted response of these sensors is non-linearly transformed according to the divisive normalization gain control, $\mathbf{R}$ [Heeger, 1992, Foley, 1994, Watson and Solomon, 1997, Carandini and Heeger, 1994, Carandini et al., 1997]:

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

In this scheme, the set of local-frequency analyzers (matrix $\mathbf{T}$) and the slopes of their responses (matrix $\mathbf{S}$) constitute the linear part of the model. The rows of the matrix $\mathbf{T}$ contain the linear receptive fields of V1 neurons. In this paper we used an orthogonal 4-scales QMF wavelet transform[1] [Simoncelli and Adelson, 1990] to model such receptive fields. $\mathbf{S}$ is a diagonal matrix containing the linear gains to model the CSF. Finally, $\mathbf{R}$ is the divisive normalization response which describes the non-linear behavior:

$$\mathbf{R}(\mathbf{w}')_i = r_i = \mathrm{sign}(w'_i)\frac{|S_{ii} \cdot w_i|^\gamma}{\beta_i^\gamma + \sum_{k=1}^{n} H_{ik}|S_{kk} \cdot w_k|^\gamma} \tag{2}$$

where H is a kernel matrix that controls how the responses of neighboring linear sensors, k, affect the non-linear response of sensor i. The constants β_i determine the minimum contrast for significant response saturation.

[1] http://www.cns.nyu.edu/∼lcv/software.php

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 [Martinez-Uriegas, 1997]. In particular, we used the standard YUV (luminance, yellow-blue, red-green) representation [Pratt, 1991]. According to the well known differences in frequency sensitivity in the opponent channels [Mullen, 1985], we will allow for different matrices $\mathbf{S}$ in each channel. 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 [Martinez-Uriegas, 1997].

The natural way to set the parameters of the model is empirical: by fitting low-level perception data, either physiological recordings [Heeger, 1992] or threshold psychophysics [Watson and Solomon, 1997]. This low-level approach is not straightforward because the experimental literature is often interested in a subset of the parameters, and a variety of experimental settings is used (e.g. different stimuli, different contrast definitions, etc.). As a result, it is not easy to unify the wide range of data into a common computational framework. Alternative (theoretical) approaches involve using image statistics and the efficient encoding hypothesis to derive the parameters [Olshausen and Field, 1996, Schwartz and Simoncelli, 2001, Malo and Gutiérrez, 2006]. Obviously, this is not an option in our case since our aim is assessing the statistical efficiency of a non-statistically optimized model.

Instead, in this work we used an empirical but *indirect* approach: we set the parameters of the model to reproduce experimental (but higher-level) visual results such as image quality assessment as in [Watson and Malo, 2002]. In particular, we optimized the V1 model to obtain an image distortion metric that maximizes the correlation with the subjective ratings of a small subset of the LIVE Quality Assessment Database[2] [Sheikh et al., 2006]. Appendix A gives further details on the parametrization and the optimization process.

Figure 1 shows the optimal values for the linear gains $\mathbf{S}$, the saturation constants, β^γ, and the interaction kernel H. The optimal value for the excitation and inhibition exponent was $\gamma = 1.7$. The Matlab implementation of the proposed model is available on-line[3].

Appendix B shows that the obtained model simultaneously accounts for a wide variety of suprathreshold distortions as well as for the basic trends of threshold psychophysics (e.g. frequency sensitivity and contrast masking).

3 PDF factorization through V1 Divisive Normalization

In this section we assume a plausible joint PDF model for natural images in the wavelet domain and we show that this PDF is factorized by a divisive normalization transform, given that some conditions apply. The analytical results shown here predict quite characteristic marginal PDFs in the transformed domain. In section 4 we will empirically check the predictions made here by applying the normalization model proposed above to a set of natural images.

[2] http://live.ece.utexas.edu/research/quality/

[3] http://www.uv.es/vista/vistavalencia/standard_V1_model/

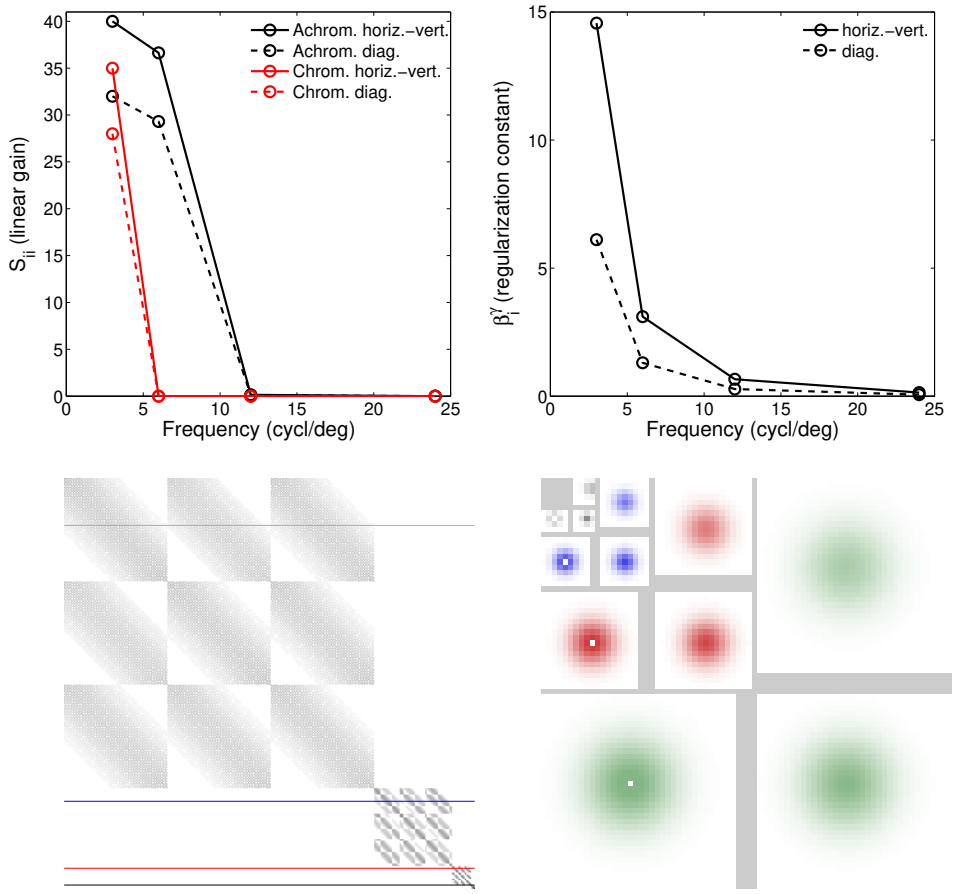


Fig. 1. *Linear gains S (top left), saturation constants β^γ (top right), and kernel H (bottom left). The particular structure of the interaction kernel comes from the particular arrangement of wavelet coefficients used in the transform [Simoncelli and Adelson, 1990]. The bottom right figure shows the individual rows highlighted in different colors in the kernel figure. Each row corresponds to the particular coefficients in white in the bottom right figure. The different shades of color represent the interaction intensity with the spatial and orientation neighbors. In this example we assumed 72×72 discrete images sampled at 64 cycles per degree. According to this, the spatial extent of the subbands is 1.125 degrees.*

3.1 Image model

It is widely known that natural images display a quite characteristic behavior in the wavelet domain: on the one hand, they show heavy-tailed marginal PDFs, $P_{w_i'}(w_i')$ (see Fig. 2), and, on the other hand, the variance of one particular coefficient is related to the variance of the neighbors. These relations are easy to see by looking at the so called bow-tie plot: the conditional probability of a coefficient given the values of some of its neighbors, $P(w_j'|w_i')$, normalized by the maximum of the function for each value of w_i' (see Fig. 2). In this representation tilting of the conditional density suggests that the coefficients are correlated,

but more importantly, it can be seen that the variance of one coefficient strongly depends on the variance of the neighbor. These observations on the marginal and conditional PDFs have been used to propose leptokurtotic functions to model the marginal PDFs [Simoncelli, 1997, 1999, Hyvärinen, 1999] and models of the conditional PDFs in which the variance of one coefficient depends on the variance of the neighbors [Buccigrossi and Simoncelli, 1999, Schwartz and Simoncelli, 2001].

Inspired on these conditional models, we propose the following joint PDF (for the N-dimensional vectors $\mathbf{w}'$), in which, each element of the diagonal matrix, Σ, depends on the neighbors:

$$P_{\mathbf{w}'}(\mathbf{w}') = \frac{1}{Z} \frac{1}{|\Sigma(\mathbf{w}')|^{1/2}} e^{-\frac{1}{2}\,\mathbf{w}'^{T}\cdot\Sigma(\mathbf{w}')^{-1}\cdot\mathbf{w}'} \tag{3}$$

where,

$$\Sigma_{ii}(\mathbf{w}') = (\beta_i^{\gamma} + \sum_j H_{ij} \cdot |w_j'|^{\gamma})^{\frac{2}{\gamma}} \tag{4}$$

and Z is simply a normalization constant to ensure that $\int_{\mathbf{w}'} P_{\mathbf{w}'}(\mathbf{w}')\, d\mathbf{w}' = 1$. Appendix C shows that the normalization constant, Z, is bounded.

The diagonal matrix $\Sigma(\mathbf{w}')$ can be thought as playing similar role as the covariance matrix in a regular Gaussian PDF. However, note that $\Sigma(\mathbf{w}')$ is point dependent (i.e. it is not a covariance matrix), and even though it is diagonal, it introduces relations among the energies of neighbor coefficients (see eq. 4). Therefore, this joint PDF *is not* Gaussian, and the coefficients of the wavelet transform are not independent since the joint PDF, $P_{\mathbf{w}'}(\mathbf{w}')$, cannot be factorized by its marginal PDFs, $P_{w_i'}(w_i')$.

The proposed PDF is inspired by the models used in [Buccigrossi and Simoncelli, 1999, Schwartz and Simoncelli, 2001] since it tries to describe the relations among neighbor coefficients in wavelet domains using linear combinations of them. The differences include (1) the specific exponent, a sort of norm, γ, applied to the coefficients of the wavelet transform used in the linear combination (whether you consider amplitudes, $\gamma = 1$, as in [Buccigrossi and Simoncelli, 1999]; energy, $\gamma = 2$, as in [Schwartz and Simoncelli, 2001]; or some generic γ, here), and (2) the fact that here we are proposing a joint PDF model while in those cases the model was conditional.

A 2D example using the above joint PDF illustrates its suitability to capture the reported marginal and conditional behavior of wavelet coefficients: see the predictions shown in Fig. 2.

3.2 V1 normalized components are approximately independent

Here we compute the PDF of the natural images in the divisive normalized representation assuming (1) the above image model, and (2) the match between the parameters of the V1 representation and the parameters of the image model. Specifically, the match between the denominator in the perceptual response (eq. 2) and the matrix Σ in the image model (eq. 4).

We will use the fact that given the PDF of a random variable, $\mathbf{w}'$, and some transform, $\mathbf{r} = \mathbf{R}(\mathbf{w}')$, the PDF of the transformed variable can be computed by [Stark and Woods, 1994],

$$P_{\mathbf{r}}(\mathbf{r}) = P_{\mathbf{w}'}(\mathbf{R}^{-1}(\mathbf{r})) \cdot |\nabla_{\mathbf{r}}\mathbf{R}^{-1}| \tag{5}$$

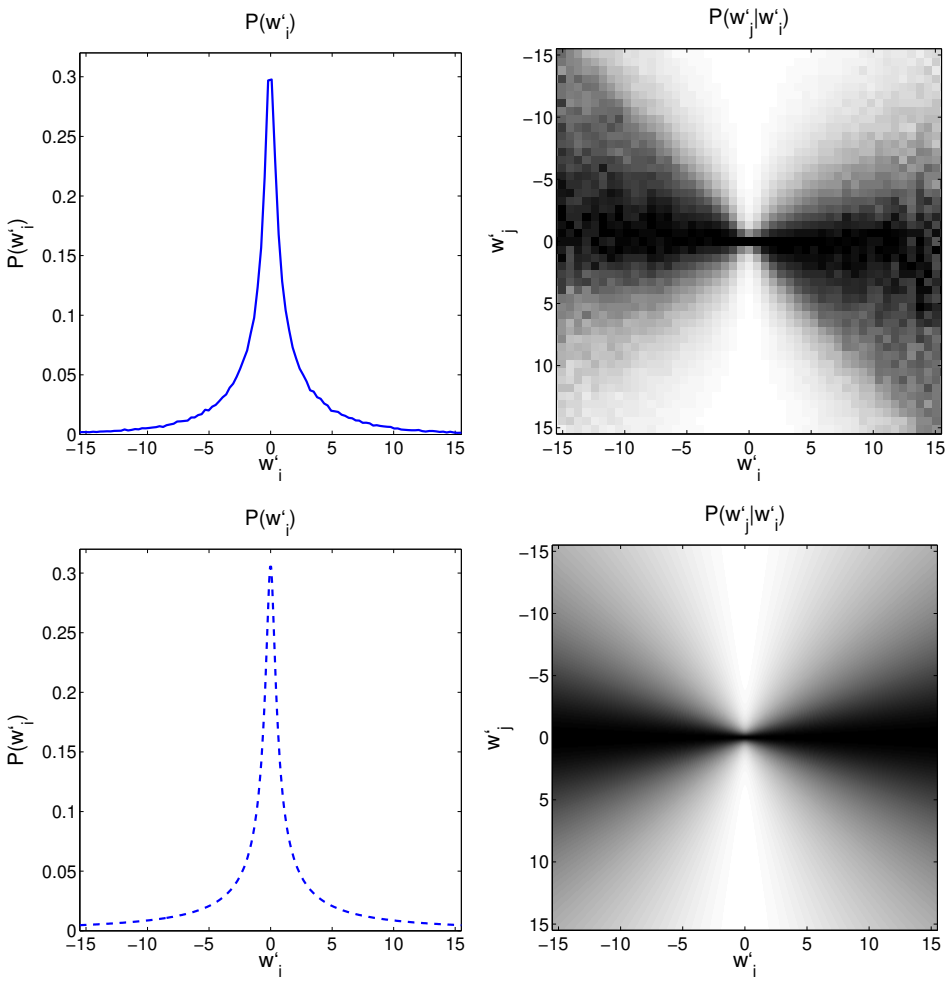


Fig. 2. *Top: empirical behavior of wavelet coefficients of natural images (marginal PDF -left- and conditional PDF -right-). Darker values indicate higher probability. Bottom: simulated behavior according to the proposed model. In this 2D experiment we considered two coefficients of the second scale of* $\mathbf{w}'$ *(computed for 10000 images of the database [Olmos and Kingdom, 2004], using* $3 \cdot 10^6$ *samples). We used* $S_{ii} = 0.14$, $\beta_i = 0.4$, $H_{ii} = 0.7$ *and* $H_{ij} = 0.3$ *and* $\gamma = 1.7$, *according to the psychophysically fitted model.*

Considering that the divisive normalization (in vector notation) is just:

$$\mathbf{r} = sign(\mathbf{w}')\, \Sigma(\mathbf{w}')^{-\frac{\gamma}{2}} \cdot |\mathbf{w}'|^{\gamma} \tag{6}$$

where $|\cdot|^{\gamma}$ is an element-by-element exponentiation, the inverse, $\mathbf{R}^{-1}$, can be obtained from one of these (equivalent) expressions [Malo et al., 2006]:

$$|\mathbf{w}'|^{\gamma} = (I - D_{|\mathbf{r}|}H)^{-1} \cdot D_{\beta^{\gamma}} \cdot |\mathbf{r}| \tag{7}$$

$$\mathbf{w}' = sign(\mathbf{r})\Sigma(\mathbf{w}')^{\frac{1}{2}} \cdot |\mathbf{r}|^{\frac{1}{\gamma}} \tag{8}$$

where D_v are diagonal matrices with the vector v in the diagonal. Plugging $\mathbf{w}'$, eq. 8, into the image model we have,

$$P_{\mathbf{w}'}(\mathbf{R}^{-1}(\mathbf{r})) = \frac{1}{Z}\,\frac{1}{|\Sigma(\mathbf{w}')|^{1/2}}\,e^{-\frac{1}{2}\,(|\mathbf{r}|^{1/\gamma})^T\cdot I\cdot(|\mathbf{r}|^{1/\gamma})} \tag{9}$$

Taking derivatives on the inverse, eq. 7, the determinant of the Jacobian is:

$$|\nabla_{\mathbf{r}}\mathbf{R}^{-1}| = det\left(\frac{1}{\gamma}\,\Sigma(\mathbf{w}')^{1/2}\cdot D_{|\mathbf{r}|^{\frac{1}{\gamma}-1}}\cdot\left(I+\underbrace{D_{\beta-\gamma}\cdot H\cdot(I-D_{|\mathbf{r}|}H)^{-1}\cdot D_{\beta\gamma}\cdot D_{|\mathbf{r}|}}_{M(\mathbf{r})}\right)\right)$$

$$|\nabla_{\mathbf{r}}\mathbf{R}^{-1}| = det\left(\frac{1}{\gamma}\,\Sigma(\mathbf{w}')^{1/2}\cdot D_{|\mathbf{r}|^{\frac{1}{\gamma}-1}}\cdot(I+M(\mathbf{r}))\right)$$

$$|\nabla_{\mathbf{r}}\mathbf{R}^{-1}| = \;|\Sigma(\mathbf{w}')|^{1/2}\cdot\prod_{i=1}^{N}\frac{1}{\gamma}|r_i|^{\frac{1}{\gamma}-1}det(I+M(\mathbf{r}))^{\frac{1}{N}}$$

Since $det(I+M(\mathbf{r}))^{\frac{1}{N}}\approx 1$ in natural images[4], it follows,

$$|\nabla_{\mathbf{r}}\mathbf{R}^{-1}| \approx \;|\Sigma(\mathbf{w}')|^{1/2}\cdot\prod_{i=1}^{N}\frac{1}{\gamma}|r_i|^{\frac{1}{\gamma}-1} \tag{10}$$

Therefore, from Eqs. 5, 9 and 10, it follows that the joint PDF of the normalized signal is just the product of N functions that depend solely on r_i:

$$P_{\mathbf{r}}(\mathbf{r}) \approx \prod_{i=1}^{N}\frac{1}{\gamma\,Z^{1/N}}\;|r_i|^{\frac{1}{\gamma}-1}\;e^{-\frac{|r_i|^{2/\gamma}}{2}} = \prod_{i=1}^{N}P_{r_i}(r_i) \tag{11}$$

i.e., we have factorized the joint PDF into its marginal PDFs.

Even though factorization of the PDF does not depend on the particular γ, (provided the normalization transform uses the appropriate γ value), this exponent determines the shape of the marginal PDFs (see Fig. 3). In particular, if the appropriate value were $\gamma = 1$, the transform would give rise to Gaussian marginal PDFs thus becoming similar to Radial Gaussianization transforms as suggested in [Lyu and Simoncelli, 2009]. However, note that different values of γ in the transform would imply a better (or worse) match between the denominator of the normalization and the matrix Σ of the image model. This match is required to achieve the factorization result in eq. 11.

4 Results

This section assesses the component independence performance of the psychophysically fitted V1 representation (i.e. the validity of Eq. 11) by (1) analyzing the marginal and conditional probabilities of the transformed coefficients, and (2)

[4] We found that the average value and standard deviation of this determinant on 10000 images taken from McGill calibrated image dataset [Olmos and Kingdom, 2004] is: $\langle det(I+M(\mathbf{r}))^{\frac{1}{N}}\rangle = 1.013 \pm 0.003$.

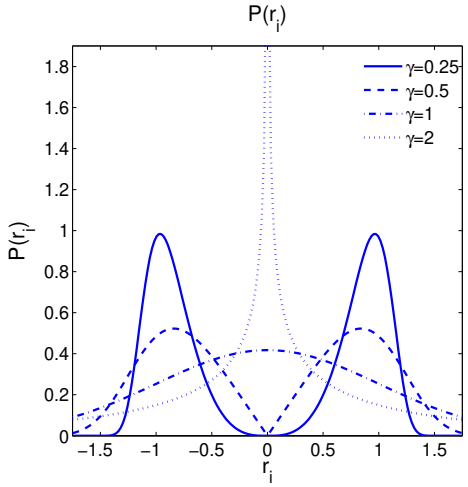


Fig. 3. *Family of marginal PDFs of the normalized coefficients r_i as a function of γ.*

by mutual information measures. To do so, 10000 image patches of size 72×72 from the McGill database [Olmos and Kingdom, 2004] were considered and transformed to the linear V1 representation (the wavelet domain), and to the non-linear V1 representation.

4.1 Marginal and conditional PDFs

Figure 4 shows the experimental and the predicted marginal and conditional PDFs in the normalized domain. These results correspond to two spatial neighbors of the second scale and horizontal orientation ($3 \cdot 10^6$ 2D samples). Similar results are obtained for other subbands. For the sake of illustration, in the case of the marginal PDFs, we show the results for different values of the exponent γ in the transform: the psychophysically optimal value, $\gamma = 1.7$, and other values, $\gamma = 0.5$ and $\gamma = 0.25$, due to the characteristic bimodal shape of the predicted marginal PDFs in those cases (see Fig. 3).

Bimodal results are obtained in the marginal PDFs for the (psychophysically non-optimal) values of γ as predicted by the theory. However, note that the agreement with the theoretical prediction is much better for the psychophysically optimal exponent, thus indicating the match of the psychophysical vision model to image statistics.

The result for the conditional probability shows that the vision model substantially reduces the redundancy among neighbor coefficients with regard to the linear wavelet representation: note that the bow-tie has practically disappeared (compare with the equivalent result in fig. 2), in close agreement with the theoretical prediction.

4.2 Mutual Information measures

Mutual information (MI) between pairs of neighbor coefficients of image samples in the spatial domain, in the linear V1 image representation (wavelet domain),

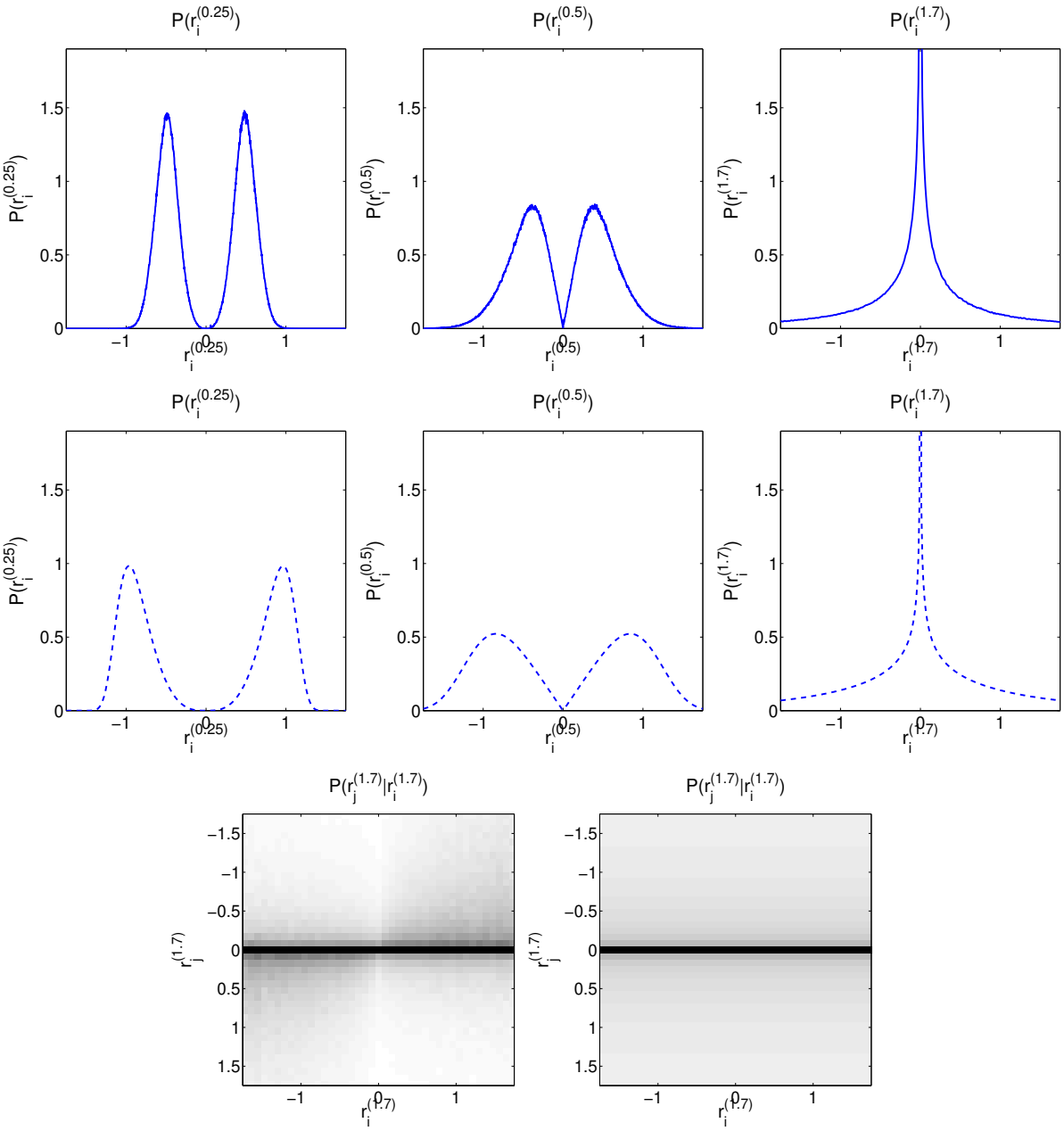


Fig. 4. *Marginal and conditional PDFs in the response domain. The first row shows the experimental marginal PDF of the responses for illustrative values of the exponent* $\gamma = 0.25, 0.5, 1.7$. *The second row shows the corresponding predictions according to the theoretical results in section 3. The third row shows the experimental (left) and the theoretical (right) conditional distributions for pairs of coefficients of the psychophysically optimal V1 model* ($\gamma = 1.7$).

and in the V1 non-linear representation were computed. The eventual reduction of MI values would point out the redundancy reduction along the visual pathway. In order to assess the magnitude of the achieved reductions we also include the results of two non-linear statistically-based techniques designed to give rise to independent components in images: Radial Gaussianization using L_2 norm as in [Lyu and Simoncelli, 2009], and L_p norm as in [Eichhorn et al., 2009]. These transforms start from a whitened linear representation of image vectors followed by an univariate (radial) non-linear transform tuned to obtain Gaussian distribution of the L_2 or L_p lengths of the vectors. In order to make the comparison easier, we used the same initial linear stage (wavelets) in those non-linear transforms. It is true that orthogonal wavelets may not be the best linear transform to achieve independence, but it is important to stress that (1) the selected linear stage is not critical for the final independence results obtained by using Radial Gaussianization techniques as pointed out in [Eichhorn et al., 2009], and (2) the aim of this work is not looking for the ultimate transform to achieve independence, but to show that the brain substantially reduces redundancy through the gain control non-linearity. The second non-linear stage in these illustrative Gaussianization techniques was performed by equalizing the L_2 and L_p lengths respectively, as done in [Lyu and Simoncelli, 2009]. In our simulations we used $p = 1.2$ in the L_p norm according to the results in [Eichhorn et al., 2009]. This is the optimal norm for ICA, while other linear representations are optimal for exponents in the range [1.2, 2]. As stated above, choosing different linear representations with norm exponents in the cited range gives rise to similar independence results [Eichhorn et al., 2009].

Appendix D gives the details on the used MI estimator and its errors: it shows that the errors are small compared to the MI differences presented in this section, thus ensuring the significance of the differences.

We performed two experiments. The first one tries to obtain a rough estimate of the global redundancy reduction ability of the linear (wavelet) and the non-linear (divisive normalization) stages of the V1 model. In this experiment we computed the MI among one coefficient and all the other coefficients (both in the spatial domain and in the local frequency domains, also including Radial Gaussianization using both L_2 and L_p).

The second experiment consists of a more accurate analysis of the different possible relations in the local frequency representations, $\mathbf{w}$, $\mathbf{r}$, and Radial Gaussianization using L_2 and L_p: (1) *intra-band*, measuring the MI of one coefficient with its $9 \times 9 - 1$ neighbors of the same subband, (2) *inter-orientation*, measuring the MI of one coefficient with its corresponding 5×5 spatial neighbors in a subband of the same scale but different orientation, and (3) *inter-scale*, measuring the MI of one coefficient in a coarser scale with its 2×2 sons in the corresponding finer scale.

Figure 5 shows representative results of the first experiment. In each MI computation 10^4 2D samples from the McGill database [Olmos and Kingdom, 2004] were used. The MI values in the spatial domain monotonically decrease with distance, as previously reported in [Lyu and Simoncelli, 2009]. The MI values among neighbors in the local frequency domains decrease as the distance in space, orientation and scale increases. The behavior is similar for coefficients of other scales and orientations.

As expected, the statistically tuned Gaussianization techniques obtain quite good independence results on the considered data set. Interestingly, the psychophysically tuned transform (that uses no statistical optimization at all) obtains very similar results in redundancy reduction. These results show that about 86% of the average MI in the spatial domain is reduced by the linear wavelet transform, while the non-linear psychophysical transform further reduces an additional 82% of the remaining MI in the linear wavelet domain. As a consequence, the non-linear V1 representation reduces about 98% of the average MI in the spatial domain, which is comparable to the reductions achieved by the statistically tuned Radial Gaussianization techniques using L_2 norm (99.2%) and L_p norm (99.5%).

Figure 6 shows a representative subset of the results of the second experiment: intra-band and inter-orientation MI values for the different orientations of the second scale, and inter-scale MI values for parents of the third scale and the corresponding sons of the second scale. Overlapping blocks of the different subbands were used to obtain more samples for a reliable MI estimation. The intra-scale, inter-orientation and inter-scale results were computed using $0.8 \cdot 10^6$, $1.3 \cdot 10^6$, and $0.7 \cdot 10^6$ 2D samples respectively.

Again, the results show that the statistically tuned Radial Gaussianization transforms substantially reduce the redundancy among the different neighbor coefficients with regard to the linear wavelet representation. The psychophysically optimal divisive normalized representation (second column) achieves very similar results. This means that the redundancy removal obtained through the psychophysical transform is significant. This is consistent with the removal of the bow-tie relations in the conditional probability plots (Figure 4).

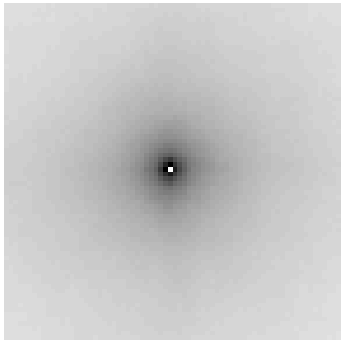
Moreover, it is interesting to note the similarity between the exponents to be used in the L_p norm in Eichhorn et al. [2009], and the psychophysical value for γ in the V1 normalization (which normalizes each wavelet coefficient by a sort of γ norm of its neighbors). In the first case, the value for better independence is in the range [1.2, 2]. In the psychophysically tuned V1 transform, $\gamma = 1.7$.

The fact that relatively more redundancy is reduced in the intra-band and inter-orientation cases may be due to the quantization of the masking kernel which was necessary for practical computational reasons (see comment in Appendix A). The quantization of the kernel in divisive normalization removes inter-scale interactions so the normalization is not as effective in that situation. Note that the optimal kernel in figure 1 does not reflect inter-scale interactions.

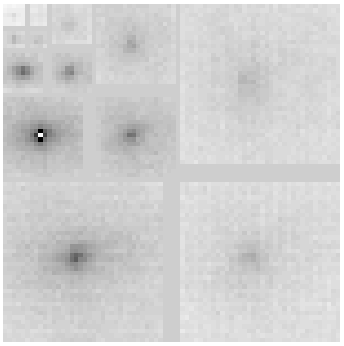
Summarizing, the agreement between the experimental and the predicted marginal and conditional PDFs of $\mathbf{r}$, and the substantial MI reduction with regard to the linear wavelet domain confirm the theoretical result in eq. 11: the psychophysically optimal divisive normalization is well matched to image statistics and approximately factorizes the PDF of natural images.

5 Conclusions

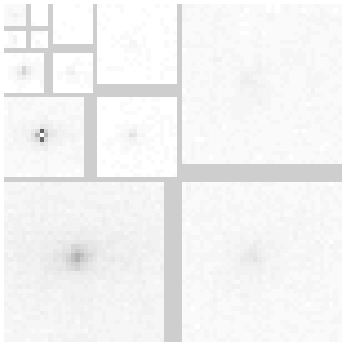
Here we showed that the non-linear stage of the standard V1 cortex model optimized to reproduce image quality psychophysics substantially increases the independence of the image coefficients obtained in the linear stage. Theoretical results (confirmed by experiments) show that the psychophysically tuned



$MI_{\mathbf{x}}$: *0.38*, [0.21, 1.70] bits



$MI_{\mathbf{w}}$: *0.052*, [10^{-3}, 0.27] bits



$MI_{\mathbf{r}}$: *0.009*, [10^{-4}, 0.15] bits

$MI_{RG\,L_2}$: *0.003*, [$6 \cdot 10^{-5}$, 0.06] bits

$MI_{RG\,L_p}$: *0.002*, [10^{-5}, 0.05] bits

Fig. 5. *MI results between one coefficient (the one in white) and its neighbors in the spatial domain (top) the linear V1 response, wavelet domain (middle left); the non-linear V1 response domain (middle right); the Radial Gaussianization using L_2 norm (bottom left); and the Radial Gaussianization using L_p norm (bottom right). The numbers in each case represent the average and the range of MI values found in bits. The top figure is scaled so that the black and white correspond to the maximum MI value in the spatial domain, 1.70 bits, and 0 bits respectively. All the other figures are scaled with regard to the maximum MI value in the wavelet domain: white and black correspond to the limits of the range [0, 0.27] bits.*

V1 model approximately factorizes a plausible joint PDF for natural images in

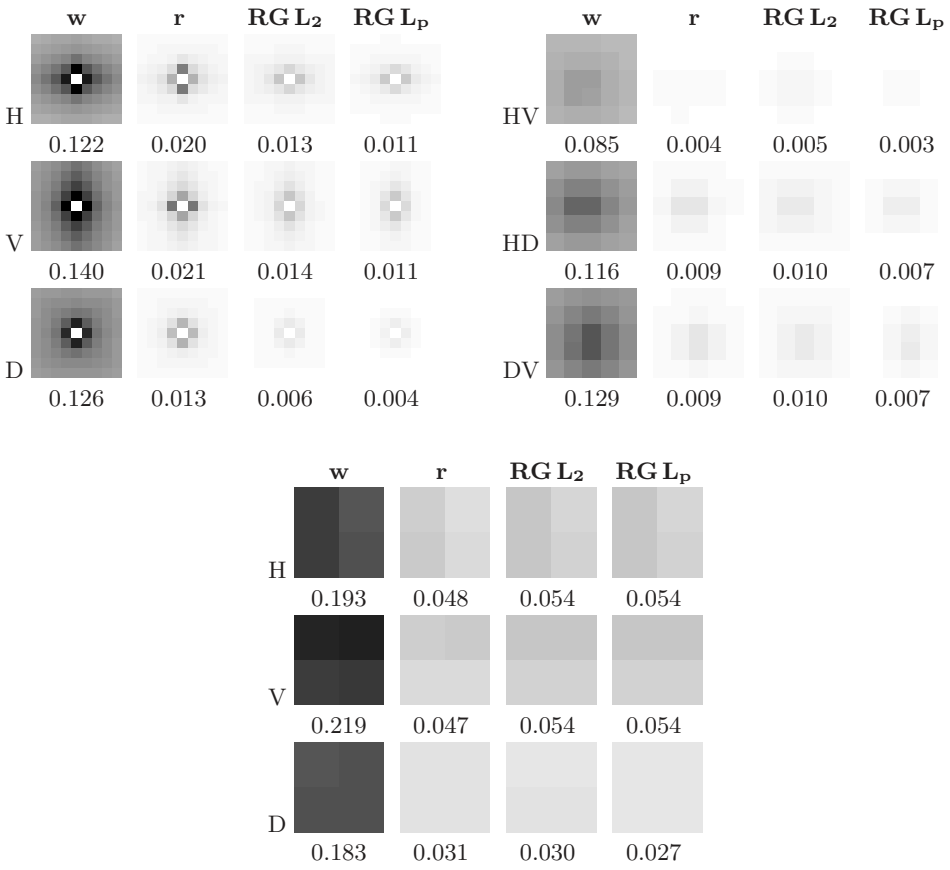


Fig. 6. *MI (in bits) between pairs of coefficients in the linear V1 representation (wavelet,* **w***) and in the non-linear V1 representation (normalized response,* **r***). The last two columns in each panel show the results of Radial Gaussianization techniques using L_2 norm and L_p norm respectively. The top left panel shows intra-band relations within 2nd scale subbands of different orientation. The top right panel shows inter-orientation relations for the 2nd scale coefficients. The bottom panel shows inter-scale relations of coefficients of the 3rd scale with their sons in the 2nd scale for different orientations. All images are scaled so that back and white correspond to the maximum MI value and 0 bits respectively. The numbers represent the average MI value (in bits) in each image.*

the wavelet domain: bow-tie dependencies are almost removed and redundancy among coefficients is substantially reduced.

Therefore, the results presented here confirm the efficient encoding hypothesis in a novel direction: *from perception to image statistics*. These results complement the standard approach to validate the hypothesis (e.g. *from image statistics to perception*) taken in [Olshausen and Field, 1996, Schwartz and Simoncelli, 2001, Kayser et al., 2003, Malo and Gutiérrez, 2006].

It is true that redundancy reduction is not the only goal in early visual processing [Barlow, 2001], but the results presented here suggest that this initial set of perceptual transforms performs a sort of non-linear independent components extraction.

Further work should address the issue of redundancy on the sign (or phase) of the wavelet coefficients. This information is not taken into account in the image model since the PDF is symmetric around the origin, and signs (and their eventual relations) are not modeled in any way. The proposed divisive normalization model does not take this issue into account either since it acts on the amplitude of the wavelet coefficients. A separate or complementary model for the signs of image coefficients is needed. Extensions of the perception model could be fitted by using the specific distortions in subjectively rated image databases consisting of phase alteration (e.g. fast fading or JPEG2000 transmission errors).

A Setting the V1 model parameters

Here we give the complete parametrization of the model and describe the optimization process.

The diagonal in $\mathbf{S}$, that accounts for contrast sensitivity, 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, $\mathbf{p}$:

$$S_{ii} = S_{(e,o,\mathbf{p})} = A_o \cdot \exp\left(-\frac{(4-e)^\theta}{s^\theta}\right) \tag{12}$$

where A_o is the maximum gain for the considered orientation, s controls the bandwidth of the frequency response, and θ determines the sharpness of the decay with spatial frequency. This parametrization describes the shape of the CSF in the wavelet domain [Malo, 1997]. As stated above, we will allow for different linear gains in the different chromatic channels YUV. In particular, we will allow for different gains (A_{oY}, $A_{oU} = A_{oV}$) and different bandwidths (s_Y, $s_U = s_V$).

We use the Gaussian interaction kernel proposed in [Watson and Solomon, 1997], which has been successfully used in image processing applications in the block-DCT domain [Malo et al., 2006, Gutiérrez et al., 2006, Camps et al., 2008]. In the wavelet domain this reduces to:

$$\begin{aligned} H_{ik} &= H_{(e,o,\mathbf{p}),(e',o',\mathbf{p}')} = \\ &= K \cdot \exp\left(-\left(\frac{(e-e')^2}{\sigma_e^2} + \frac{(o-o')^2}{\sigma_o^2} + \frac{(\mathbf{p}-\mathbf{p}')^2}{\sigma_p^2}\right)\right) \end{aligned} \tag{13}$$

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, σ_e is the kernel width in the scale dimension, σ_o is the kernel width in the orientation dimension, and σ_p is the kernel width in the spatial dimension, and K is a normalization factor to ensure $\sum_k H_{ik} = 1$. In order to reduce the storage requirements of the kernels, we quantized them to obtain sparse matrices. This implies neglecting small interactions among coefficients.

In our implementation of the model we set an initial guess of the profile of the saturation 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 [Schwartz and Simoncelli, 2001]. This profile β_i (computed from 100 images of a calibrated image data base [Olmos and Kingdom, 2004]) was scaled by a constant b to be set in the psychophysical 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 $\Delta\mathbf{r}$ [Laparra et al., 2010]. Non-quadratic pooling norms have been reported [Ahumada, 1993, Watson and Solomon, 1997, Watson and Malo, 2002]. Moreover, different summation exponents, for the pooling across spatial position, q_p, and frequency, q_f, may be used [Laparra et al., 2010]. The parameters of the model (including the pooling exponents) can be optimized to maximize the correlation among the predicted distortion and the perceived distortion on a subjectively rated database. The model was tuned to reproduce the subjective quality data of three images of the LIVE database [Sheikh et al., 2006], which includes 5 kinds of distortion (i.e. it was optimized for a total of 83 distorted images).

Assuming the same behavior in the horizontal and vertical directions ($o = 1, 3$), and assuming that the oblique effect in the frequency sensitivity [Watson and Ramirez, 2000] 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:

$$\Omega \equiv \{ A_{1Y}, d, A_{1UV}, s_Y, s_{UV}, \theta, \gamma, b, \sigma_e, \sigma_o, \sigma_{\mathbf{p}}, q_p, q_f \}. \tag{14}$$

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 $\Omega_3 \equiv \{ q_p, q_f \}$.

The explored ranges for the parameters and the optimal values found are shown in Table 1. The optimal summation strategy consist of pooling first over the frequency dimensions and then over the spatial dimensions.

The parameters found are consistent with previously reported results in psychophysical and physiological literature. First, the acrhomatic linear gain has bigger peak sensitivity than the chromatic linear gain ($A_Y > A_{UV}$), and its bandwidth is also bigger ($s_Y > s_{UV}$), which is consistent with the experimental results on achromatic and chromatic CSFs [Mullen, 1985]. The reduction in sensitivity in oblique directions ($d < 1$) is also consistent with models based on low-level psychophysics [Watson and Ramirez, 2000, Wuerger et al., 2002]. As a

result, the basic trends of the CSF can be reproduced with the model (see CSF reproduction compared to the one of the Standard Spatial Observer in appendix B). The value of the excitatory and inhibitory exponent, $\gamma = 1.7$, is close to the values reported in the literature: in [Watson and Solomon, 1997], they use values in the range [2, 2.3] to fit contrast incremental threshold data. In [Carandini et al., 1997] values in the range [1,3] are considered for V1 cells, and they finally chose a quadratic exponent in the simulations. Note that in [Watson and Solomon, 1997, Carandini et al., 1997] the models are slightly different since they allow for different exponents in the numerator (excitation) and the denominator (inhibition) of the normalization. The width of the pooling kernel is also consistent with previously reported results: in [Watson and Solomon, 1997] they found $\sigma_p \approx 0.5$ degrees, while in [Watson and Malo, 2002] a smaller pooling area was found, $\sigma_p \approx 0.1$ degrees (twice the spatial width of the CSF filter). In our case, when fitting image distortion results, we found $\sigma_p = 0.25$ degrees. Pooling in scale was not considered, $\sigma_e = 0$, in [Watson and Solomon, 1997], which is consistent with the small value we found in our experiments. In fact, in our case, the effect is so small that these inter scale interactions are neglected when simplifying the kernel by quantization (see figure 1). On the contrary, pooling in orientation is wide: we found $\sigma_o = 3$, that is there is a strong interaction between subbands that are 90 degrees apart (see figure 1), which is consistent with the results in [Watson and Solomon, 1997] that found $\sigma_o \approx 85$ degrees. Finally the summation exponents for the distance computation ($q_p = 2.2$ and $q_f = 4.5$, table 1) are consistent with the values found in [Watson and Malo, 2002], $q_p = 2.9$, and with the summation exponent found in [Watson and Solomon, 1997], about 5.1. However, note that the redundancy reduction properties of the representation do not depend on these summation exponents.

Parameter	**Meaning**	**Range**	**Optimal**	**Correlation**
A_Y	Amplitude of achrom. CSF	$30, \ldots, 60$	**40**	
s_Y	Bandwidth of achrom. CSF	$0.25, \ldots, 3$	**1.5**	
θ	Sharpness of CSFs	$2, \ldots, 8$	**6**	
γ	Excit.& inhibit. exp.	$0.5, \ldots, 3$	**1.7**	
b	Scaling of regularizat.	$0.5, \ldots, 8$	**2**	
σ_e	Scale width of masking	$0.15, \ldots, 3$	**0.25**	
σ_o	Orientat. width of masking	$0.15, \ldots, 3$	**3**	
$\sigma_{\mathbf{p}}$	Spat. width of masking	$0.03, \ldots, 0.4$	**0.25** (in deg)	$\rho_p = 0.916$
A_{UV}	Amplitude of chrom. CSF	$30, \ldots, 40$	**35**	
s_{UV}	Bandwidth of chrom. CSF	$0.25, \ldots, 1.5$	**0.5**	
d	Oblique factor	$0.6, \ldots, 1.4$	**0.8**	$\rho_p = 0.922$
q_p	Spatial pooling exp.	$0.5, \ldots, 6$	**2.2**	
q_f	Freq. pooling exp.	$0.5, \ldots, 6$	**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.*

B Reproducing low-level and high-level psychophysics

In this section we show that the model optimized to account for (high-level) image quality opinion also accounts for the fundamental trends of (low-level) threshold psychophysics.

Figures 7-9 show the results of three experiments: (1) reproduction of subjectively rated distortion, (2) reproduction of frequency-dependent threshold contrast sensitivity, and (3) reproduction of contrast masking non-linearities.

In the first experiment, the generalization ability and robustness of the model to account for a wide variety of suprathreshold distortions is assessed by checking its performance on a more general image quality database (with more images and distortions of different nature). Here we applied the model (optimized for 83 images) to predict distortions on the whole LIVE database (779 distorted images), plus on the whole TID database [Ponomarenko et al., 2008]. The extension to the TID database is challenging since it not only contains different images, but more importantly, it includes 12 kinds of distortion not included in the LIVE database. The model was finally applied to 2479 distorted images. The performance of the proposed model (figure 7.a) can be compared to the performance of the state-of-the-art Visual Information Fidelity index (VIF) [Sheikh and Bovik, 2006] of the same authors as the LIVE database (figure 7.b). Note that the VIF metric fails to account for some of the distortions in the TID database (represented by different symbols/colors in the plots) while the proposed V1 image representation model obtains significantly better correlation when considering a wide range of distortions (see the Pearson and Spearman correlation coefficients at the plots). More details on the performance of the proposed model as image quality metric are given in [Laparra et al., 2010]. The Matlab implementation of the metric is available on-line[5].

The second experiment shows how the model accounts for the threshold frequency sensitivity. 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 CSF can be simulated by computing the above distances between sinusoids with fixed contrast, but different frequencies and orientations, and a uniform gray background. Figures 8.a and 8.b compare 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 [Watson and Ramirez, 2000]. Note that the model approximately reproduces the band pass behavior, the overall bandwidth, and the oblique effect.

The third experiment simulates contrast masking results. In order to do so, the contrast of a Gabor patch is increased on top of different backgrounds (sinusoids with different contrasts and orientations). As widely known [Watson and Solomon, 1997, Foley, 1994], 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 when the the background has the same orientation as the target. Figures 9.a and 9.b show the response curves of the model to vertical targets for the different background sets: vertical (left) and horizontal (right). The model response to the target is a saturating non-linearity when the target is shown on top of no background

[5] `http://www.uv.es/vista/vistavalencia/div_norm_metric/`

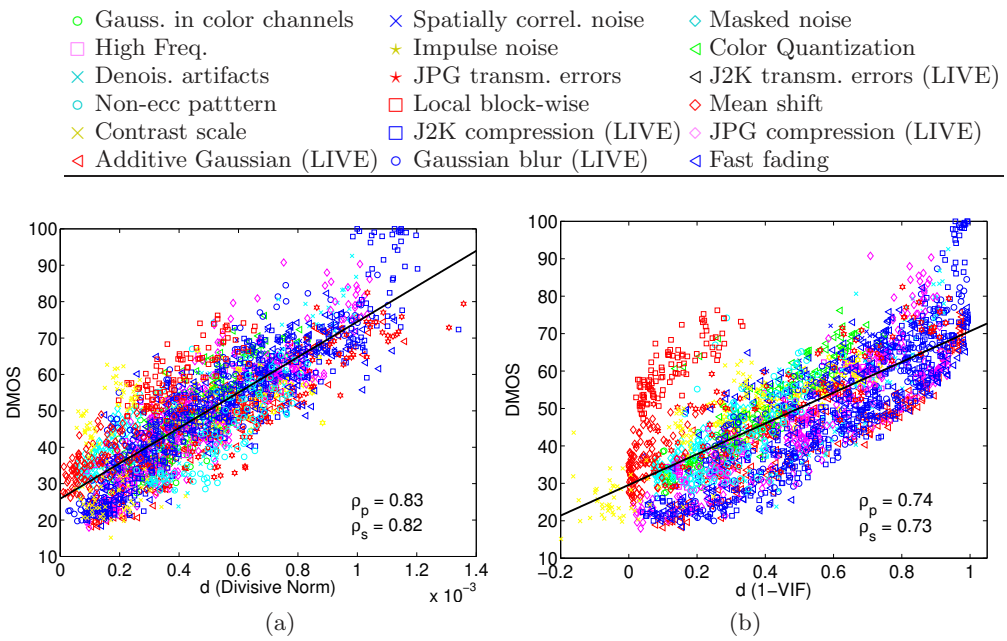


Fig. 7. *Reproduction of high-level perception results. The figures show the correlation among the predicted distortion, d, and the observers opinion, DMOS, for the distance in the proposed V1 image representation (a), and the state-of-the-art VIF metric (b). The different symbols in the plot and legend represent images with distortions of different nature. For details on the different distortions see [Sheikh et al., 2006, Ponomarenko et al., 2008].*

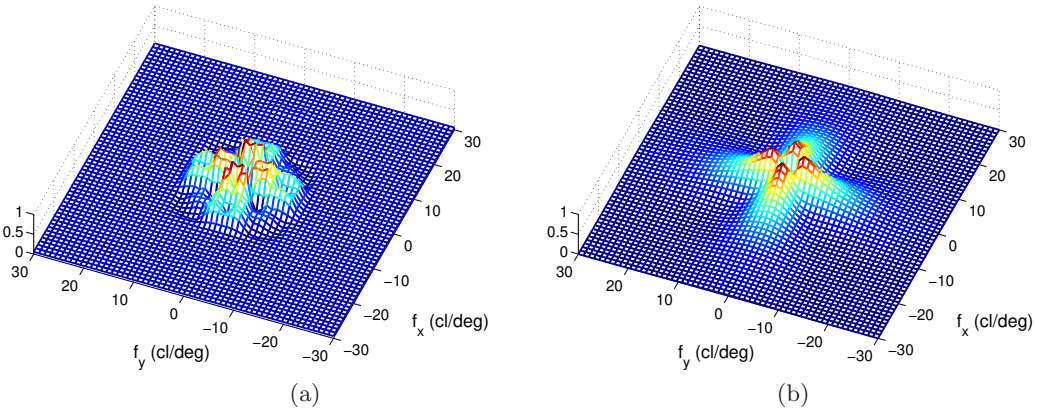


Fig. 8. *Reproduction of (low-level) frequency-dependent sensitivity. In the plots, the achromatic CSF as predicted by the proposed V1 model (a) is compared to the Standard Spatial Observer CSF (b).*

(auto-masking). The model predicts the reduction of the response 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 (vertical target and vertical background). Therefore, the behavior of the model with the proposed parameters is compatible with the low-level behavior of human observers reported elsewhere [Watson and Solomon, 1997].

Figures 9.c and 9.d show contrast incremental thresholds ΔC for non-zero mask contrast as a function of the test contrast. These plots have been obtained from the previous response curves (with $C_{mask} = 0.1$) looking for the amount of contrast deviation needed to obtain a constant increment in the response (or distance). The left plot corresponds to target and background of the same frequency and orientation while the right plot corresponds to the orthogonal orientation situation. In both cases the thresholds increase with contrast (as expected from saturating responses). However, when target and background have the same orientation the sensitivity is reduced (thresholds increase faster). Figures 9.e and 9.f show equivalent experimental results by Foley [Foley, 1994] explicitly reproduced from [Watson and Solomon, 1997], which display the same behavior.

To summarize, the results in this section show that the divisive normalization model optimized to reproduce high level distortions (such as those in the LIVE database) can simultaneously reproduce the basic features of low-level psychophysics (e.g. frequency sensitivity and contrast masking), while being robust enough to account for a wider range of suprathreshold distortion data (TID database).

C Normalization constant of the proposed PDF

The normalization constant, Z, in eq. 3 is:

$$Z = \int_{\mathbf{w}'} \frac{1}{|\Sigma(\mathbf{w}')|^{1/2}} \, e^{-\frac{1}{2}\,\mathbf{w}'^{T}\cdot\Sigma(\mathbf{w}')^{-1}\cdot\mathbf{w}'} \, d\mathbf{w}' \tag{15}$$

The proposed PDF integrates to 1 if Z is bounded. Intuitively, this is the case in practical situations, since the linearly weighted wavelet domain (the range of possible values of $\mathbf{w}'$) is limited for images of finite energy and extent, and the diagonal matrix, $\Sigma(\mathbf{w}')$, is not singular for realistic (strictly positive) values of the parameters (see eq. 4), so the integrand is bounded (and strongly decays with $|\mathbf{w}'|$).

In addition to the above physical constraints, in this appendix we give an approximated bound for Z (reasoning in the linear domain, $\mathbf{w}'$), and we derive an explicit value for Z (reasoning in the non-linear domain, $\mathbf{r}$).

Note that Z is just a scalar that does not depend on $\mathbf{w}'$ so the factorization result in section 3.2 does not depend on the particular value of this constant.

The bound for Z is based on the Laplace method to approximate integrals of exponential functions [MacKay, 2002]. The idea is approximating the exponent by its expansion up to 2nd order around the maximum, $\mathbf{w}'_{\mathbf{0}}$. In our case, the peak of the exponent, $-1/2\ \mathbf{w}'^{T}\cdot\Sigma(\mathbf{w}')^{-1}\cdot\mathbf{w}' = -f(\mathbf{w}')$, is at the origin, $\mathbf{w}'_{\mathbf{0}} = 0$:

$$\begin{aligned}\int_{\mathbf{w}'} e^{-f(\mathbf{w}')}\, d\mathbf{w}' &= \int_{\mathbf{w}'} e^{-(f(0)+\frac{1}{2}\mathbf{w}'^{T}\cdot\nabla^2 f(0)\cdot\mathbf{w}'+\cdots)}\, d\mathbf{w}' \\ &\approx \int_{\mathbf{w}'} e^{-\frac{1}{2}\mathbf{w}'^{T}\cdot\nabla^2 f(0)\cdot\mathbf{w}'}\, d\mathbf{w}' = (2\pi)^{N/2}\,|\nabla^2 f(0)|^{-1/2}\end{aligned} \tag{16}$$

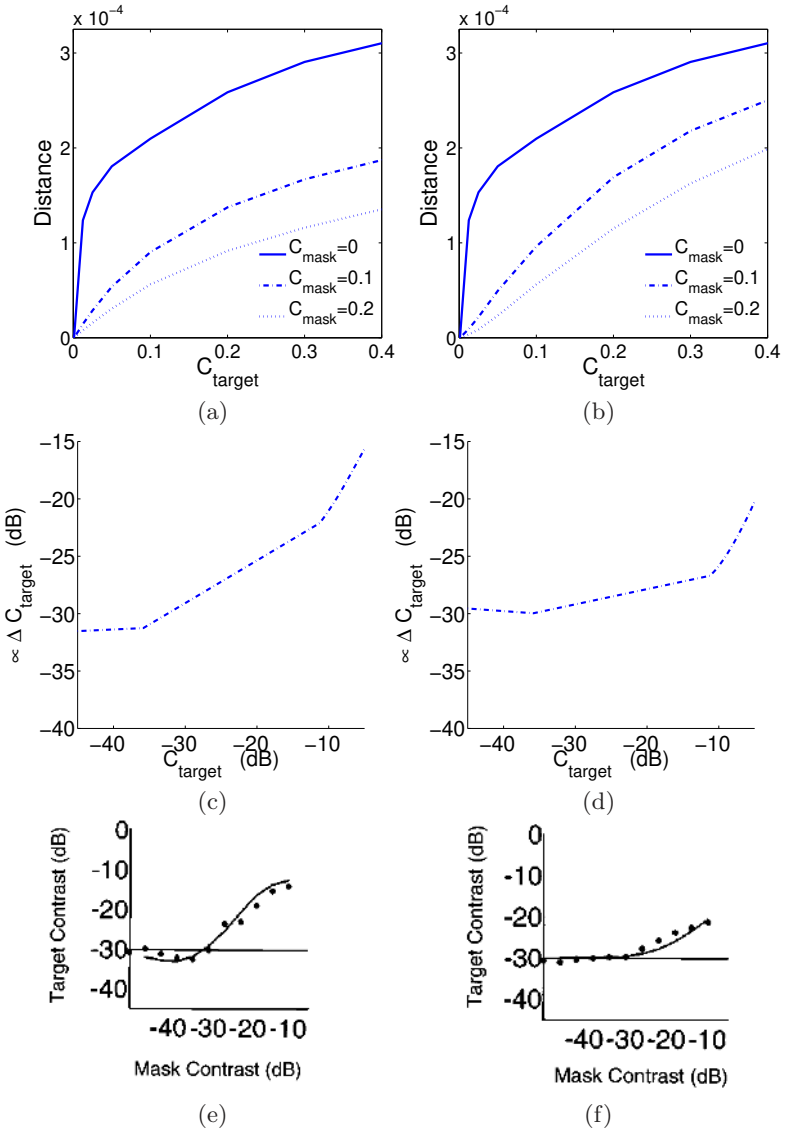


Fig. 9. *Reproduction of (low-level) masking non-linearities and contrast incremental thresholds. Top row non-linear response: (a) response to Gabor targets of increasing contrast seen on top of sinusoids of the same frequency and orientation, and (b) equivalent responses on top of orthogonal sinusoids. Middle row: contrast incremental thresholds ΔC as a function of the test contrast when mask and test have the same orientation (c) and orthogonal orientations (d). Bottom row: equivalent ΔC psychophysical data by Foley [Foley, 1994], as reported in [Watson and Solomon, 1997]. In the middle and bottom rows contrast is expressed in dB: $C_{dB} = 20 \log_{10} C$.*

Computing the matrix of second derivatives of $f(\mathbf{w}')$, it can be seen that $\nabla^2 f(0) = \Sigma(0)^{-1}$. Therefore:

$$\int_{\mathbf{w}'} \frac{1}{|\Sigma(0)|^{1/2}} e^{-\frac{1}{2}\mathbf{w}'^T \cdot \Sigma(\mathbf{w}')^{-1} \cdot \mathbf{w}'} \, d\mathbf{w}' \approx (2\pi)^{N/2} \tag{17}$$

Taking into account eqs. 15 and 17, and the fact that $|\Sigma(\mathbf{w}')| > |\Sigma(0)|$ for $\mathbf{w}' \neq 0$, it follows:

$$Z < \int_{\mathbf{w}'} \frac{1}{|\Sigma(0)|^{1/2}} e^{-\frac{1}{2}\mathbf{w}'^T \cdot \Sigma(\mathbf{w}')^{-1} \cdot \mathbf{w}'} d\mathbf{w}' \approx (2\pi)^{N/2} \tag{18}$$

More accurate estimations of the bound can be obtained by considering higher order terms in the expansion using the generalized Laplace method [Fog, 2008].

An explicit value for Z can be easily obtained in the non-linearly transformed domain $\mathbf{r}$. Under the assumptions detailed in section 3.2, the joint PDF after the divisive normalization of eq. 6 is given by the product of the marginals,

$$P_{r_i}(r_i) = \frac{1}{\gamma\, Z^{1/N}} \, |r_i|^{\frac{1}{\gamma}-1} \, e^{-\frac{|r_i|^{2/\gamma}}{2}} \tag{19}$$

therefore, an alternative equation for the normalization constant is,

$$Z = \left(\frac{2}{\gamma} \int_0^\infty r_i^{\frac{1}{\gamma}-1} \, e^{-\frac{r_i^{2/\gamma}}{2}} \, dr_i \right)^N \tag{20}$$

Using the change of variable, $u = r_i^{\frac{1}{\gamma}}$, and $du = \frac{1}{\gamma} r_i^{\frac{1}{\gamma}-1} dr_i$, it follows,

$$Z = \left(2 \int_0^\infty e^{-\frac{u^2}{2}} \, du \right)^N = (2\pi)^{N/2} \tag{21}$$

D Measuring mutual information

Mutual information (MI) between two random variables, $MI(v_1, v_2)$, is defined as the difference between the sum of marginal entropies and the joint entropy [Cover and Tomas, 1991]:

$$MI(v_1, v_2) = h(v_1) + h(v_2) - h(v_1, v_2) \tag{22}$$

Since MI is invariant under point-wise transforms [Cover and Tomas, 1991], our MI estimator first equalizes the marginal PDF of each coefficient to obtain uniform densities in the range $[0, 1]$. Then, the joint entropy is computed by using the 2D histogram and the Miller-Madow correction [Miller, 1955]. In our implementation, the total number of bins in the 2D histogram was set to be the square root of the number of available samples. In our case, the marginal entropies are zero due to the uniformization step. Therefore the MI is equal to minus the joint entropy.

In order to assess the accuracy of the above estimator we tested it for two particular densities of known MI: (1) Gaussian densities, whose MI can be computed in closed-form [Cover and Tomas, 1991], and (2) the image model in the wavelet domain, Eq. 3, whose MI can be obtained by numerical integration of the joint PDF.

In Table 2 we show the mean and the standard deviation of the percentage of error for 2D PDFs of different MI as a function of the number of samples used in the estimation. The explored range of MI values is $[0.01, 0.32]$ bits, and the number of samples is in the range $[10^4, 10^6]$. These error percentages have been obtained with 100 different realizations for each sample size.

These results ensure that the estimation error is always below the MI differences shown in section 4.

Gaussian PDFs						
	Number of samples ($\times 10^4$)					
MI	1	2.5	6.3	15.8	39.8	100
0.01	14 ± 18	9 ± 10	5 ± 6	3 ± 3	2 ± 2	1.3 ± 1.5
0.04	7 ± 7	4 ± 4	4 ± 3	2 ± 2	2.1 ± 1.3	1.6 ± 0.7
0.09	8 ± 5	5 ± 3	4 ± 2	2.8 ± 1.3	2.4 ± 0.8	1.7 ± 0.5
0.14	8 ± 4	5 ± 2	4 ± 1.5	3.3 ± 1.0	2.5 ± 0.6	1.9 ± 0.4
0.19	8 ± 3	6 ± 2	5 ± 1.5	3.5 ± 0.9	2.6 ± 0.6	1.9 ± 0.3
0.24	8 ± 3	6 ± 2	5 ± 1.2	3.3 ± 0.7	2.7 ± 0.5	2.0 ± 0.3
0.28	8 ± 2	6 ± 1.8	5 ± 0.9	3.7 ± 0.7	2.6 ± 0.4	2.0 ± 0.3
0.31	8 ± 2	6 ± 1.6	5 ± 1.1	3.7 ± 0.6	2.8 ± 0.4	2.0 ± 0.2
0.32	9 ± 2	6 ± 1.6	5 ± 1.0	3.7 ± 0.6	2.8 ± 0.4	2.1 ± 0.2
Image PDF model in the wavelet domain (Section 3.1)						
	Number of samples ($\times 10^4$)					
MI	1	2.5	6.3	15.8	39.8	100
0.21	8 ± 3	5 ± 2	2.9 ± 1.4	1.5 ± 0.6	1.5 ± 0.6	1.0 ± 0.3

Table 2. *Relative error (in %) of the mutual information estimator on Gaussian densities and on the proposed image model in the wavelet domain.*

Bibliography

A.J. Ahumada. Computational image quality metrics: A review. In J. Morreale, editor, *Intl. Symp. Dig. of Tech. Papers, Sta. Ana CA*, volume 25 of *Proceedings of the SID*, pages 305–308, 1993.

H.B. Barlow. Possible principles underlying the transformation of sensory messages. In WA Rosenblith, editor, *Sensory Communication*, pages 217–234. MIT Press, Cambridge, MA, 1961.

H.B. Barlow. Redundancy reduction revisited. *Network: Computation in Neural Systems*, 12:241–253, 2001.

A.J. Bell and T.J. Sejnowski. The 'independent components' of natural scenes are edge filters. *Vision Research*, 37(23):3327–3338, 1997.

R.W. Buccigrossi and E.P. Simoncelli. Image compression via joint statistical characterization in the wavelet domain. *IEEE Tr. Im. Proc.*, 8(12):1688–1701, 1999.

F.W. Campbell and J.G. Robson. Application of Fourier analysis to the visibility of gratings. *Journal of Physiology*, 197:551–566, 1968.

G. Camps, J. Gutiérrez, G. Gómez, and J. Malo. On the suitable domain for SVM training in image coding. *JMLR*, 9:49–66, 2008.

M. Carandini and D. Heeger. Summation and division by neurons in visual cortex. *Science*, 264(5163):1333–6, 1994.

M. Carandini, D. J. Heeger, and J. A. Movshon. Linearity and normalization in simple cells of the macaque primary visual cortex. *J Neurosci*, 17(21): 8621–8644, November 1997. ISSN 0270-6474.

T.M. Cover and J.A. Tomas. *Elements of Information Theory*. John Wiley & Sons, New York, 1991.

J.G. Daugman. Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research*, 20:847–856, 1980.

Jan Eichhorn, Fabian Sinz, and Matthias Bethge. Natural image coding in v1: How much use is orientation selectivity? *PLoS Comput Biol*, 5(4), April 2009.

A. Fog. Calculation Methods for Wallenius Noncentral Hypergeometric Distribution. *Comm. Stat. Simul. Comp.*, 37:258–273, 2008.

P. Foldiak. Adaptive network for optimal linear feature extraction. In *Neural Networks, 1989. IJCNN., International Joint Conference on*, volume 1, pages 401–405 vol.1, Jun 1989.

J.M. Foley. Human luminance pattern mechanisms: Masking experiments require a new model. *Journal of the Optical Society of America A*, 11(6):1710–1719, 1994.

J. Gutiérrez, F. Ferri, and J. Malo. Regularization operators for natural images based on nonlinear perception models. *IEEE Tr. Im. Proc.*, 15(1):189–200, 2006.

D. J. Heeger. Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9:181–198, 1992.

A. Hyvärinen. Sparse code shrinkage: Denoising of nongaussian data by ML estimation. *Neur. Comp.*, pages 1739–1768, 1999.

C. Kayser, K.P. Körding, and P. König. Learning the nonlinearity of neurons from natural visual stimuli. *Neural Computation*, 15:1751–1759, 2003.

V. Laparra, J. Muñoz Marí, and J. Malo. Divisive normalization image quality metric revisited. *JOSA A*, 27(4):852–864, 2010.

R Linsker. From basic network principles to neural architecture: emergence of orientation-selective cells. *Proceedings of the National Academy of Sciences of the United States of America*, 83(21):8390–8394, 1986.

S Lyu and E P Simoncelli. Nonlinear extraction of 'independent components' of natural images using radial Gaussianization. *Neural Computation*, 21(6): 1485–1519, May 2009.

David J. C. MacKay. *Information Theory, Inference & Learning Algorithms*. Cambridge University Press, 1st edition, June 2002.

J. Malo. Characterization of HVS threshold performance by a weighting function in the Gabor domain. *J. Mod. Opt.*, 44(1):127–148, 1997.

J. Malo and J. Gutiérrez. V1 non-linear properties emerge from local-to-global non-linear ICA. *Network: Computation in Neural Systems*, 17:85–102, 2006.

J. Malo, I. Epifanio, R. Navarro, and E. Simoncelli. Non-linear image representation for efficient perceptual coding. *IEEE Transactions on Image Processing*, 15(1):68–80, 2006.

E. Martinez-Uriegas. Color detection and color contrast discrimination thresholds. In *Proc. OSA Meeting*, page 81, 1997.

G. Miller. Note on the bias of information estimates. *Information Theory in Psychology*, II-b:95–100, June 1955.

K. T. Mullen. The CSF of human colour vision to red-green and yellow-blue chromatic gratings. *J. Physiol.*, 359:381–400, 1985.

A. Olmos and F. A. A. Kingdom. McGill calibrated colour image database. *http://tabby.vision.mcgill.ca.*, 2004.

B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.

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*, pages 403–408, Oct. 2008.

W.K. Pratt. *Digital Image Processing*, chapter 3: *Photometry and Colorimetry*. John Wiley & Sons, New York, 1991.

R. P. Rao and D. H. Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, 2(1):79–87, January 1999.

T. D Sanger. Analysis of the two-dimensional receptive fields learned by the generalized hebbian algorithm in response to random input. *Biological cybernetics*, 63:221–228, 1990.

T.D. Sanger. Optimal unsupervised learning in a single-layer network. *Neural Networks*, 2:459–473, 1989.

O. Schwartz and E.P. Simoncelli. Natural signal statistics and sensory gain control. *Nat. Neurosci.*, 4(8):819–825, 2001.

H.R. Sheikh and A.C. Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2):430–444, Feb 2006.

H.R. Sheikh, M.F. Sabir, and A.C. Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on Image Processing*, 15(11):3440–3451, November 2006.

E. Simoncelli. Bayesian denoising of visual images in the wavelet domain. In *Bayesian Inference in Wavelet Based Models*, pages 291–308. Springer-Verlag, New York, 1999.

E.P. Simoncelli. Statistical models for images: Compression, restoration and synthesis. In *31st Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA.*, 1997.

E.P. Simoncelli. Vision and the statistics of the visual environment. *Current Opinion in Neurobiology*, 13:144–149, 2003.

E.P. Simoncelli and E.H. Adelson. *Subband Image Coding*, chapter Subband Transforms, pages 143–192. Kluwer Academic Publishers, Norwell, MA, 1990.

H. Stark and J. Woods. *Probability, Random Processes, and Estimation Theory for Engineers*. Prentice Hall, NJ, 1994.

J. H Van Hateren. A theory of maximizing sensory information. *Biological Cybernetics*, 68(1):23–29, 1992.

J. H Van Hateren. Spatiotemporal contrast sensitivity of early vision. *Vision Research*, 33(2):257 – 267, 1993.

J.H. van Hateren and A. van der Schaaf. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc.R.Soc.Lond. B*, 265:359–366, 1998.

A.B. Watson. Detection and recognition of simple spatial forms. In O.J. Braddick and A.C. Sleigh, editors, *Physical and Biological Processing of Images*, volume 11 of *Springer Series on Information Sciences*, pages 100–114, Berlin, 1983. Springer Verlag.

A.B. Watson. Efficiency of a model human image code. *Journal of Optical Society of America A*, 4(12):2401–2417, 1987.

A.B. Watson and J. Malo. Video quality measures based on the standard spatial observer. *Proc. IEEE ICIP*, 3:41–44, 2002.

A.B. Watson and C.V. Ramirez. A Standard Observer for Spatial Vision. *Investig. Opht. and Vis. Sci.*, 41(4):S713, 2000.

A.B. Watson and J.A. Solomon. A model of visual contrast gain control and pattern masking. *JOSA A*, 14:2379–2391, 1997.

S. Wuerger, A.B. Watson, and A.J. Ahumada. Toward a standard observer for spatio-chromatic detection. In *Proc SPIE*, volume 4662, 2002.