Differential inversion of V1 non-linearities

I. Epifanio, J. Malo*

Dept. de Matemàtiques, Universitat Jaume I
Campus del Riu Sec, 12071 Castelló, Spain

(*) Dept. d'Òptica, Universitat de Valncia
Dr. Moliner 50, 46100 Burjassot, Valncia, Spain

Jesus.Malo@uv.es

November 2, 2004

Abstract

The current models of early visual processing in the human brain consist of a linear wavelet-like transform of the image followed by a compressive non-linear mapping that accounts for the masking and adaptation phenomena. The use of this set of transforms in both directions, direct and inverse, has a variety of applications in neuroscience and image processing.

However, the non-linear part of the model is particularly difficult to invert.

In this work we analyze rigorously a differential method to invert the non-linear part of the current V1 model, originally proposed in [1]. The existence and uniqueness of the solution, the conditions to ensure the invertibility and the convergence rates with regard to conventional gradient-based methods to solve non-linear equations are analyzed in the context of the divisive normalization model of the non-linearity. However, the most appealing feature of the differential approach considered here is that it does not depend on the particular functional form of the non-linearity but on its local (differential) properties. Therefore it can be applied either to recently proposed models which are non-parametric or to more complex parametric models that may appear in the future.

1 Introduction

The standard model of human low-level image analysis has two basic stages [2, 3, 4],

$$a \xrightarrow{T} c \xrightarrow{R} r \tag{1}$$

in which the input image, a, (array of luminances in the *spatial domain*) is first transformed into a vector, $c = T \cdot a$ (with components c_f, $f = 1 \dots M$), in a local frequency domain (the *transform domain*) using a linear filter bank, T [5, 6, 7]. Then, a set of mechanisms responds to each coefficient of the transformed signal giving an output, $r = R(c)$, which is the image representation in the *response domain* [8, 9, 10, 4, 2].

The linear transform should be a local-frequency representation as is commonly used in transform coding (e.g., block-DCT or a wavelet filterbank). It is well known that the perception of errors in coefficients of local frequency or wavelet representations is not independent, a phenomenon known in the perceptual literature as *masking* [2]. Specifically, the presence of large coefficients can reduce the visibility of errors in coefficients that are nearby in position, orientation and scale.

The linear coefficients may be modified so as to more accurately represent perceptual distances by *normalizing* (dividing) each coefficient by a gain signal obtained from a combination of adjacent coefficients [11, 12, 2, 1, 13]. This is consistent with recent models of neurons in visual cortex, in which primarily linear neural responses are modulated by a gain signal computed from a combination of other neural responses [14, 8, 9, 10, 4]. So, the second stage, the divisive normalization stage describes the gain control mechanisms normalizing the energy of each linear coefficient by a linear combination of its neighbors in space, orientation and scale:

$$r_i = \frac{\mathrm{sgn}(c_i) \quad |c_i|^\gamma}{\beta_i + \sum_j h_{ij}|c_j|^\gamma}. \tag{2}$$

Particularly, this response model, the energy-normalization model, is that of Refs. [4, 15, 2, 16]. Each coefficient c_i is first rectified and exponentiated. Each of the resulting values are then divided by a weighted sum of the others, where h_{ij} is the set of weights that specify the *interactions* between all the coefficients of the vector c and coefficient c_i. The sign (or phase, in the case of a complex-valued transform) of each normalized coefficient, is inherited from the sign of the corresponding linear coefficient, $\mathrm{sgn}(c_i)$.

Other models could be also considered, both more complex parametric models that may appear in the future and models which are non-parametric, as those recently proposed [17].

In general, these non-linear models share a common problem: the difficulty in inverting them, which is necessary in variety of applications in neuroscience and image processing, such as image compression.

In this work we analyze in detail a differential method to invert the non-linear part of the current V1 model (eq. 2), originally proposed in [1]. In section 2 the method is presented in detail. The existence and uniqueness of the solution and the conditions to ensure the invertibility are also studied. The method is analyzed in the context of the previous divisive normalization model and it is compared with conventional gradient-based methods to solve non-linear equations in section 3. The paper finishes in section 4 with some concluding remarks.

2 Differential inversion

The idea here is putting the inversion as an *initial value problem* [18]: given a known stimulus-response pair, (c_0, r_0), the stimulus -the transform-, c, that gives rise to a particular response, r, can be obtained propagating the *initial value* (c_0, r_0) using the differential behavior,

$$dc = \nabla R^{-1}(r') \cdot dr' \tag{3}$$

i.e., solving the integral:

$$c = c_0 + \int_{r_0}^{r} \nabla R^{-1}(r') \cdot dr' \tag{4}$$

where ∇R stands for the Jacobian matrix of the transformation R.

This *initial value problem* can be solved because for each particular point in the integration path, the *unknown* jacobian, $\nabla R^{-1}(r')$, can be computed using $\nabla R^{-1}(r') = \nabla R(c')^{-1}$.
Even though an accurate Runge-Kutta integration of eqs. 3 and 4 is needed, it is worth to take a look at a naive Euler integration of eq. 3 in M steps:

$$c_{(n)} = c_{(n-1)} + \nabla R(c_{(n-1)})^{-1} \cdot dr \tag{5}$$

where the increments in r are given by $dr = (r - r_0)/M$. We quote this result to stress the similarities and differences with the standard gradient-based search procedures.

2.1 General invertibility condition

The *differential* method requires the differential equation 3 to be integrable. The existence and uniqueness of an initial value problem is guaranteed if the jacobian to be integrated, ∇R^{-1}, is bounded for every step in the integration path [18].

3 Inversion of the divisive normalization

In order to check the behavior of the proposed method, the procedure is used to invert the previous non-linear model (eq. 2).

Firstly, the parameters values used are introduced in section 3.1. Secondly, the conditions for invertibility of the divisive normalization are studied in section 3.2. Finally, the results are compared with a conventional gradient-based method to solve non-linear equations in section 3.3.

3.1 Parameters of the divisive normalization

For this paper, we use a 16×16-point block DCT for the transformation T. The main results are general, and would apply to wavelet-style filter bank representations as well.

There are three basic sources from which one can obtain the normalization parameters: psychophysics [19, 2, 20], electrophysiology [4] and image statistics [21, 22]. In the psychophysically-inspired divisive normalization considered in this paper, the parameters are chosen by fitting data from human vision experiments, using a method similar to that of [19, 2]. As in [2], we augment the standard DCT with an additional scalar weighting parameter, α, accounting for the global sensitivity to the frequency range represented by each basis function (the CSF [23]). Thus, the transform coefficients, c_i, are given by:

$$c_i = \alpha_i \cdot \sum_{j=1}^{N^2} T_{ij}\, a_j$$

where T_{ij} are the basis functions of the linear transform that analyzes the image a_j. The amplitudes of the DCT are expressed in *contrast* dividing the coefficients by the local luminance. Similar contrast measures have been proposed in the literature in the context of pyramidal decompositions [24, 25, 26].

The parameters of the normalization are determined by fitting the slopes of the normalization function in eq. 2 to the inverses of the psychophysically measured contrast incremental thresholds for gratings [19, 2]. The values of α, β and h that fit the experimental responses of isolated sinusoidal gratings [27] are shown in figure 1. In the same way, the exponent was found to be $\gamma = 0.98$.

Given an image, a, of size $N \times N$, if T corresponds to a non-redundant basis, the size of the vectors c, r, α, β is N^2. The size of the matrix, h_{ij}, is $N^2 \times N^2$. For redundant bases the dimensions will be bigger. Considering these sizes, an arbitrary interaction pattern in the matrix h would imply an explicit (expensive) computation of the products $\sum_j h_{ij}|c_j|^\gamma$. Fortunately, the nature of the interactions between the coefficients is *local* [2, 28], as shown in figure 1. This fact induces a sparse structure in h and allows a very efficient computation of $\sum_j h_{ij}|c_j|^\gamma$ using simple convolutions. Since our experimental data don't constrain the shape of the

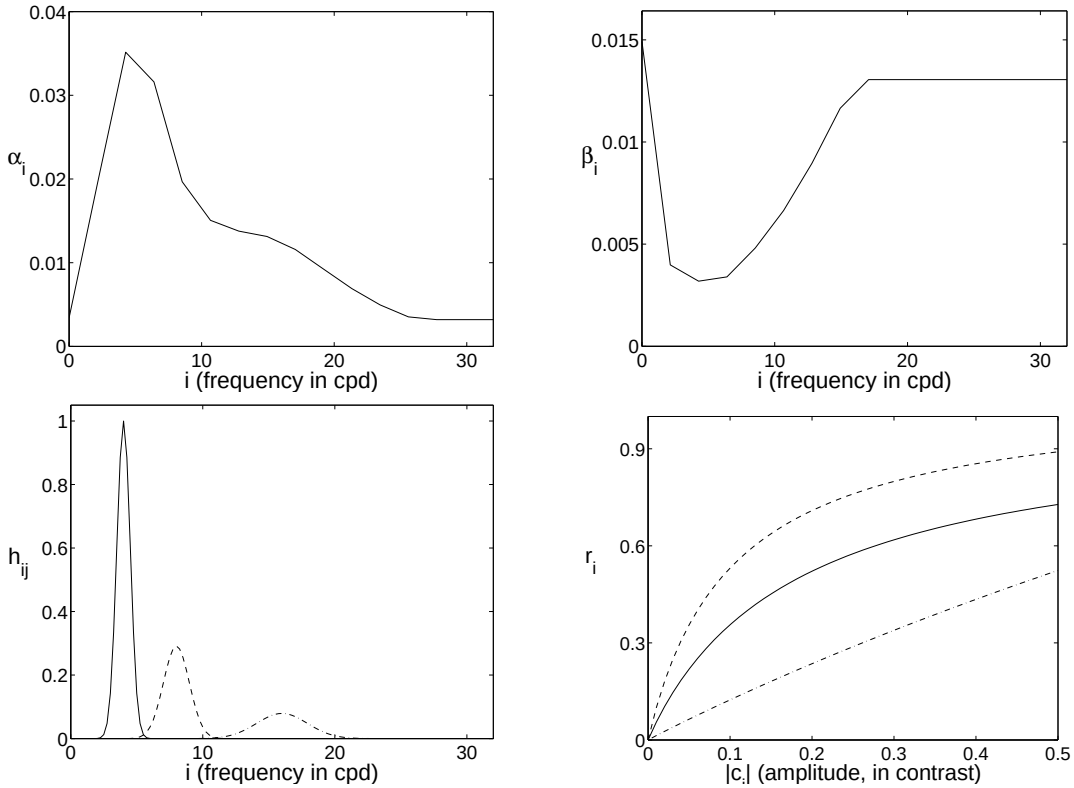

Figure 1: *Parameters α, β and (three examples of the interaction kernels in) h that fit the contrast incremental threshold data for the DCT case. The different line styles represent different frequencies: 4 cpd (solid), 8 cpd (dashed) and 16 cpd (dash-dot). The bottom right figure shows some examples of the normalized response as a function of coefficient amplitude, on a zero background.*

interaction function, following [2] we assume that each row of the matrix h has a Gaussian form over scale (spatial frequency) and orientation. Specifically, we set the kernels h_{ij} as,

$$\begin{aligned} h_{ij} &= exp(-|f_i - f_j|^2/\sigma_{f_i}^2) \\ \sigma_{f_i} &= \frac{1}{6}|f_i| + 0.05 \end{aligned}$$

where f_i and f_j are 2D vectors that represent the frequency (in cycles per degree) of the elements i and j. The distance $|f_i - f_j|^2$ determines a circular neighborhood in the 2D frequency domain that relates coefficients of similar frequency and orientation.

For other bases of interest such as wavelets, the normalization model can be extended introducing spatial interactions in the Gaussian kernels using the results reported in [2] or [28]: the spatial extent of the interactions is about twice the size of the impulse response of the CSF.

3.2 Invertibility of the divisive normalization

Let D_r and D_β be diagonal matrices with the absolute value of the elements of r and β in the diagonal, then from eq. 2 it follows:

$$|c|^\gamma = (I - D_r \cdot h)^{-1} \cdot D_\beta \cdot |r| \tag{6}$$

where the absolute value and the exponent γ are applied in a element-by-element basis. As in eq. 2, the sign (or phase) of each coefficient, c_i, is inherited from the sign (or phase) of the corresponding normalized coefficient, r_i.

Note that an analytic solution is possible in this particular case. However, the proposed method is applicable to more complex non-linearities, R, even if they are not analytically invertible.

As said, the *differential* method requires the differential equation 3 to be integrable. In our case from eq. 6, we have:

$$\nabla R^{-1}_{ij} = \frac{\partial R^{-1}(r)_i}{\partial r_j} = \mathrm{sgn}(r_i)\frac{1}{\gamma}|c_i|^{1-\gamma}\left((I - D_r \cdot h)^{-1}_{ij}\beta_j + \sum_{l=1}^{N^2}\frac{\partial (I - D_r \cdot h)^{-1}_{il}}{\partial r_j}\beta_l |r_l|\right) \tag{7}$$

which involves the matrix $(I - D_r \cdot h)^{-1}$.

Therefore, to ensure the existence of the solution, the following condition has to hold.
Let V and λ be the eigenvector and eigenvalue matrix decomposition of $D_r \cdot h$:

$$D_r \cdot h = V \cdot \lambda \cdot V^{-1}$$

As we show below, the invertibility condition turns out to be:

$$\lambda_{max} = max(\lambda_i) < 1 \tag{8}$$

According to eq. 7 the matrix $(I - D_r \cdot h)$ has to be invertible, i.e. $det(I - D_r \cdot h) \neq 0$. However if some eigenvalue, λ_i, is equal to one, then $det(\lambda_i I - D_r \cdot h) = 0$. In theory, it would be enough to ensure that $\lambda_i \neq 1$, but in practice, as the spectrum of $D_r \cdot h$ is almost continuous (see the examples of Fig. 2), the matrix is likely to be ill-conditioned if the condition 8 does not hold.
We have empirically checked the invertibility of the normalization that uses *psychophysically inspired* parameters for the local-DCT by computing the maximum eigenvalue of $D_r \cdot h$ over 25600 blocks randomly taken from the Van Hateren natural image data set [29]. Figure 2.a shows the average eigenvalues spectrum and figure 2.b the PDF of the maximum eigenvalue. In this experiment on a large natural data base the maximum eigenvalues are always far enough from 1. These results suggest that the normalization with these parameters will be invertible, and it will remain invertible even if the responses r undergo small distortions such as quantization.

3.3 Experimental results

In order to check the performance of the differential inversion, several experiments are carried out. Moreover, a traditional gradient-guided search is also quoted here to stress the differences of the differential procedure with a standard root-finding method that uses equivalent information.

3.3.1 Standard gradient-based search: root finding in non-linear equations

Obtaining c from eq. 2 can be also formulated as a non-linear root finding problem: we have to find the vector c that minimizes the error $|r - R(c)|$. The standard techniques to solve these problems involve a gradient-based iterative search such as the steepest descent or the Newton-Raphson methods [30]. For example, the evolution of the steepest descent solution is:

$$c_{(n)} = c_{(n-1)} + \mu^{opt}_{(n-1)} \cdot \nabla R(c_{(n-1)}) \cdot (r - R(c_{(n-1)})) \tag{9}$$

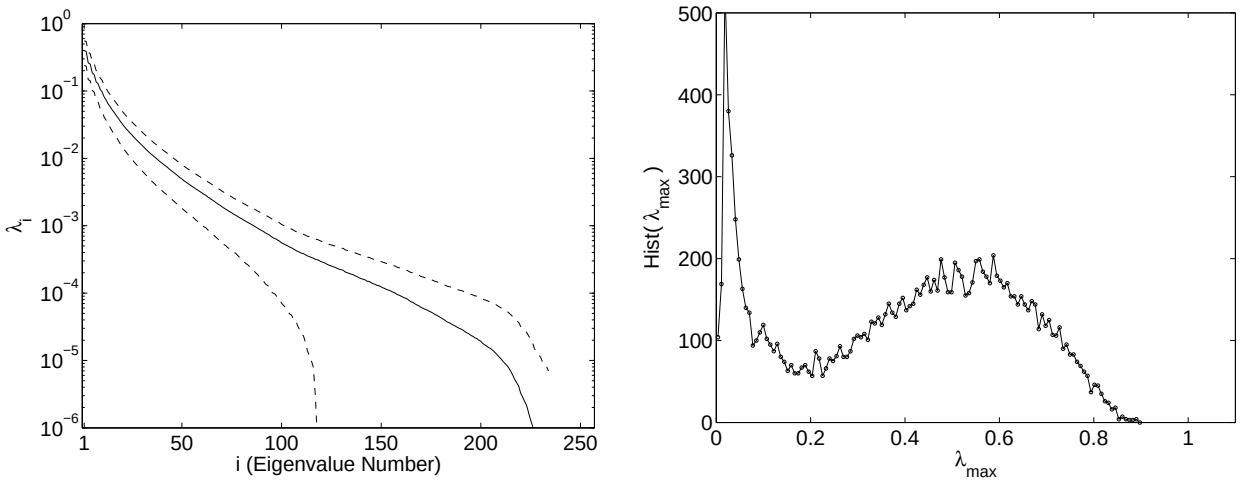

Figure 2: *Behavior of $D_r \cdot h$ for a set of 25600 blocks taken from the Van Hateren data base [29]. (a) Average eigenvalues spectrum. Dashed lines represent the standard deviation. (b) PDF of $\lambda_{\max}$*

where $\mu_{(n)}^{opt}$ is the factor that minimizes the error in each step.

The differential technique to solve $R^{-1}(r)$ is an alternative to these standard methods. The way in which the problem is posed and the fact that the convergence to the solution is theoretically guaranteed (provided some condition holds, see section 3.2) make a difference in the proposed technique. However, the resemblance of the Euler integration, eq. 5, with a gradient-guided search, eq. 9, suggests that the proposed method should be compared with the minimization techniques that use the same information. On the other hand, the proposed procedure would be pointless if the problem of the inversion were so easy that could be solved by a straightforward minimization. To prove that this is not the case, next we compare the performance of the differential technique with a standard steepest descent method.

3.3.2 Inversion results

Figure 3 shows the average error of the reconstructed image-block as a function of the number of gradient computations (each step of the 4^{th}-order Runge-Kutta method uses 4 gradient computations). For this test we used 2560 blocks and different initial guesses.

This experiment shows that the steepest descent method cannot successfully solve the problem because it gets trapped in local minima (far from an acceptable solution). This suggests that the inversion problem is not trivial. The error surface may have multiple different local minima as suggested by the fact that the solution depends on the initial guess. On the contrary, the proposed technique achieves better solutions, faster, and regardless of the initial conditions.

The convergence of the proposed method can also be seen through the inverted images for different number of steps: 1, 2, 4, 6, 8, 14 19 and 25. Figures 4, 5, 6 and 7 show the inversion for different well-known images, when the initial condition is the average DCT block in the training set.

According to these experiments, we could conclude that the proposed method achieves the inversion (with a non-significant distortion) of all blocks, in few steps and not depending on the initial conditions.

For the majority of blocks, an acceptable inverse is obtained with a small number

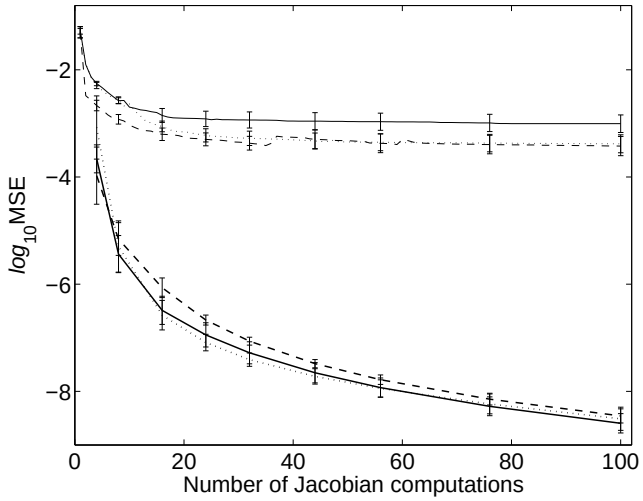

Figure 3: *Error of the inverse using the steepest descent method (upper thin lines) and the differential method (lower thick lines). The different line styles indicate different initial conditions: average DCT block in the training set (dotted lines), flat spectrum (dashed lines) and* $1/f$ *spectrum (solid lines).*

of steps (around 4), it does not matter the initial conditions. Only few blocks need a higher number of steps, but 25 steps have never been exceeded. Therefore, the method does not seem sensitive to the initial conditions, the final distortion depends on the block itself.

4 Concluding remarks

In this paper, a differential method for inverting the non-linear part of the current V1 model is proposed. This numerical inversion algorithm is general and applicable to more complex non-linearities, R, even if they are not analytically invertible. This allows to use the ideas proposed here with more complex perceptual models that may appear in the future. The only requirement is to compute the jacobian ∇R.

The method has been applied to invert the divisive normalization (eq. 2). The results indicate that the inverse is achieved in few steps, not depending on the initial conditions. Moreover, the error of the inverse using our method is dramatically reduced compared with when the steepest descent method is used.

References

[1] 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.*, 1876:658–667, 2000.

[2] A.B. Watson and J.A. Solomon. A model of visual contrast gain control and pattern masking. *Journal of the Optical Society of America A*, 14:2379–2391, 1997.

[3] H.R. Wilson. Pattern discrimination, visual filters and spatial sampling irregularities. In M.S. Landy and J.A. Movshon, editors, *Computational Models of Visual Processing*, pages 153–168, Massachussets, 1991. MIT Press.

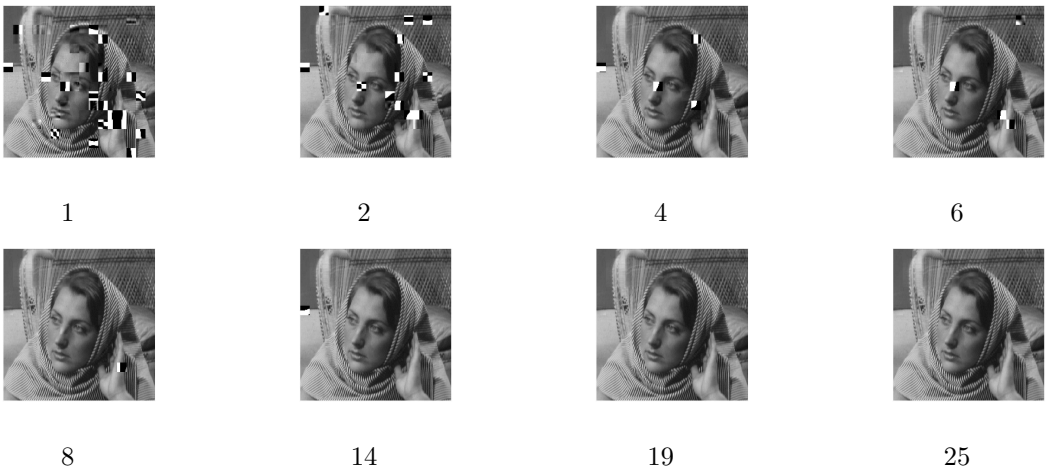

Figure 4: Inversion of Barbara with different number of steps

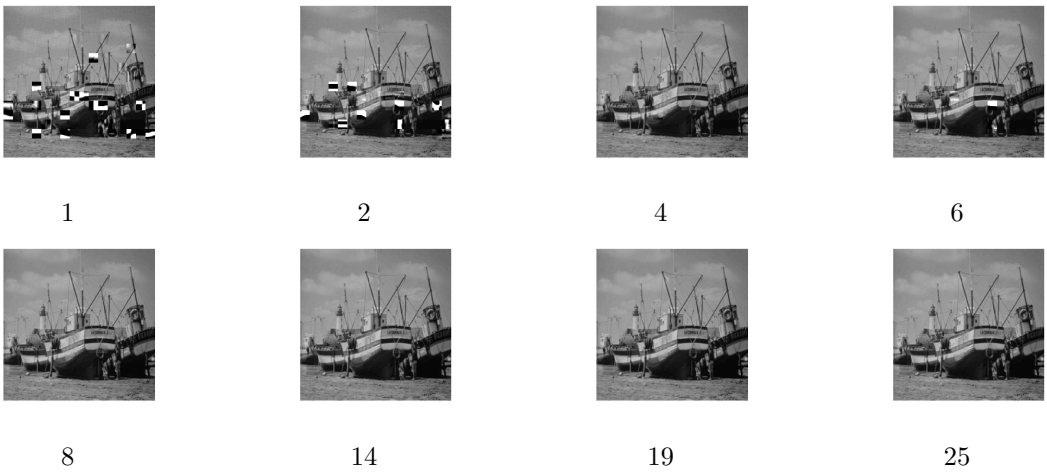

Figure 5: Inversion of Boats with different number of steps

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

[5] S. Marcelja. Mathematical description of the response of simple cortical cells. *Journal of the Optical Society of America*, 70(11):1297–1300, 1980.

[6] A.B. Watson. The cortex transform: Rapid computation of simulated neural images. *Computer Vision, Graphics and Image Processing*, 39:311–327, 1987.

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

[8] A.B. Bonds. Temporal dynamics of contrast gain in single cells of the cat striate cortex. *Visual Neuroscience*, 6:239–255, 1991.

[9] W S Geisler and D G Albrecht. Cortical neurons: Isolation of contrast gain control. *Vision Research*, 8:1409–1410, 1992.

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

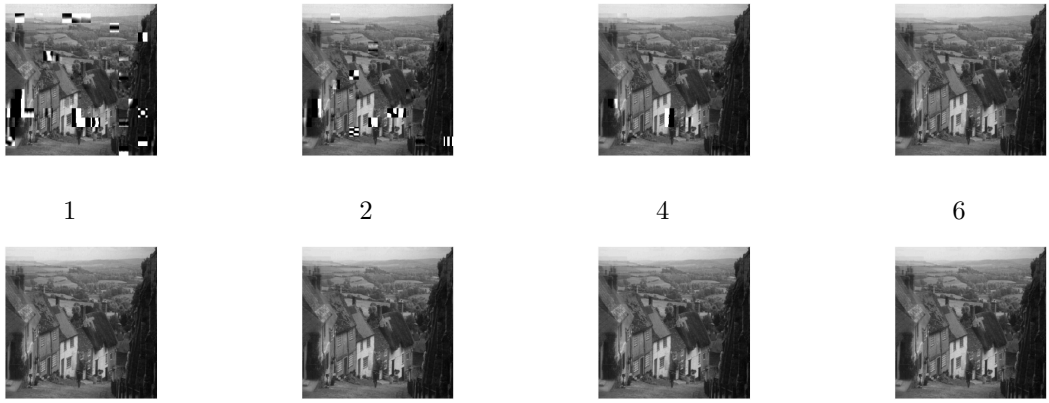

Figure 6: Inversion of Golden Hill with different number of steps

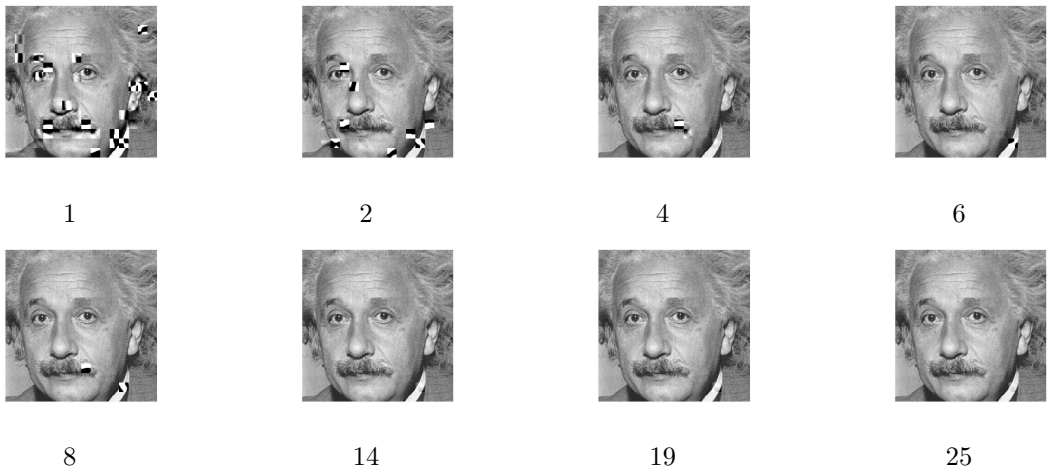

Figure 7: Inversion of Einstein with different number of steps

[11] P.C. Teo and D.J. Heeger. Perceptual image distortion. *Proceedings of the SPIE*, 2179:127–141, 1994.

[12] 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.

[13] I. Epifanio, J. Gutirrez, and J.Malo. Linear transform for simultaneous diagonalization of covariance and perceptual metric matrix in image coding. *Pattern Recognition*, 36:1799–1811, 2003.

[14] I. Ohzawa, G. Sclar, and R.D. Freeman. Contrast gain control in the cat's visual system. *Nature*, 298:266–268, 1982.

[15] P.C. Teo and D.J. Heeger. Perceptual image distortion. *Proc. of the First IEEE Intl. Conf. Im. Proc.*, 2:982–986, 1994.

[16] E.P. Simoncelli and O. Schwartz. Modeling surround suppression in V1 neurons with a statistically-derived normalization model. In M.S. Kearns, editor, *Adv. in Neural Inf. Proc. Syst.*, volume 11. MIT Press, 1999.

[17] J. Malo, J. Gutirrez, and J. Rovira. V1 non-linearities from local linear ICA. *Submmited to: Network Inf. Neur. Sys. (special issue on Sensory Coding and the Natural Environment)*, Dec 2004.

[18] D.J. Logan. *Introduction to non-linear partial differential equations.* Wiley&Sons, NY, 1994.

[19] G.E Legge. A power law for contrast discrimination. *Vision Research*, 18:68–91, 1981.

[20] P. Neri. Estimation of nonlinear psychophysical kernels. *Journal of Vision*, 4(2):82–91, 2004.

[21] D. L. Ruderman. The statistics of natural images. *Network: Computation in Neural Systems*, 5(4):517–48, 1994.

[22] O. Schwartz and E.P. Simoncelli. Natural signal statistics and sensory gain control. *Nature Neuroscience*, 4(8):819–825, 2001.

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

[24] E. Peli. Contrast in complex images. *JOSA A*, 7:2032–2040, 1990.

[25] M. Duval-Destin, M.A. Muschietti, and B. Torrésani. Continuous wavelet decompositions: Multiresolution and contrast analysis. *SIAM J. Math. Anal.*, 24:739–755, 1993.

[26] Vandergheynst P and Gerek O. Nonlinear pyramidal image decomposition based on local contrast parameters. In *Proceedings of the IEEE-Eurasip Workshop on Nonlinear Signal and Image Processing (NSIP99)*, volume 2, pages 770–773. IEEE, June 1999.

[27] A.M. Pons. *Estudio de las Funciones de Respuesta al Contraste del Sistema Visual.* PhD thesis, Dpt. d'Òptica, Facultat de Física, Universitat de València, Julio 1997.

[28] A.B. Watson and J.Malo. Video quality measures based on the standard spatial observer. *Proc. IEEE Intl. Conf. Im. Proc.*, 3:41–44, 2002.

[29] 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.

[30] W.H. Press, B.P. Flannery, S.A. Teulosky, and W.T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, Cambridge, 1992.