Nice Statistical Properties of V1 Cortex Image Representation

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. Here, the standard V1 cortex model optimized to reproduce image distortion psychophysics is shown to have nice statistical properties, e.g. approximate factorization of the PDF of natural images. These results confirm the efficient encoding hypothesis that aims to explain the organization of biological sensors by information theory arguments.

1 Introduction

Barlow (Barlow, 1961) suggested that functional properties of biological vision sensors should be matched to the signal statistics faced by these sensors. The standard approach to confirm the plausibility of such hypothesis goes *from image statistics to perception*: e.g. predicting the shape of the linear receptive fields (Olshausen and Field, 1996) and the non-linear behavior in V1 (Schwartz and Simoncelli, 2001; Malo and Gutiérrez, 2006), by using image statistics, efficient encoding arguments and no physiological or psychophysical information. Nowadays, there is a productive debate in Computational Neuroscience about the generality of the original efficient encoding hypothesis, or the strict applicability of redundancy reduction arguments (Barlow, 2001; Simoncelli, 2003).

In this work we show that an alternative confirmation of the matching between biological vision systems and image statistics may be obtained by following the opposite direction: i.e. *from perception to image statistics*. Here we show that psychophysically fitted divisive normalization has appealing statistical properties (e.g. approximate factorization of the PDF of natural images and redundancy reduction) when no statistical information is used in the model.

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) method to set its parameters. Section 3 analytically shows how the proposed model may factorize a plausible PDF for natural images. Section 4 empirically shows how the proposed model achieves component independence and redundancy reduction. Finally, section 5 draws the conclusions of the work.

2 V1 visual cortex model

The image representation considered here is based on the standard psychophysical and physiological model that describes the early visual processing up to the V1 cortex (Mullen, 1985; Malo, 1997; Heeger, 1992; Watson and Solomon,

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 (Watson and Solomon, 1997). The linear sensors have a frequency dependent linear gain according to the Contrast Sensitivity Function (CSF), $\mathbf{S}_i$, (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; Watson and Solomon, 1997):

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

In this scheme, the rows of the matrix $\mathbf{T}$ contain the receptive fields of V1 neurons, here modeled by an orthogonal 4-scales QMF wavelet transform[1]. $\mathbf{S}$ is a diagonal matrix containing the linear gains to model the CSF. Finally, $\mathbf{R}$ is the Divisive Normalization response:

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

where H is a kernel matrix that controls how the responses of neighboring linear sensors, k, affect the non-linear response of sensor i. Here we use the Gaussian interaction kernel proposed by Watson and Solomon (Watson and Solomon, 1997), which has been successfully used in block-frequency domains (Malo et al., 2006; Gutiérrez et al., 2006; Camps et al., 2008). In the wavelet domain the width of the interaction kernel for spatial, orientation and scale neighbors has to be found. The resulting kernel is normalized to ensure that $\sum_k H_{ik} = 1$. 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 initial guess is consistent with the interpretation of the values β_i as priors of the amplitude of the coefficients (Schwartz and Simoncelli, 2001). This profile (computed from 100 images of a calibrated image data base[2] is further multiplied by a constant to be fitted to the psychophysical data.

The above V1 image representation induces a subjective image distortion metric. Given an input image, $\mathbf{x}$, and its distorted version, $\mathbf{x}' = \mathbf{x} + \Delta\mathbf{x}$, the model provides two response vectors, $\mathbf{r}$, and $\mathbf{r}' = \mathbf{r} + \Delta\mathbf{r}$. The perceived distortion has been proposed to be the Euclidean norm of the difference vector (Teo and Heeger, 1994), but non-quadratic pooling norms have also been reported (Watson and Solomon, 1997; Watson and Malo, 2002).

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). 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 by fitting threshold psychophysics or physiological recordings (Heeger, 1992; Watson and Solomon, 1997). This low-level approach is not straightforward because the experimental

---

[1] http://www.cns.nyu.edu/ lcv/software.php
[2] http://tabby.vision.mcgill.ca

literature is often interested in a subset of the parameters, and a variety of experimental settings is used. 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 (Olshausen and Field, 1996; Schwartz and Simoncelli, 2001; Malo and Gutiérrez, 2006), but that is not the right thing to do since we want to include no statistical information in the 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 Divisive Normalization metric to maximize the correlation with the subjective ratings of a subset of the LIVE Quality Assessment Database[3]. The range of the parameter space was set according to an initial guess obtained from threshold psychophysics (Mullen, 1985; Watson and Solomon, 1997; Malo, 1997) and previous use of similar models in image processing applications (Malo et al., 2006; Gutiérrez et al., 2006; Camps et al., 2008).

Figure 1 shows the optimal values for the linear gains $\mathbf{S}$, the regularization constants β^γ and the interaction kernel H. The particular structure of the interaction kernel comes from the particular arrangement of wavelet coefficients used in the transform. The optimal value for the excitation and inhibition exponent was $\gamma = 1.7$. The optimal values for the spatial and frequency summation exponents were $q_s = 3.5$ and $q_f = 2$, where the summation is made first over space and then over the frequency dimensions.

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 approximately 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 model proposed above to a set of natural images.

---

[3] http://live.ece.utexas.edu/research/quality/
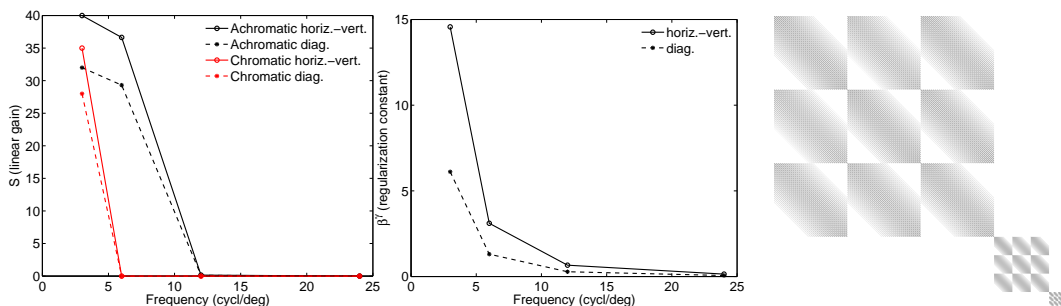
Fig. 1. *Linear gains S (left), regularization constants β^γ (center), and kernel H (right).*

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. This quite evident by looking at the so called bow-tie plot: the conditional probability of a coefficient given the values of some of its neighbors, $P(w'_i|w'_j)$, normalized by the maximum of the function for each value of w'_j (see Fig. 2). These facts have been used to propose leptokurtotic functions to model the marginal PDFs (Hyvärinen, 1999) and models of the conditional PDFs in which the variance of one coefficient depends on the variance of the neighbors (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 covariance, Σ_{ii}, depends on the neighbors:

$$P_{\mathbf{w}'}(\mathbf{w}') = \mathcal{N}(0, \Sigma(\mathbf{w}')) = \frac{1}{(2\pi)^{N/2}|\Sigma(\mathbf{w}')|^{1/2}} \, e^{-\frac{1}{2}\,\mathbf{w}'^T \cdot \Sigma(\mathbf{w}')^{-1} \cdot \mathbf{w}'} \tag{3}$$

where,

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

Note that this joint PDF *is not* Gaussian because the variance of each coefficient depends on the neighbors according to the kernel in eq. 4. Therefore, 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, $P_{w'_i}(w'_i)$.

A 2D toy example using 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 denominator of the normalization and the covariance of the image model.
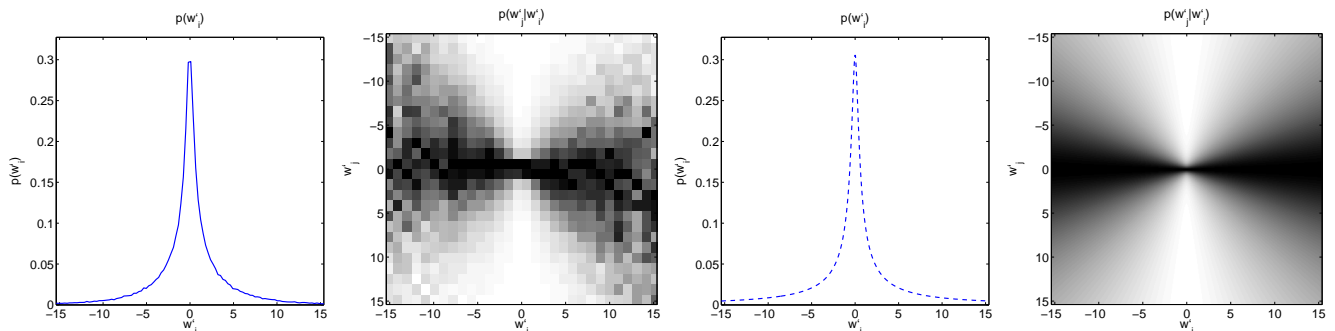
Fig. 2. *Left: empirical behavior of wavelet coefficients of natural images (marginal PDF and conditional PDF). Right: simulated behavior according to the proposed model. In this toy experiment we considered two coefficients of the second scale of $\mathbf{w}'$ (computed for 8000 images). We used $S_i = 0.14$, $\beta_i = 0.4$, $H_{ii} = 0.7$ and $H_{ij} = 0.3$ and $\gamma = 1.7$, according to the psychophysically fitted model.*

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),

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

Considering that the divisive normalization (in vector notation) is just: $\mathbf{r} = sign(\mathbf{w}') \, \Sigma(\mathbf{w}')^{-\frac{\gamma}{2}} \cdot |\mathbf{w}'|^{\gamma}$, 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):

$$|\mathbf{w}'|^{\gamma} = (I - D_{|\mathbf{r}|}H)^{-1} \cdot D_{\beta\gamma} \cdot |\mathbf{r}| \tag{5}$$

$$\mathbf{w}' = sign(\mathbf{r})\Sigma(\mathbf{w}')^{\frac{1}{2}} \cdot |\mathbf{r}|^{\frac{1}{\gamma}} \tag{6}$$

where D_v are diagonal matrices with the vector v in the diagonal. Plugging $\mathbf{w}'$ into the image model and using $|\mathbf{w}'|^{\gamma}$ to compute the Jacobian of the inverse, we have,

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

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

Assuming that the matrix in the brace is negligible[4]:

$$|\nabla_{\mathbf{r}} \mathbf{R}^{-1}| \sim \; det(\Sigma(\mathbf{w}'))^{1/2} \prod_{i=1}^{N} \frac{1}{\gamma} r_i^{\frac{1}{\gamma}-1} \tag{7}$$

it follows that the joint PDF of the normalized signal is just the product of N functions that depend solely on r_i:

$$P_{\mathbf{r}}(\mathbf{r}) = \prod_{i=1}^{N} \frac{1}{\gamma(2\pi)^{1/2}} \; r_i^{\frac{1}{\gamma}-1} \; e^{-\frac{r_i^{2/\gamma}}{2}} = \prod_{i=1}^{N} P_{r_i}(r_i) \tag{8}$$

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

Even though factorization of the PDF does not depend on γ, it determines the shape of the marginal PDFs (see Fig. 3). However, note that different values of γ would imply a better (or worse) match between the denominator of the normalization and the covariance of the image model.

4   Results

This section assesses the component independence performance of the psychophysically fitted V1 image representation (i.e. the validity of Eq. 3) by (1) Mutual

---

[4] This approximation was validated by (1) computing the average value of the matrix on a set of 8000 normalized images (results not shown here), and by (2) the agreement between predictions and empirical behavior shown in Section 4.
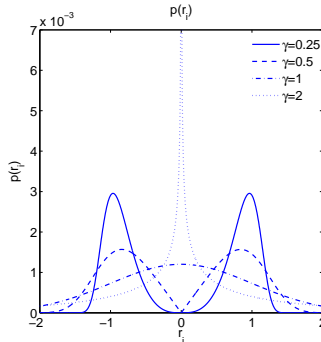
Fig. 3. *Family of marginal PDFs of the normalized coefficients r_i as a function of γ.*

Information (MI) measures, and (2) by analyzing the conditional probabilities of the transformed coefficients. To do so, 8000 image patches, **x**, of size 72×72 were considered and transformed to the linear wavelet domain, **w**, and to the non-linear V1 representation, **r**. For the sake of illustration, the results for two values of the exponent γ are used in the divisive normalization: the psychophysically optimal value $\gamma = 1.7$, and $\gamma = 0.5$ due to the (predicted) characteristic shape of the marginal PDFs in that case (see Fig. 3).

4.1 Mutual Information measures

Table 1 shows the MI results (in bits) for pairs of coefficients in **w** and **r**. 120000 pairs of coefficients were used in each estimation. Two kinds of MI estimators were used: (1) direct computation of MI, which involves 2D histogram estimation (Cover and Tomas, 1991), and (2) estimation of MI by PCA-based Gaussianization (GPCA) (Laparra et al., 2009), which only involves univariate histogram estimations.

These results show that the wavelet representation removes about 92% of the redundancy in the spatial domain, and divisive normalization further reduces about 69% of the remaining redundancy. This suggests that one of the goals of the psychophysical V1 image representation is redundancy reduction.

	w	$\mathbf{r^{(0.5)}}$	$\mathbf{r^{(1.7)}}$
Intraband (scale = 2)	0.29 (0.27)	0.17 (0.17)	0.16 (0.15)
Intraband (scale = 3)	0.24 (0.22)	0.08 (0.09)	0.09 (0.09)
Inter-scale, scales = (1,2)	0.17 (0.17)	0.10 (0.11)	0.08 (0.08)
Inter-scale, scales = (2,3)	0.17 (0.15)	0.04 (0.04)	0.04 (0.04)
Inter-scale, scales = (3,4)	0.09 (0.07)	0.01 (0.01)	0.01 (0.01)
Inter-orientation (H-V), scale = 2	0.10 (0.08)	0.01 (0.01)	0.01 (0.01)
Inter-orientation (H-V), scale = 3	0.08 (0.06)	0.01 (0.01)	0.01 (0.01)
Inter-orientation (H-D), scale = 2	0.16 (0.15)	0.04 (0.04)	0.03 (0.03)
Inter-orientation (H-D), scale = 3	0.15 (0.14)	0.01 (0.01)	0.02 (0.02)

Table 1. *MI measures in bits. GPCA MI estimations are shown in parenthesis. Just for reference, the MI among luminance values in the spatial domain is 2.12 (2.14) bits.*

4.2 Marginal and conditional PDFs

Figure 4 shows the predicted and the experimental marginal PDFs in the normalized domain and the experimental conditional PDFs. The resemblance among theory and experiments confirms the theoretical results in section 3. Note also that the PDF of one coefficient given the neighbor is more independent of the neighbor value than in the wavelet domain (Fig. 2). This is particularly true in the case of using the optimal value $\gamma = 1.7$, thus indicating the match of the physchophysically optimal vision model and image statistics. Note also that the agreement between the marginal PDFs and the theoretical prediction is better for the optimal exponent.

5 Conclusions

Here we showed that the standard V1 cortex model optimized to reproduce image quality psychophysics increases the independence of the image coefficients obtained by linear ICA (wavelet-like) filters. Theoretical results (confirmed by experiments) show that the V1 model approximately factorizes a plausible joint PDF in the wavelet domain: bow-tie dependencies are almost removed and redundancy is substantially reduced.

The results presented here confirm the efficient encoding hypothesis in a novel direction: *from perception to image statistics.*
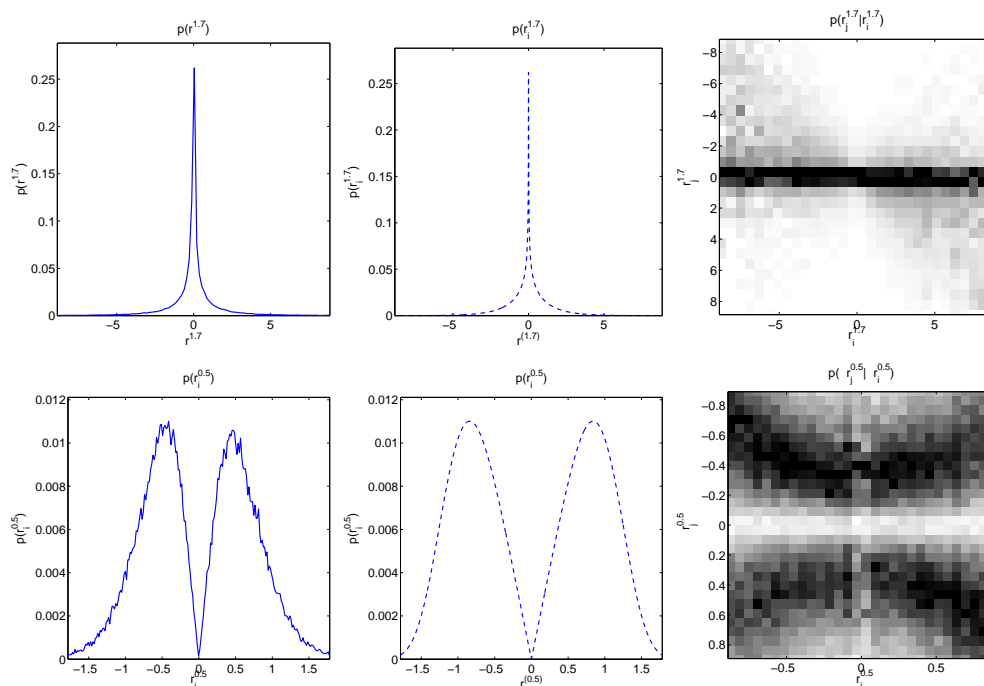
Fig. 4. *Experimental marginal PDF (left), theoretical prediction (center), and bow-tie plots (right) for* **r** *coefficients using the optimal value of $\gamma = 1.7$ (top) and other illustrative value, $\gamma = 0.5$ (bottom).*

Bibliography

[Barlow, 1961]Barlow, H. (1961). Possible principles underlying the transformation of sensory messages. In Rosenblith, W., editor, *Sensory Communication*, pages 217–234. MIT Press, Cambridge, MA.

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

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

[Cover and Tomas, 1991]Cover, T. and Tomas, J. (1991). *Elements of Information Theory*. John Wiley & Sons, New York.

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

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

[Hyvärinen, 1999]Hyvärinen, A. (1999). Sparse code shrinkage: Denoising of non-gaussian data by ML estimation. *Neur. Comp.*, pages 1739–1768.

[Laparra et al., 2009]Laparra, V., Camps, G., and Malo, J. (2009). PCA gaussianization for image processing. In *Submitted to: IEEE ICIP*.

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

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

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

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

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

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

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

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

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

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

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

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