ICA FEATURES OF COLOUR AND STEREO IMAGES

Patrik O. Hoyer and Aapo Hyvärinen

Neural Networks Research Centre
Helsinki University of Technology
P.O. Box 5400, FIN-02015 HUT, Finland
http://www.cis.hut.fi/projects/ica/

ABSTRACT

Previous work has shown that independent component analysis (ICA) applied to natural image data yields features resembling Gabor functions and simple-cell receptive fields. This article considers the effects of including chromatic and stereo information. The inclusion of colour leads to features divided into separate red/green, blue/yellow, and bright/dark channels. Stereo image data, on the other hand, leads to binocular receptive fields which are tuned to various disparities. The similarities between these results and observed properties of simple cells in primary visual cortex are further evidence for the hypothesis that visual cortical neurons perform some type of redundancy reduction, which was one of the original motivations for ICA in the first place. In addition, ICA provides a principled method for feature extraction from colour and stereo images; such features could be used in image processing operations such as denoising, compression, and pattern recognition.

1. INTRODUCTION

Ever since Hubel and Wiesel's classic experiments [16] a number of studies have investigated the receptive field properties of neurons in the primary visual cortex. Receptive fields have been found to be localised in space and time, have band-pass characteristics in spatial and temporal frequency and are selective to some preferred orientation (see, e.g. [10, 9]). In addition, many are selective to direction of movement, chromatic contrast, and/or binocular disparity [10, 25, 4].

Why do the neurons respond the way they do? Barlow [3] has proposed that the neurons form a representation with independent components, whereas Field [13] has argued for a 'sparse coding' of the visual input, which is in many ways equivalent to Barlow's proposal.

Recently, these theories have been tested experimentally. Olshausen and Field [26, 27] applied a sparseness maximization network to input data consisting of image patches from natural images. Basically, one attempts to represent each image patch as a linear combination of 'basis' patches, such that the mixing coefficients are as sparse as possible. In other words,

$$\mathbf{x} = \mathbf{A}\mathbf{s} = \sum_{i=1}^{m} \mathbf{a}_i s_i, \qquad (1)$$

where we have denoted the input image by $\mathbf{x}$, and the basis images are the $\mathbf{a}_i$, the columns of $\mathbf{A}$. One then optimizes the $\mathbf{a}_i$ such that for typical $\mathbf{x}$, most of the s_i will be close to zero and only a few will have significantly non-zero values. This led to features qualitatively similar to simple cell receptive fields.

Subsequently, Bell and Sejnowski [6] as well as Hurri et al. [18] applied independent component analysis (ICA) [8] to similar data. In ICA, the decomposition is also linear as in (1), but now the purpose is to seek mutually independent components s_i. In other words, we seek a factorial code for the data. This leads to a local representation quite similar to that obtained through sparse coding. In fact, it has been shown [27] that the sparseness-maximization network and ICA are very closely related. Later, van Hateren and van der Schaaf [37] quantitatively compared the filters learned by ICA to measurements of neural receptive fields, and found a good match for most parameters.

The extension of the above experiments into the spatio-temporal domain was performed in [36]. Instead of considering only static image patches, they took sequential patches from video sequences to take into account temporal redundancy. Again, the found ICA decomposition seemed to fit the receptive field properties found in the cortex, giving features tuned to temporal as well as spatial frequencies.

This article reports on similar numerical experiments on the two remaining dimensions of our visual input: chromatic contrast and stereopsis. We extract features from natural colour and stereo images using ICA, and compare the results to known receptive field properties of neurons found in the visual cortex. A longer version of this article can be found in [15].

2. COLOUR IMAGE EXPERIMENTS

2.1. Data

In this section, we extend the gray-scale ICA image model to include colours, by using natural colour images as data. Neurons of course receive their colour information ultimately from the outputs of the cones in the retina. Thus our input data should consist of the hypothetical outputs of the three types of cones in response to our images. However, any three linear combinations of these outputs is just as good an input data when applying ICA. Linearly transforming the data transforms the mixing matrix, but does not alter the independent components (sources).

We thus choose to use standard digital red/green/blue (RGB) images as input. This is motivated by the equivalence stated above, and allows a direct comparison of the features found to features currently in use in image processing operations such as compression or denoising, and could straightforwardly be applied in such tasks. Our data simply consists of colour versions of natural scenes (depicting forest, wildlife, rocks, etc.) which we have used in previous work as well [20, 21]. The data is in the form of 20 RGB images (of size 384×256-pixels) in standard TIFF format.

The drawback of using RGB values as inputs is of course that any nonlinearities inherent in the conversion from RGB to cone responses will affect the ICA result and a comparison to properties of neurons may not be warranted. We have found that nonlinearities do not seem to play a major role; see [15] for details.

2.2. Preprocessing

From the images, a total of 50,000 12-by-12 pixel image patches were sampled randomly. Since each channel yields 144 pixels, the dimensionality was now $3 \times 144 = 432$. The mean was subtracted componentwise, centering the dataset on the origin (this is a standard preprocessing step in ICA). Then, we calculated the correlation matrix and it's eigenvectors. The eigenvectors consist of global features, resembling 2D Fourier bases. The variance decreases with increasing spatial frequency, and when going from grayscale to blue/yellow to red/green features (for details see [15]). These results were also observed in [29].

We project the data onto the 160 first principal components before whitening.[1] This is done for two reasons. First, something similar is probably done in real neurons, as amplifying directions with small variance would be disastrous in terms of signal-to-noise ratio [2]. Second, the dimension is dropped to lower computational costs. After whitening, the FastICA algorithm [22, 19] is run to estimate the ICA basis matrix, starting from a random initial matrix and using function $g_1(u) = \tanh(u)$ [19].

2.3. Results

The estimated basis, in full colour, can be found at

`www.cis.hut.fi/phoyer/papers/colouricabasis.jpg`

The first thing to note is that each basis patch mainly codes one of three channels: bright/dark variations, blue/yellow features, or red/green information. (The only exceptions are the three basis vectors coding the average colour in the patch.) Thus we can, without losing much information, display the basis achromatically by dividing the features according to colour channel. This is done in Figure 1.

Examining the basis closely reveals that the features found are very similar to earlier results [26, 6] on gray-scale image data, i.e. the patches resemble Gabor-functions. Most units are (mainly) achromatic, so they only represent brightness variations. This is in agreement with the finding that a large part of the neurons in the primary visual cortex seem to respond equally well to different coloured stimuli, i.e. are not selective to colour [17, 25]. In addition, there is a small number of red/green and blue/yellow patches.[2] These are also oriented, but of much lower spatial frequency, similar to the gray-scale patches of lowest frequency. One could think that the low frequency patches together form a 'colour' (including brightness) system, and the high-frequency gray-scale patches a channel analysing form.

Although early results [25] on the chromatic properties of neurons suggested that most colour-sensitive cells were unoriented, and exhibited center-surround receptive fields, more recent studies have indicated that there are also oriented colour-selective neurons [33]. The fact that our colour-features are mostly oriented is thus at least in partial agreement with neurophysiological data.

In any case, there is some agreement that most neurons are not selective to chromatic contrast, rather are more concerned about form [17, 25, 34]. Our basis

[1] The exact number (160) of retained components is here selected purely based on considerations of computational complexity. We have experimented with various choices and found the qualitative aspects of the colour ICA basis fairly robust to changes in the amount of components retained.

[2] The colour content of the features is discussed in more detail in [15].

is in agreement with these findings. In addition, the cytochrome oxidase blobs which have been linked to colour processing [25] have also been associated with low spatial frequency tuning [32, 30]. In other words, colour selective cells should be expected to be tuned to lower spatial frequencies. This is also seen in our features.

3. STEREO IMAGE EXPERIMENTS

3.1. Data

Another interesting extension of the basic gray-scale image ICA model can be made by modeling stereopsis. Now, our artificial neurons are attempting to learn the dependencies of corresponding patches from natural stereo images. Again, the choice of data is an important step for us to get realistic results. Previous studies have used different approaches. In some early work, a binocular correlation function was estimated from actual stereo image data, and subsequently analyzed [24]. In addition, at least one investigation of receptive field development used artificially generated disparity from monocular images [31]. We have chosen to use 11 images from a commercial collection[3] of stereo images of natural scenes.

To simulate the workings of the eyes, we selected 5 focus points at random from each image and estimated the disparities at these points. We then randomly sampled 16×16-pixel corresponding image patches in an area of 300×300 pixels centered on each focus point, obtaining a total of 50,000 samples. Because of the local fluctuations in disparity (due to the 3D imaging geometry) corresponding image patches often contained similar, but horizontally shifted features; this is of course the basis of stereopsis.

Note that in reality the 'sampling' is quite different. Each neuron sees a certain area of the visual field which is relatively constant with respect to the focus point. Thus a more realistic sampling would be to randomly select 50,000 focus points and from each take corresponding image patches at some given constant positional offset. However, the binocular matching is computationally slow and we thus opted for the easier approach, which should give the same distribution of disparities.

3.2. Preprocessing

The same kind of preprocessing was used in these experiments as was used for colour. Since each sample consisted of corresponding left and right 16×16-patches our original data was 512 dimensional. First, the local mean was subtracted componentwise, to center the data on the origin. Then, we calculated the correlation matrix of the data, and projected the data onto the first 160 principal components. The data was then whitened by normalizing the variances of the components. Again, dropping the dimension is plausible as a coding strategy for neurons, but is mainly done to lower the computational expenses and thus running time and memory consumption. Due to the structure of the correlation matrix, dropping the dimension to 160 is similar to low-pass filtering.

[3] `http://members.home.net/holographics/cd~1.htm`

3.3. Results

Figure 2 shows the estimated basis vectors after convergence of the FastICA algorithm [22, 19]. Each pair of patches represents one basis vector $\mathbf{a}_i$. First, note that pairs have varying degrees of binocularity. Many of our 'model neurons' respond equally well to stimulation from both eyes, but there are also many which respond much better to stimulation of one eye than to stimulation of the other, i.e. the features exhibit varying degrees of ocular dominance. A quantitative analysis of this is given in [15].

Taking a closer look at the binocular pairs reveals that for most pairs the left patch is similar to the right patch both in orientation and spatial frequency. The positions of the features inside the patches are close, when not identical. In some pairs the phases are very similar, while in others they are quite different, even completely opposite. These properties make the features sensitive to different degrees of binocular disparity. Identical left-right receptive fields make the feature most responsive to zero disparity, while receptive fields identical except for a phase reversal show strong inhibition to zero disparity.

The disparity tuning was quantitatively analysed in [15]. It was shown that most of the features fit neatly into the categories 'tuned excitatory', 'tuned inhibitory', 'near', or 'far', which have been identified in physiological experiments [28, 14, 23]. Tuned excitatory units showed a strong peak at zero, usually with smaller inhibition at either side. Tuned inhibitory units on the other hand showed a marked inhibition (cancelling) at zero disparity, with excitation at small positive or negative disparities. Features classified as 'near' showed a clear positive peak at crossed (positive) disparity while those grouped as 'far' a peak for uncrossed (negative) disparity.

4. RELATION TO OTHER WORK

Although there have been numerous studies of learning gray-scale receptive fields, not many have attempted the analysis of colour or stereopsis. There has been some work concerning the second-order statistics of colour [1, 35, 29]. In addition, coloured input was used in [5] to emerge a topographic map of receptive fields. Again, that work is indirectly about only the second-order structure of the data, as the correlation-based learning used relies only on this information. The current work is thus the first (to the knowledge of the authors) to consider higher-order statistics of colour images.

Emerging receptive fields from stereo input has been considered in [24, 31, 11, 12]. As with colour, most studies have explicitly or implicitly used only second-order statistics [24, 11, 12]. The exception is [31] which used the BCM learning rule [7] which is a type of projection pursuit learning closely linked to ICA. The main difference between their work and ours is that we use data from actual stereo images whereas they used horizontally shifted (misaligned) data from regular images. In addition, we estimate a complete basis for the data, whereas they studied only single receptive fields.

5. CONCLUSIONS

We have investigated the use of independent component analysis for decomposing natural colour and stereo images. ICA applied to colour images yields basis vectors which resemble Gabor functions, with most features achromatic, and the rest red/green- or blue/yellow-opponent. When ICA is applied on stereo images we obtain feature pairs which exhibit various degrees of ocular dominance and are tuned to various disparities.

These results are significant for two reasons. First, the features learned by ICA could be straightforwardly applied in denoising, compression, or pattern recognition of colour or stereo data. In each of these tasks it is important to model the statistical structure of the data; ICA has been successfully used to model that structure [20]. Second, ICA can be used to model computational properties of V1 cells. The similarity of the ICA features to optimal stimuli measured for neurons in the primary visual cortex using single-cell recordings suggest that these neurons perform some form of redundancy reduction, as proposed by Barlow [3]. It seems likely that information processing strategies successful in the primary visual cortex would also be useful in higher visual processing, and indeed in processing of other sensory signals; thus it seems probable that ICA could be applied in modeling these functions as well.

6. REFERENCES

[1] J. J. Atick, Z. Li, and A. N. Redlich. Understanding retinal color coding from first principles. *Neural Computation*, 4:559–572, 1992.

[2] J. J. Atick and A. N. Redlich. What does the retina know about natural scenes? *Neural Computation*, 4:196–210, 1992.

[3] H. B. Barlow. Unsupervised learning. *Neural Computation*, 1:295–311, 1989.

[4] H. B. Barlow, C. Blakemore, and J. D. Pettigrew. The neural mechanism of binocular depth discrimination. *Journal of Physiology*, 193:327–342, 1967.

[5] H. G. Barrow, A. J. Bray, and J. M. L. Budd. A self-organized model of 'color blob' formation. *Neural Computation*, 8:1427–1448, 1996.

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

[7] E. L. Bienenstock, L. N. Cooper, and P. W. Munro. Theory for the development of neuron selectivity: orientation specificity and binocular interation in visual cortex. *Journal of Neuroscience*, 2:32–48, 1982.

[8] P. Comon. Independent component analysis – a new concept? *Signal Processing*, 36:287–314, 1994.

[9] G. C. DeAngelis, I. Ohzawa, and R. D. Freeman. Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development. *Journal of Neurophysiology*, 69:1091–1117, 1993.

[10] R. L. DeValois, E. W. Lund, and N. Hepler. The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research*, 22:531–544, 1982.

[11] E. Erwin and K. D. Miller. Modeling joint development of ocular dominance and orientation maps in primary visual cortex. In J. M. Bower, editor, *Computational neuroscience: Trends in research 1995*, pages 179–184. New York: Academic Press, 1996.

[12] E. Erwin and K. D. Miller. Correlation-based development of ocularly-matched orientation and ocular dominance maps: determination of required input activities. *Journal of Neuroscience*, 18:5908–5927, 1998.

[13] D. J. Field. What is the goal of sensory coding? *Neural Computation*, 6:559–601, 1994.

[14] B. Fischer and J. Kruger. Disparity tuning and binocularity of single neurons in cat visual cortex. *Experimental Brain Research*, 35:1–8, 1979.

[15] P. O. Hoyer and A. Hyvärinen. Independent component analysis applied to feature extraction from colour and stereo images. 2000. Submitted, available at http://www.cis.hut.fi/phoyer/papers/colourstereo.ps.

[16] D. Hubel and T. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160:106–154, 1962.

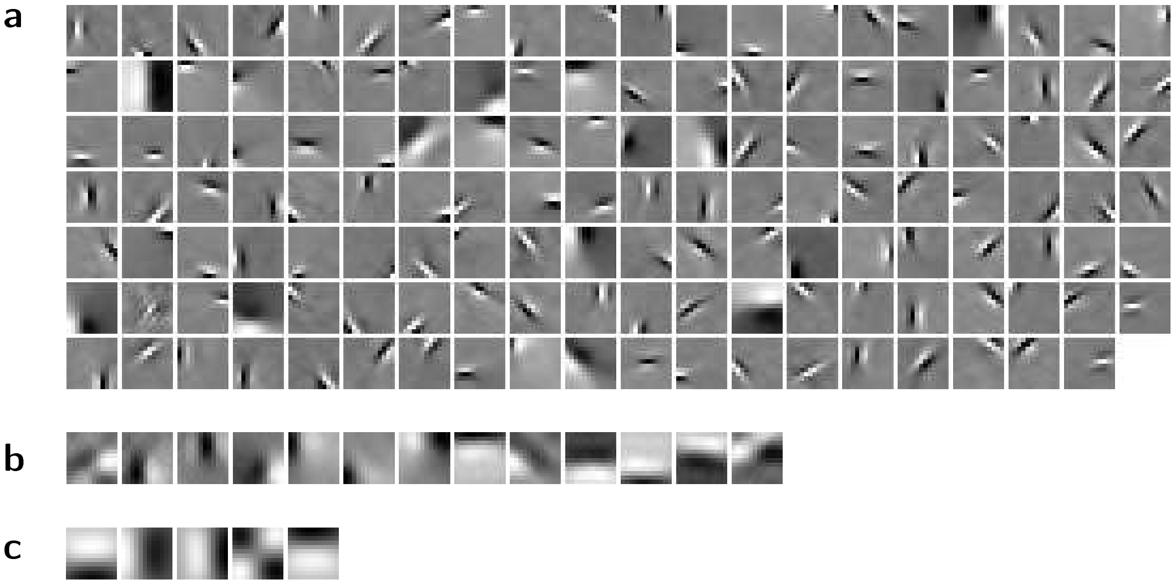

Figure 1: The ICA basis of colour images, with bright/dark, blue/yellow, and red/green features displayed separately. (a) The features (basis patches) coding variations in luminance (bright/dark channel). Altogether 139 basis vectors belonged in this group. (b) Features coding for blue/yellow variations (13 basis vectors). Black represents blue and white represents yellow. (c) Basis vectors representing red/green features (5). Black represents green and white red. (Note that the signs of the basis vectors are arbitrary, as we are assuming symmetric sources in the ICA model.) Not shown are the three basis vectors needed to code the average colour of the patch. The basis can be viewed in full colour at `http://www.cis.hut.fi/phoyer/papers/colouricabasis.jpg`.

[17] D. Hubel and T. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, 195:215–243, 1968.

[18] J. Hurri, A. Hyvärinen, and E. Oja. Wavelets and natural image statistics. In *Proc. Scandinavian Conf. on Image Analysis '97*, Lappenranta, Finland, 1997.

[19] A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. on Neural Networks*, 10(3):626–634, 1999.

[20] A. Hyvärinen. Sparse code shrinkage: Denoising of nongaussian data by maximum likelihood estimation. *Neural Computation*, 11(7):1739–1768, 1999.

[21] A. Hyvärinen and P. O. Hoyer. Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces. *Neural Computation*, 2000. (in press).

[22] A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483–1492, 1997.

[23] S. LeVay and T. Voigt. Ocular dominance and disparity coding in cat visual cortex. *Visual Neuroscience*, 1:395–414, 1988.

[24] Z. Li and J. J. Atick. Efficient stereo coding in the multiscale representation. *Network: Computation in Neural Systems*, 5:157–174, 1994.

[25] M. S. Livingstone and D. H. Hubel. Anatomy and physiology of a color system in the primate visual cortex. *Journal of Neuroscience*, 4:309–356, 1984.

[26] 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.

[27] B. A. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, 37:3311–3325, 1997.

[28] G. F. Poggio and B. Fischer. Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. *Journal of Neurophysiology*, 40:1392–1405, 1977.

[29] D. L. Ruderman, T. W. Cronin, and C. Chiao. Statistics of cone responses to natural images: Implications for visual coding. *Journal of the Optical Society of America A*, 15(8):2036–2045, 1998.

[30] D. Shoham, M. Hübener, S. Schulze, A. Grinvald, and T. Bonhoeffer. Spatio-temporal frequency domains and their relation to cytochrome oxidase staining in cat visual cortex. *Nature*, 385:529–533, 1997.

[31] H. Shouval, N. Intrator, C. C. Law, and L. N. Cooper. Effect of binocular cortical misalignment on ocular dominance and orientation selectivity. *Neural Computation*, 8:1021–1040, 1996.

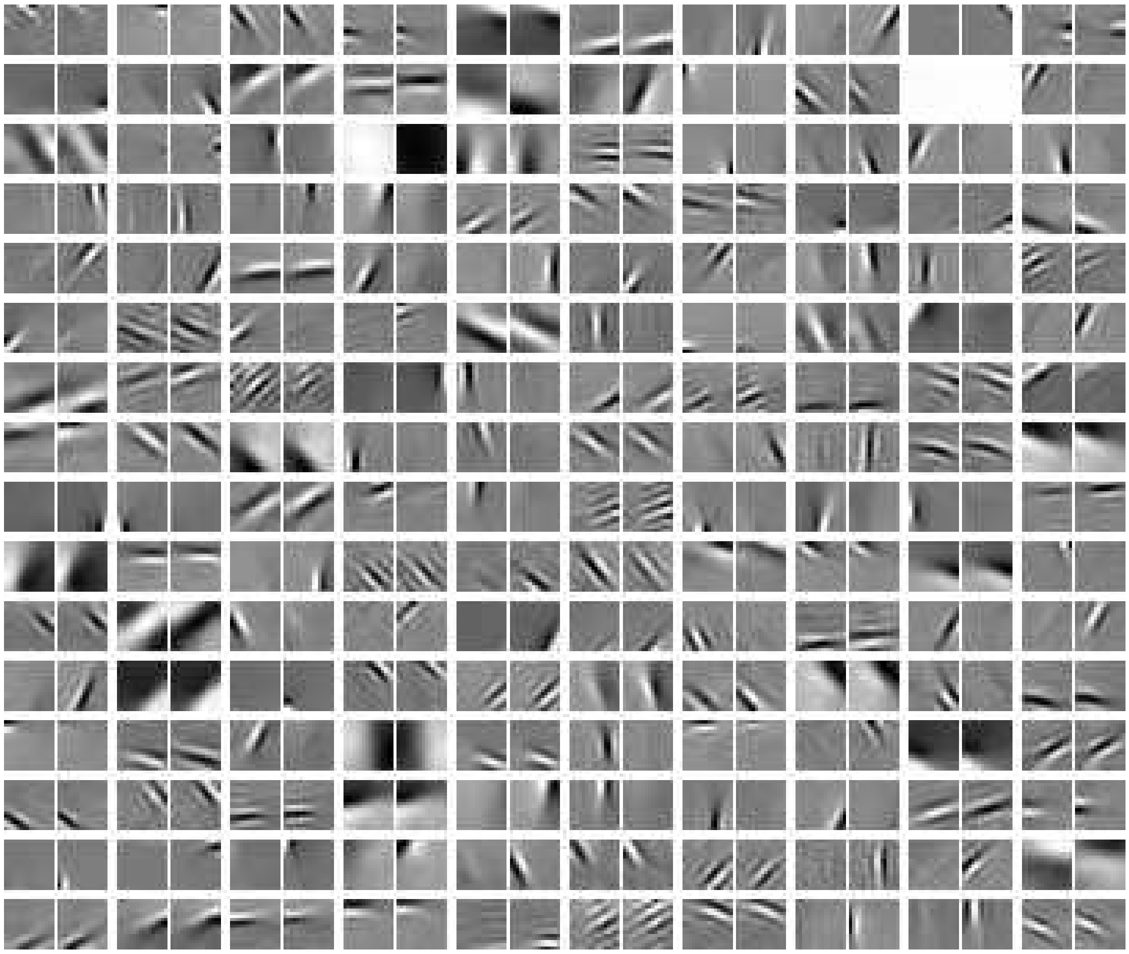

Figure 2: The ICA basis of stereo image patches. Each pair of patches represents one basis vector $\mathbf{a}_i$ of the estimated mixing matrix.

[32] R. B. H. Tootell, M. S. Silverman, S. L. Hamilton, E. Switkes, and R. L. De Valois. Functional anatomy of macaque striate cortex. V. Spatial frequency. *Journal of Neuroscience*, 8:1610–1624, 1988.

[33] D. Y. Ts'o and C. D. Gilbert. The organization of chromatic and spatial interactions in the primate striate cortex. *Journal of Neuroscience*, 8(5):1712–1727, 1988.

[34] D. Y. Ts'o and A. W. Roe. Functional compartments in visual cortex: Segregation and interaction. In M. S. Gazzaniga, editor, *The Cognitive Neurosciences*, pages 325–337. MIT Press, 1995.

[35] J. H. van Hateren. Spatial, temporal and spectral preprocessing for colour vision. *Proc. Royal Society ser. B*, 251:61–68, 1993.

[36] J. H. van Hateren and D. L. Ruderman. Independent component analysis of natural image sequences yields spatiotemporal filters similar to simple cells in primary visual cortex. *Proc. Royal Society ser. B*, 265:2315–2320, 1998.

[37] J. H. van Hateren and A. van der Schaaf. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. Royal Society ser. B*, 265:359–366, 1998.