

CN530, Spring 2004
Final Report
Independent Component Analysis and Receptive Fields

Madhusudana Shashanka
shashanka@cns.bu.edu

Abstract

This report surveys the literature that uses Independent Component Analysis on natural images and natural image sequences to understand receptive field properties of cells in the visual cortex. I present studies which compare ICA filters/basis functions with spatial, spatio-temporal, spatio-chromatic and stereo properties of receptive fields of simple and complex cells. The intuition behind the ICA algorithm is explained. Finally, I present results of some neurophysiological experiments which conclude that the present comparison of the results of ICA with receptive fields is not justified and more careful studies are needed.

Introduction

It has long been assumed that neurons are adapted, through both evolutionary and developmental processes, to the statistical properties of the signals to which they are exposed. Because not all signals are equally likely, it is natural to assume that perceptual systems should be able to best process those signals that occur most frequently. Thus, the statistical properties of the environment are relevant for sensory processing. In recent years, this *information theoretic* framework for studying vision has gained in popularity. According to this view, sensory input is a signal that carries *information* about the outside world and the visual system somehow recodes these inputs to *reduce redundancy*, providing an ‘economical’ description of the signals. This hypothesis is called *redundancy reduction*, or *efficient coding*.

There are two main approaches to test the efficient coding hypothesis. A review of research in this area can be found in [11]. One possibility is to record from neurons at various levels of the visual system and try to directly estimate the information versus redundancy in their patterns of firing. The other approach is to consider the statistics of the typical sensory input and then ‘derive’ a model for efficient coding of this input. Then, the model is compared to known physiology of the early visual system. In the following sections, we focus on the latter approach, especially a method called Independent Component Analysis (ICA). We examine how it has been used to model early visual processing. Many researchers have compared the results of ICA processing with receptive fields of cells in the visual cortex. We examine these studies and see to what extent these claims are supported by experimental results.

Independent Component Analysis

Background

The basic efficient coding hypothesis states that sensory neurons should be adapted to transmit the maximum amount of information about the natural environment, given limited resources. Using information theory, under certain assumptions it can be shown that the mutual information between the sensory input and the neural response is maximized when the entropy of the response is at a maximum (see [6]). In case of many neurons, maximal response entropy requires that the responses of the neurons are statistically independent.

The question now is whether we can find a transformation from an input consisting of natural image data that would give responses which are mutually independent. In the general case, this is a complicated problem and researchers have mainly focused on the special case where the mapping is constrained to be linear. The problem is to find such a mapping. Several algorithms have been developed for this purpose and *Independent Component Analysis* is a class of such algorithms.

Mathematical Formulation

Hyvärinen and Oja ([7]) give a detailed account of ICA algorithms and applications. The mathematical formulation we present here is based on this paper. To define ICA, we use a statistical “latent variables” model. Assume that we observe n linear mixtures x_1, \dots, x_n of n independent components

$$x_j = a_{j1}s_1 + a_{j2}s_2 + \dots + a_{jn}s_n, \quad \text{for all } j. \quad (1)$$

Let us denote by \mathbf{x} the random vector whose elements are the mixtures x_1, \dots, x_n , and likewise by \mathbf{s} the random vector with elements s_1, \dots, s_n . Let us denote by \mathbf{A} the matrix with elements a_{ij} . Then, the model can be written as

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad \text{or} \quad \mathbf{x} = \sum_{i=1}^n \mathbf{a}_i s_i \quad (2)$$

where \mathbf{a}_j represents the j th column of the matrix \mathbf{A} . The statistical model in the above equation is called independent component analysis, or ICA model. The ICA model is a generative model, which means that it describes how the observed data are generated by a process of mixing of the components s_i . The independent components are latent variables, meaning that they cannot be observed. The mixing matrix is also unknown. We must estimate \mathbf{A} and \mathbf{s} using the only observable vector \mathbf{x} and this must be done under as general assumptions as possible.

The first assumption is that the components s_i are statistically *independent*. We must also assume that the independent components have *non-Gaussian* distributions. But notice that we do *not* assume these distributions known. After estimating the matrix \mathbf{A} , we can compute its inverse, say \mathbf{W} , and obtain the independent components simply by $\mathbf{s} = \mathbf{W}\mathbf{x}$.

Principles of ICA Estimation

The key to estimating the ICA model is non-Gaussianity. According to the central limit theorem, a classical result of probability theory, the distribution of a sum of independent random variables tends toward a Gaussian distribution under certain conditions. Thus, a sum of two independent random variables (non-gaussian) usually has a distribution that is closer to Gaussian than any of the two original variables.

Consider a data vector \mathbf{x} which is a mixture of independent components. For simplicity of explanation, we assume that all the independent components have identical distributions. To estimate one of the independent components, we consider a linear combination of the x_i ; let us denote this by $y = \mathbf{w}^T \mathbf{x} = \sum_i w_i x_i$, where \mathbf{w} is a vector to be determined. In practice, we cannot determine such a \mathbf{w} exactly, because we have no knowledge of matrix \mathbf{A} , but we can find an estimator that gives a good approximation. We introduce a new variable $\mathbf{z} = \mathbf{A}^T \mathbf{w}$. Then we have $y = \mathbf{w}^T \mathbf{x} = \mathbf{w}^T \mathbf{A}\mathbf{s} = \mathbf{z}^T \mathbf{s}$. y is thus a linear combination of s_i , with weights given by z_i . Since a sum of even two independent random variables is more Gaussian than the original variables, $\mathbf{z}^T \mathbf{s}$ is more Gaussian than any of the s_i and becomes least Gaussian when it in fact equals one of the s_i . Therefore, we could take as \mathbf{w} a vector that *maximizes the non-Gaussianity* of $\mathbf{w}^T \mathbf{x}$. Such a vector $\mathbf{w}^T \mathbf{x} = \mathbf{z}^T \mathbf{s}$ would be equal to one of the independent components. This is the main idea of the algorithm.

Different quantitative measures for non-Gaussianity can be used. They include *kurtosis*, *negentropy* and *approximations of negentropy using higher-order moments*. Other approaches for ICA estimation include *minimization of mutual information* and *maximum likelihood estimation*. The interested reader is referred to [7] for more details.

Natural Images and ICA

In this section, we examine various studies connecting ICA and early visual processing. But before we proceed, we shall present a brief background of the development of this idea in vision science. We have already introduced the idea of information theoretic framework earlier. Here we introduce the *analysis-by-synthesis* approach and *latent variable* models.

Different modeling approaches

The traditional computational approach to vision focuses on how, from the image data \mathbf{x} , one can compute quantities of interest, which we will call \mathbf{s} . In other words, the emphasis is on a function f that transforms images into object information, as in $\mathbf{s} = f(\mathbf{x})$. This operation might be called image *analysis*. But consider the opposite operation called image *synthesis*. The mapping g that generates the image given the state of the world ($\mathbf{x} = g(\mathbf{s})$), is considerably easier to work with than the mapping f . One can search for the parameters $\hat{\mathbf{s}}$ that produce an image $\hat{\mathbf{x}} = g(\hat{\mathbf{s}})$ which, as well as possible matches the observed image \mathbf{x} . This approach to vision is known as *analysis-by-synthesis*.

In the probabilistic approach to vision, analysis-by-synthesis is naturally implemented in the framework of *latent variable models*. Latent variable models attempt to explain observed data by some underlying hidden causes or factors that we have only indirect information about. In the previous discussion of ICA, the variables s_i correspond to the hidden factors determining the data x_i (see equation (1)). There are two fundamental operations, *inference* and *learning*. Inference refers to estimating the hidden variables \mathbf{s} given an observed data vector \mathbf{x} . Usually, one must be content with a probability density $p(\mathbf{s}|\mathbf{x})$, which can be calculated using the Bayes rule. The latent variable model specifies how data vectors are synthesized (generated) from the ‘causes’ \mathbf{s} by $p(\mathbf{x}|\mathbf{s})$. These causes are selected so that the fit to the observed data is maximized. The second fundamental operation in latent variable models is the estimation, or learning, of the model from observed data. The dependencies are parametrized and the model parameters are adapted so that they best fit the observed data. Thus, considering equation (1), we can summarize that inference consists of estimating the hidden variables s_j that generated the given data vector \mathbf{x} , while learning is the estimation of the model parameters a_{ij} from a large set of data vectors.

Basic facts of neurophysiology

The receptive field properties of simple cells¹ in mammalian primary visual cortex have been studied extensively. The receptive fields are localized in space and time, have band-pass characteristics in the spatial and temporal frequency domains, are oriented, and are often sensitive to the direction of motion of a stimulus. Several hypotheses as to the function of these cells have been proposed. As the cells preferentially respond to oriented edges or lines, they can be viewed as edge or line detectors. Their joint localization in both the spatial domain and the spatial frequency domain has led to the suggestion that they mimic Gabor filters, minimizing uncertainty in both domains.

ICA and receptive fields

With this background, we now proceed to present some results that have been reported connecting the results of ICA on natural images and receptive fields of cells in the visual

¹The notions of ‘receptive field’, ‘simple’ and ‘complex’ cells are not precise. See [4], [8].

cortex.

In [9], Olshausen and Field described how a simple neural network performing sparse coding learned features that were qualitatively very similar to the receptive fields of V1 simple cells. This was the first study to show how all the basic spatial properties of simple cell classical receptive fields (localization in space and in spatial frequency, and in orientation tuning) could emerge in an unsupervised manner from natural images.

Bell and Sejnowski describe their results of ICA transformation on natural scenes in [1] and [2]. They present an analysis of the problem of learning a single layer of linear filters based on an ensemble of natural images. They used an unsupervised learning algorithm based on information maximization which performed ICA. Their simulations showed that filters found by the ICA algorithm are localized, oriented, and produced output distributions of very high sparseness while the independent components resembled edges. They interpreted the results as supporting the theory that visual cortex performs redundancy reduction.

In 1998, van Hateren and van der Schaaf (see [14]) extended the result by performing ICA on a large set of calibrated images, and comparing a series of properties of the resulting receptive fields (filters) with those of receptive fields measured in simple cells. While the earlier studies by Bell and Sejnowski established a *qualitative* correspondence, this study tried to examine the existence of a *quantitative* match. It was found that there was a good correspondence between the distributions for spatial frequency bandwidth, orientation tuning bandwidth, aspect ratio and receptive field length. However, results deviated strongly for the peak of the spatial frequency sensitivity. Whereas simple cells have receptive fields acting on different spatial scales (i.e. they show spatial scaling), the IC filters showed much less variability. The authors suggested some approaches to resolve this “discrepancy”. One possibility was that spatial scaling should be imposed as an extra constraint, as one could argue that spatial scaling is a useful property for higher visual processing. They mentioned a few other possibilities but believed that including the time domain would help. They concluded that their results strengthened the hypothesis that cortical simple cells strive to produce a representation of natural images with independent variables, each having a highly kurtotic amplitude distribution.

Both studies mentioned above considered purely spatial response characteristics. But much is also known about how simple cells respond to spatiotemporal, chromatic and binocular stimuli. There were several studies which tried to ascertain if the ICA model could account for these properties.

In one such study, van Hateren and Ruderman ([13]) considered video sequences of natural scenes. They found that ICA on such video yields ICs that resemble edges or bars, moving with a fixed velocity perpendicular to their main axis of orientation. The corresponding IC filters (ICFs) move with similar velocity, but at higher spatial and temporal frequencies. The spatiotemporal properties of the ICFs closely resembled those of receptive fields measured in simple cells. Both are confined to a limited region in space and time and both resemble an undulation traveling through a steady envelope. Both are found for a range of velocities, spatial frequencies and spatial scales. The first difference is that ICFs are centered at different positions and times. The authors consider ICFs that are spatially identical but centered at different times to be representing a single continuously acting filter. A second difference with the receptive fields of simple cells is that ICFs are more symmetrical in time and more narrowly tuned in temporal frequency.

Now, let us examine studies which considered chromatic properties. Taylor et al. ([12]) present results of applying linear ICA to color images of natural scenes. They found that the resulting ICFs separate into either luminance or color filters. The luminance filters were localized and oriented edge detectors as found in other studies. The color filters resembled either blue-yellow double-opponent (BYDO) or red-green

double-opponent (RGDO) receptive fields with various orientations. An equal number of each type of filter (luminance, red-green and blue-yellow) was obtained. Hence, the results may suggest the possibility that spatiochromatic information is processed together with orientation. This however would contradict the physiological evidence that color information is coded in parallel to but largely separate from orientation. The model generates double opponent receptive fields but as pointed out by the authors themselves, recent physiological studies have failed to demonstrate clearly discernible double-opponent receptive fields in V1. Despite the lack of strong physiological evidence, the authors conclude that ICA decomposition offers a near-optimum means of coding natural images. Other studies ([15]) also found qualitatively similar results.

In another study ([5]), Hoyer and Hyvärinen considered spatiochromatic properties. Rather than comparing IC filters with receptive fields, the authors compared the basis vectors with cell receptive fields. They found that most units were achromatic and a small number were red/green and blue/yellow patches. The color patches were oriented, but of much lower spatial frequency, similar to the gray-scale patches of the lowest intensity. The authors say that one could think that the low frequency patches together form a ‘color’ (brightness) system, and the high frequency patches, a channel analyzing form. Similar to the study mentioned earlier, they also observed the ICA patches to be double-opponent. To check whether the results depended on the image data set used, the study was conducted on two different data sets. Though there were quantitative differences between the results, they were qualitatively similar.

In [3], Doi et al. go one step further. They point out that in all earlier studies which considered spatio-chromatic properties, images were sampled on a regular grid, with the color at each location represented as a vector of three elements. But in the retina, cone photoreceptors are arranged in a mosaic, namely, only one cone at each location. Hence, they considered a model that incorporated the cone mosaic found in the trichromatic foveal region of primates. They adopted a hierarchical model that consists of a decorrelating stage corresponding to LGN and a subsequent ICA stage corresponding to V1. They assumed that early stages of visual processing are linear. Their results showed a majority (around 95%) of “luminance” type (achromatic) receptive fields and some “color-selective” receptive fields. The luminance type units resembled simple-cell receptive fields with high spatial-frequency selectivity suggesting that they represent spatial information (without color). The color-selective units were characterized by cone-type specific antagonistic regions in their receptive fields, and classified into two subtypes: “Y/B” type of yellow-blue selectivity, and “R/G” type of red-green selectivity. All the units exhibited color-opponency and most showed orientation selectivity. Also, these color-selective units had larger receptive fields compared to luminance units, consistent with the low spatial-frequency selectivity of the color mechanism. The authors hence say that “form” and “color” channels of the early visual system can be derived from the statistics of the sensory signals. They conclude that their results strengthen the redundancy reduction hypothesis.

In [5], Hoyer and Hyvärinen provide results of some experiments with stereo images where they tried to model stereopsis. They observed some interesting results after performing ICA on such images. Though many of their “model neurons” responded equally well to stimulation from both eyes, there were also many which responded much better to stimulation of one eye than to stimulation of the other. Binocular units had interocularly matched orientations and spatial frequencies, as has been observed for real binocular neurons. It was also found that both interocular phase and position differences existed. The simulated disparity tuning curves of the found features were observed to be similar to tuning curves measured in physiological experiments.

In [17], Zhang and Mei simulated the growth of young animals’ visual cortex in

special visual environment, using ICA. They applied ICA to three kind of images – natural images, topographic images and horizontal stripe images. They wanted to see if it can adapt to variations of environment as animal vision does. Its a known result (Blakemore and Cooper) that if a young cat grows in a horizontal stripe visual environment, simple cells in the cat’s visual cortex will almost only respond to stimuli along the horizontal edge. They applied ICA to images of horizontal stripes and found that all the ICA filters obtained were sensitive to horizontal orientation. They gave a quantitative comparison between the ICA filters and Gabor function by estimating the error between them using the simplex method. They conclude that ICA mimics the information processing of the visual system.

All the studies mentioned so far assume that early stages of visual processing are linear. There have been some studies recently which also explore non-linear and extra-classical receptive field properties. In [16], Zetzsche and Röhrbein considered how basic nonlinearities of cortical neurons - gain control and ON/OFF half-wave rectification - can exploit higher order statistical dependencies in the natural environment. They used both PCA (principal components analysis) and ICA and found that while PCA explores part of the redundancies, ICA could exploit third and higher order redundancies. They found that both schemes yield a variety of nonlinear units comprising the typical nonlinear processing properties such as end-stopping, side-stopping, complex-cell properties and extra-classical receptive field properties. However, the authors do mention that since they used only the most basic cortical nonlinearity (ON/OFF rectification), their approach could only approximate the actual nonlinear processing strategies by which the visual cortex exploits higher-order redundancies.

Recent results from neurophysiological experiments

Recently, Ringach ([10]) published some results of his experiments where he measured receptive fields of simple cells in macaque primary visual cortex. He compared his results with the filter shapes predicted by the two theories, ICA and sparse coding. Previous comparisons between predictions of such theories and experimental data were scarce and limited to one-dimensional methods (based on the line-weighting function). He observed that both these theories predict receptive fields with a larger number of subfields than observed in the experimental data. Also he observed receptive fields that were broadly tuned in orientation and low-pass in spatial frequency. But both these theories do not generate such receptive fields.

In the paper, Ringach discusses possible reasons behind the failure of ICA and SC (sparse coding) in explaining the distribution of filter shapes in V1. We reproduce them here. One possibility is that the function of simple cells is not to generate a full representation of the image, as suggested by these theories. It could also be that mean-squared error is not the measure being optimized. He mentions the possibility that considering minimization of other measures (such as the L_1 norm of the error) could produce receptive fields that match experimental data better. Another caveat about these comparisons is that the RF predictions are dependent on the preprocessing of the images used to train the algorithms, the statistics of the images used for training, the actual algorithm used to optimize the basis set, and whether the analysis is done on static images or image sequences. Since the dependence of the shapes of the predicted RFs on these implementation details has not been explored thoroughly yet, Ringach concludes that it would be premature to argue that the basic principles put forward by these theories are unsuitable for explaining simple-cell function in V1. He is hopeful that future realizations of these ideas might be able to account for the experimental data.

Problems with this approach

There are several problems with this approach. We first describe some specific issues mentioned by the authors of studies that we have covered in previous sections. We then consider general arguments advocated against such information theoretic and optimization approaches.

In [1] and [2], the authors mention that the transformation from retina to V1 is much more complicated than their simple matrix operator. One of the several objections to their scheme of representation is the fact that filters obtained were predominantly of high spatial frequency, unlike the several octave spread seen in the cortex. In [14], the authors concede that their model is not a full model of simple cells. They say that as their model is linear and non-adaptive, many aspects of simple cells are ignored, such as contrast adaptation, contrast normalization, nonlinearities involved in orientation tuning, adaptation to various stimulus statistics and so on. In [13], the authors give a more detailed discussion about the limitations of this approach. They say that the linear IC model for natural scenes can only be approximate at best because objects in scenes do not superimpose linearly, but by occlusion. Responses of simple cells are not expected to be completely independent of each other, because they appear to achieve a strongly overcomplete representation. Hence they caution that their analysis is too limited to make a full quantitative comparison possible. They also caution that systematic variations in the statistics of different scenes cause variability in ICF properties which should be taken into account when comparing with cortical receptive fields. This is not usually done in the kind of studies mentioned in the earlier sections. In [5], the authors say that “The striking resemblance of the ICA features to receptive fields of neurons in primary visual cortex suggests that the neurons do indeed perform some type of independent component analysis, and that the receptive fields are optimized for processing natural images. It is important to understand that this does not mean that the ‘learning rule’ or ‘algorithm’ actually operating in the cortex is anything like ours; rather it lets us understand the *purpose* of the computations as finding independent components of the input data. In the terminology of Marr, we model the computational level instead of the algorithmic or implementation levels.”

If we consider any optimization approach in general, the first thing that is suspect is the objective function. Assuming that nature indeed does some kind of optimization, one can still see that it will be very hard to characterize the optimization function accurately. Then, one has to also consider the constraints under which these properties evolved in biological systems. In almost all cases, we fail to take into account factors like genetic cost and ecological advantage. These factors are very important but at present there is no way we can quantify them. Also, some simple rhetorical questions can be raised to test the validity of the whole approach. For example, it is well known that V1 contains many more neurons (by orders of magnitude) than what is present in the LGN. One can take the optimization argument further and argue that the presence of so many neurons contradicts the optimization hypothesis.

In all the previous sections, we have discussed about natural images. But the question of how to define a ‘natural image’ is not a trivial one at all (see [11]). One should also be clear what one means by ‘receptive field’ and this is not an easy question to answer either (e.g. [4]). Assuming that these questions have been answered, one then has to confront the question of categorizing cells into simple and complex cells. Recent studies ([8]) have shown that there is no independent support for the simple/complex dichotomy. These studies suggest that the existence of two distinct neural populations in the primary visual cortex, and the associated hierarchical model of receptive field organization, need to be re-evaluated. This result calls into question the very basic

assumption of all the results dealt with in earlier sections.

Conclusions

In this report, I have given an overview of the literature that studies natural image statistics and tries to connect it with the receptive fields of cells in the visual cortex, with special emphasis on independent component analysis. Mathematical basis and intuition behind the ICA algorithm is explained. Studies which relate ICA with receptive fields are presented. Finally, problems with this approach are presented and highlighted. The report gives a balanced view of development of research in this area.

References

- [1] A. J. Bell and T. J. Sejnowski. Edges are the ‘independent components’ of natural scenes. In M. C. Mozer, M. J. Jordan, and T. Petsche, editors, *Advances in neural information processing systems*, volume 9, pages 831–837. MIT Press, 1997.
- [2] A. J. Bell and T. J. Sejnowski. The “independent components” of natural scenes are edge filters. *Vision Research*, 37(23):3327–3338, 1997.
- [3] E. Doi, T. Inui, Te-Won Lee, T. Wachtler, and T. J. Sejnowski. Spatiochromatic receptive field properties derived from information-theoretic analyses of cane mosaic responses to natural scenes. *Neural Computation*, 15:397–417, 2003.
- [4] David Fitzpatrick. Seeing beyond the receptive field in primary visual cortex. *Current Opinion in Neurobiology*, 10:438–443, 2000.
- [5] P. O. Hoyer and A. Hyvärinen. Independent component analysis applied to feature extraction from colour and stereo images. *Network: Computation in Neural Systems*, 11(3):191–210, 2000.
- [6] Patrik O. Hoyer. *Probabilistic Models of Early Vision*. PhD thesis, Helsinki University of Technology, 2002.
- [7] A. Hyvärinen and E. Oja. Independent component analysis: algorithms and applications. *Neural Networks*, 13:411–430, 2000.
- [8] Ferenc Mechler and Dario L. Ringach. On the classification of simple and complex cells. *Vision Research*, 42:1017–1033, 2002.
- [9] 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.
- [10] Dario L. Ringach. Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *J. Neurophysiol.*, 88:455–463, 2002.
- [11] Eero P. Simoncelli and Bruno A. Olshausen. Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24:1193–1216, 2001.
- [12] Dharmesh R. Taylor, Leif H. Finkel, and Gershon Buchsbaum. Color-opponent receptive fields derived from independent component analysis of natural images. *Vision Research*, 40:2671–2676, 2000.
- [13] J. H. van Hateren and D. L. Ruderman. Independent component analysis of natural images yields spatio-temporal filters similar to simple cells in primary visual cortex (preprint). *Proc. R. Soc. Lond.*, B 265:2315–2320, 1998.
- [14] 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.

- [15] Thomas Wachtler, Te-Won Lee, and Terrence J. Sejnowski. Chromatic structure of natural scenes. *Journal of Optical Society of America*, 18(1):65–77, January 2001.
- [16] Christoph Zetsche and Florian Röhrbein. Nonlinear and extra-classical receptive field properties and the statistics of natural scenes. *Network: Computational Neural Systems*, 12:331–350, 2001.
- [17] Liming Zhang and Jianfeng Mei. Shaping up simple cell’s receptive field of animal vision by ica and its application in navigation system. *Neural Networks*, 16:609–615, 2003.