Learning a Complex Model with structured Spatial Phases for Natural Scenes

Krishna Subramani

Electrical Engineering
Indian Institute of Technology Bombay, India

Human Visual System

- Vision can be defined as acquiring knowledge from the environment by analysis of the light reflected onto our eyes.[Statistics, 2009]
- Our Brain is somewhat adapted to the input statistics (adaptive representation). It concentrates on those aspects of data useful for further analysis
- It has evolved to solve particular kinds of problems very very efficiently(face recognition etc.)

Natural Images

- Why Natural Images - Possess Statistical Structure to which our Visual System is adapted.

Natural Images

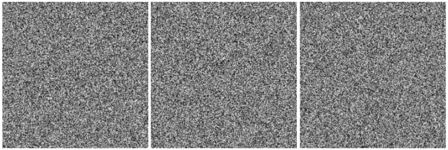
- Why Natural Images - Possess Statistical Structure to which our Visual System is adapted.

Examples of Natural Scenes

Natural Images

- Why Natural Images - Possess Statistical Structure to which our Visual System is adapted.

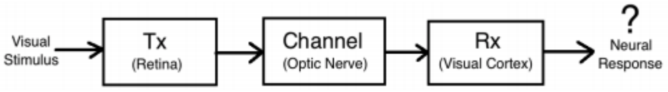
Examples of Natural Scenes



Examples of Random(Noise) Images

Efficient Coding Hypothesis[Barlow, 1961]

- One of the major task our visual system tries to do is to maximize the information of the environment/surroundings
- This hypothesis provides an Information Theoretic Perspective to our Visual System



Information Theoretic Perspective[1]

- Groups of neurons should encode information as compactly as possible in order to maximize efficient utilization of resources

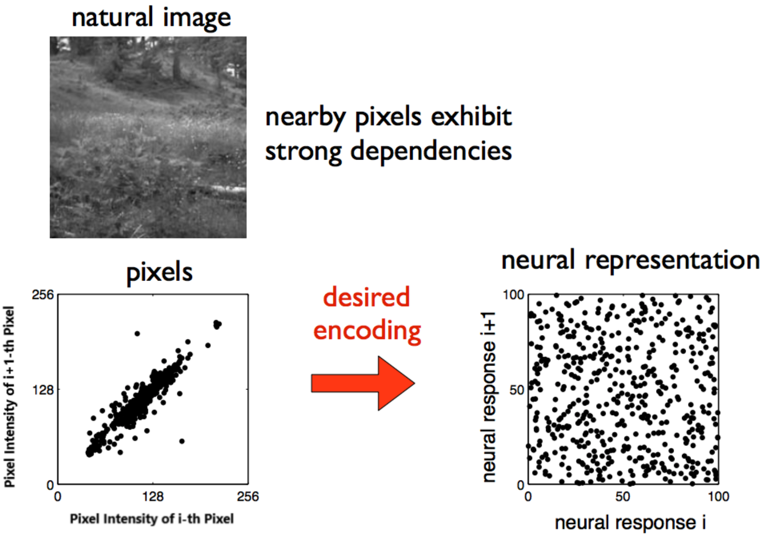
[1] [Loh and Bartulovic, 2014]

- Define Mutual Information 'I' as

$$I(\boldsymbol{X}, \boldsymbol{Y}) = H(\boldsymbol{Y}) - H(\boldsymbol{Y}/\boldsymbol{X})$$

  where H(**Y**) is the entropy of the response **Y**, and H(**Y**/**X**) is the noise entropy
- Our visual system tries to maximize the mutual information. It does this by making the neuronal responses as Independent of each other as it can so as to transmit the maximum 'Information'.

An Intuitive Approach



2

[2] Jonathan Pillow's Lecture Slides

Why understand Spatial Phase Structure?

- Phase of an image contains more perceptual information(edges and bars) than it's amplitude.[Oppenheim and Lim, 1980]

Original Image

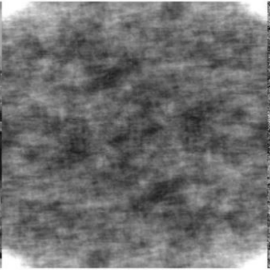
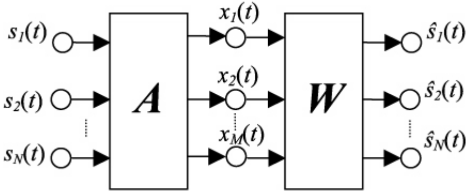


Image Reconstructed with original amplitude and random phase

Image Reconstructed with original phase and random amplitude

Independent Component Analysis(ICA) [Hyvärinen and Oja, 2000]

- One of the most widely used techniques for Blind Source Separation(BSS)[Cichocki and Amari, 2002]
- Solves the problem of BSS under the two assumptions
  1. The sources are Independent
  2. The sources are Non-Gaussian



[3]

- ICA aims to infer the de-mixing matrix 'W' and the sources 'S' under the above mentioned assumptions.

[3] Image Source

ICA and Natural Image Analysis

- Natural Scenes can be represented as a linear combination of **Features** and **Responses**

$$X = \sum_{i=1}^{N} s_i A_i$$

  Here, s_i represent the 'Responses' and A_i represents the 'Features'. In Matrix form, the above can be rewritten as

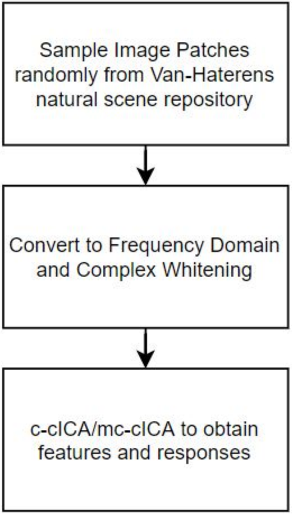
$$X = AS \; ; \; S = A^{-1}X = WX$$

- The above representation is very similar to the problem of estimation of the matrix 'W' and the vector 'S', which is what we do in ICA as well!

Complex Valued ICA

- Natural Scenes have significant perceptual information in their 'Phase Structure'. However, most conventional ICA models neglect the phase and only working with the image amplitudes.
- Local Phases of Natural Scenes can be better described by a mixture of phase distributions rather than a uniform one [MaBouDi et al., 2016]
- 'Complex ICA(cICA)' takes into account these non-uniform phase distributions as well by allowing the involved Matrices and Vectors to take on complex values as well

Proposed Algorithm



Flow of the proposed algorithm

Sampling and Pre-processing

- First, we randomly sample image patches from Hans Van Hateren's repository of Natural Scenes [Van Hateren and van der Schaaf, 1998].
- We take 'T' such samples. After obtaining the image patches of size $N \times N$, we obtain the Complex Representation of the images by taking them to the frequency domain using the Discrete Fourier Transform. We then vectorize the image to obtain a vector, which is our observation
- We then whiten(decorrelate) the data using the Complex Whitening Transform $W = (cov(X))^{-0.5}$

$$X^{obs}_{white} = WX^{obs}$$

Modelling the Density Functions

- We know that

$$X = AS \qquad ; \qquad S = A^{-1}X = WX$$

- If we know the density function $p_s(S)$, we can obtain the density of X as

$$p_X(X) = det(\overline{W})p_s(S)$$

Where $\overline{W}$ is the Jacobian of the Linear Transformation. Now, from the efficient coding hypothesis, we want the responses S to be independent of each other to maximize information transfer. Thus, $p_s(S) = \prod_{i=1}^{N} p_{s_i}(s_i)$, hence

$$p_X(X) = det(\overline{W})\prod_{i=1}^{N} p_{s_i}(s_i)$$

- 's' is a complex variable, so it can be written as $s = re^{j\phi}$ Where, $r = |s|$ and $\phi = tan^{-1}(\frac{real(s)}{imag(s)})$
- The density function of s can be written in terms of the density functions of r and ϕ as

$$p_{s_i}(s_i) = \frac{p(r_i, \phi_i)}{r_i}$$

Here, we assume that r_i and s_i are **independent** thus,

$$p(r_i, \phi_i) = p_{r_i}(r_i) p_{\phi_i}(\phi_i)$$

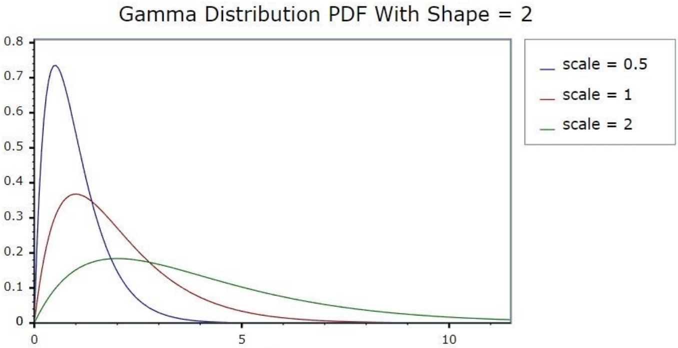
Thus,

$$p_X(X) = det(\overline{W}) \prod_{i=1}^{N} \frac{p_{r_i}(r_i) p_{\phi_i}(\phi_i)}{r_i}$$

- We model 'r' as a gamma distribution

$$p_{r_i}(r_i) = \beta_i^2 r_i e^{-\beta_i r_i}$$

With β as scale parameter and shape parameter $= 2$



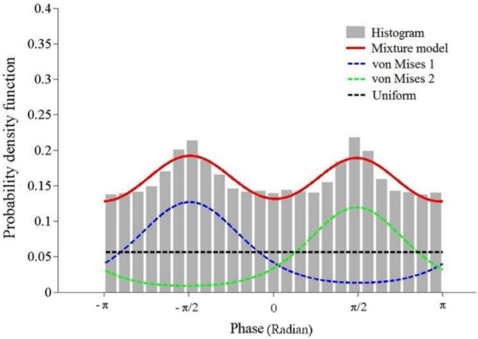
Plot of the Gamma Density Function[4]

[4]Image Source

- The phase is modelled as a mixture model based on the work in [MaBouDi et al., 2016]

$$p_{\phi_i}(\phi_i : \kappa_i) = \frac{1}{3\pi I_0(k_i)} cosh(k_i cos\phi_i) + \frac{1}{6\pi}$$

Here, I_m is Bessel function of first kind of order m



Plot of the Mixture Model Density Function[5]

[5]Image Source

- The Log-Likelihood is defined as

$$l(W, \beta, \kappa; X^{obs}) = \sum_{t=1}^{T} \sum_{i=1}^{N} log(p_{s_i}(W_i X^t)) + Tlog(det(\overline{W}))$$

- We use Maximum Likelihood Estimation for both the methods. The Likelihood is maximized using Gradient Ascent.

$$\boldsymbol{\beta} = \boldsymbol{\beta} + \eta \frac{\partial l(\boldsymbol{W}, \boldsymbol{\beta}, \boldsymbol{\kappa})}{\partial \boldsymbol{\beta}}$$

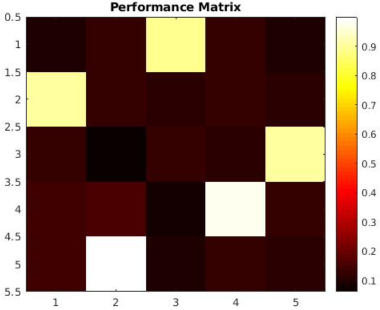
$$\boldsymbol{W} = \boldsymbol{W} + \eta \frac{\partial l(\boldsymbol{W}, \boldsymbol{\beta}, \boldsymbol{\kappa})}{\partial \boldsymbol{\beta}}$$

$$\boldsymbol{\kappa} = \boldsymbol{\kappa} + \eta \frac{\partial l(\boldsymbol{W}, \boldsymbol{\beta}, \boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}}$$

Here, η represents the learning rate

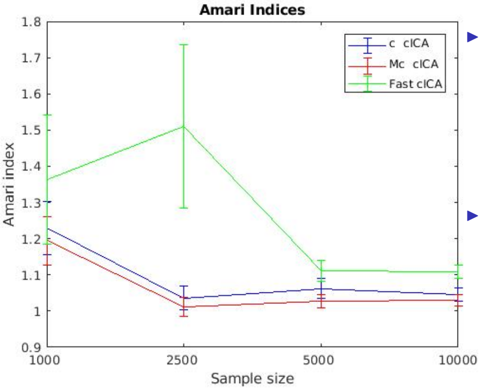
Results

1. Performance on Simulated Dataset :
   - An Artificial Dataset was generated by mixing 5 independent complex valued source signals sampled from the mixture model and using a random invertible matrix 'A'.
   - The de-mixing matrix 'W' was estimated. Ideally, *WA* should be a permutation of the Identity matrix for perfect separation.



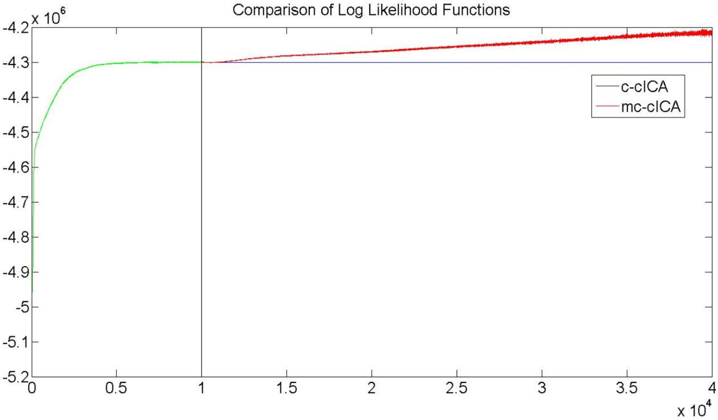
- The Amari Index was used quantify the quality of separation.

$$AI = \sum_{m=1}^{N} (\sum_{n=1}^{N} \frac{|P_{m,n}|}{\max_k |P_{m,k}|} - 1) + \sum_{n=1}^{N} (\sum_{m=1}^{N} \frac{|P_{m,n}|}{\max_k |P_{k,n}|} - 1)$$



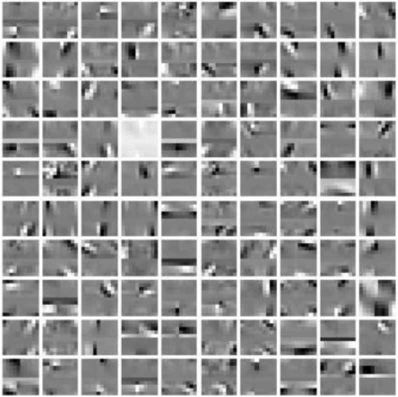
- Along with the mixture model(Mc cICA), two other models were compared :
  1. Uniform Phase(c cICA)
  2. Fast cICA
- Lower the Amari Index, better the quality of separation. Amari Index is zero if the Performance Matrix is a permutation of the Identity Matrix.s

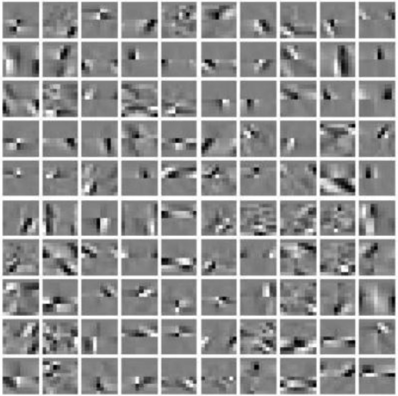
2. Performance on Natural Images
   - The model is initially trained using a uniform phase prior(c-cICA) upto 10000 iterations. Then, the performance of the Bi-Modal phase(mc-cICA) is compared with the uniform phase for the next 30000 iterations



   - The log-likelihood for the c-cICA model has saturated, but it is increasing for the mc-cICA model.

- Obtained Basis Functions - These are the Receptive Fields obtained





Future Work

1. Use Different Optimization Techniques like Conjugate Gradient, Natural Gradient[Amari, 1998] etc. to speed up the convergence.
2. Extend the model to analyze the statistics of Natural Sounds [Mlynarski, 2013].
3. Establish a Bayesian Framework for estimation in the presence of observation noise.
4. Stack multiple layers together to form a Hierarchical model for estimation.

References I

[Amari, 1998] Amari, S.-I. (1998).
Natural gradient works efficiently in learning.
Neural computation, 10(2):251–276.

[Barlow, 1961] Barlow, H. B. (1961).
Possible principles underlying the transformations of sensory messages.

[Cichocki and Amari, 2002] Cichocki, A. and Amari, S.-i. (2002).
Adaptive blind signal and image processing: learning algorithms and applications, volume 1.
John Wiley & Sons.

[Hyvärinen and Oja, 2000] Hyvärinen, A. and Oja, E. (2000).
Independent component analysis: algorithms and applications.
Neural networks, 13(4-5):411–430.

[Loh and Bartulovic, 2014] Loh, L. K. and Bartulovic, M. (2014).
Efficient coding hypothesis and an introduction to information theory.

[MaBouDi et al., 2016] MaBouDi, H., Shimazaki, H., Amari, S.-i., and Soltanian-Zadeh, H. (2016).
Representation of higher-order statistical structures in natural scenes via spatial phase distributions.
Vision research, 120:61–73.

References II

[Mlynarski, 2013] Mlynarski, W. (2013).
Sparse, complex-valued representations of natural sounds learned with phase and amplitude continuity priors.
arXiv preprint arXiv:1312.4695.

[Oppenheim and Lim, 1980] Oppenheim, A. V. and Lim, J. S. (1980).
Importance of phase in signals.
Technical report, MASSACHUSETTS INST OF TECH CAMBRIDGE RESEARCH LAB OF ELECTRONICS.

[Statistics, 2009] Statistics, N. I. (2009).
A probabilistic approach to early computational vision.

[Van Hateren and van der Schaaf, 1998] Van Hateren, J. H. and van der Schaaf, A. (1998).
Independent component filters of natural images compared with simple cells in primary visual cortex.
Proceedings of the Royal Society of London B: Biological Sciences, 265(1394):359–366.