Estimating MRF Potentials for Natural Images

Urs Köster, Jussi T. Lindgren and Aapo Hyvärinen

Neuroinformatics Group, Basic Research Unit, Helsinki Institute for Information Techonlongy, University of Helsinki, Finland

USE MARKOV RANDOM FIELDS TO GENERALIZE ICA FOR ARBITRARILY LARGE IMAGES

ABSTRACT

Markov Random Field (MRF) models with potentials learned from the data have recently received attention for learning the low-level structure of natural images.

- A MRF provides a principled model for whole images, unlike ICA, which can in practice be estimated for small patches only.
- Learning the filters in an MRF paradigm has been problematic in the past since it required computationally expensive Monte Carlo methods.
- We show how MRF potentials can be estimated using Score Matching.
- With this estimation method we can learn filters of size 12×12 pixels, considerably larger than traditional "hand-crafted" MRF potentials.
- We analyze the tuning properties of the filters in comparison to ICA filters, and show that the optimal MRF potentials are similar to the filters from an overcomplete ICA model.

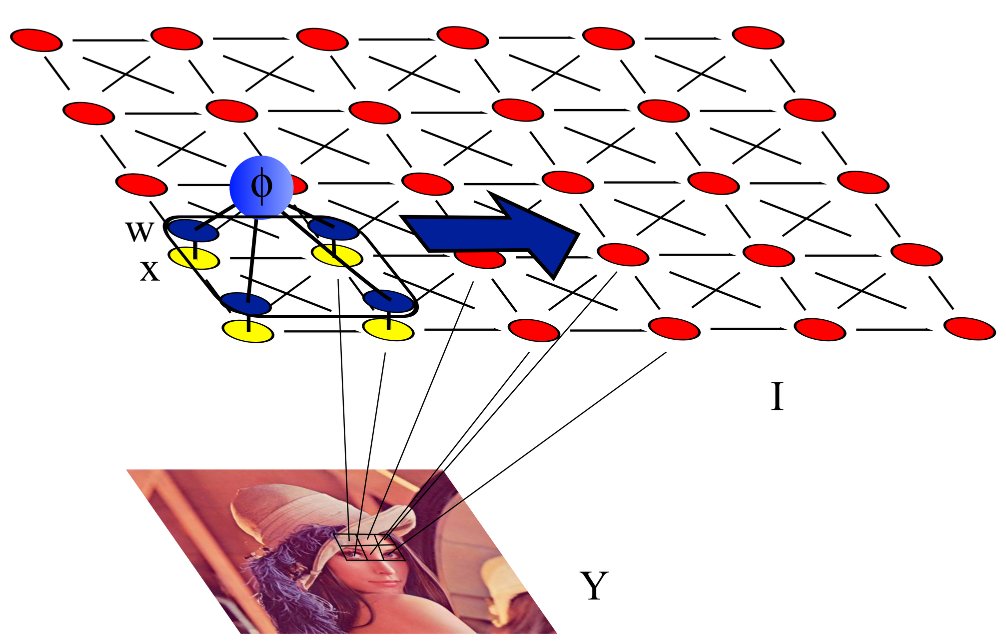

Figure 1: Sketch of a Markov Random Field:

- The MRF has maximal cliques of size 2×2 pixels; one clique $\mathbf{x}_i$ is shaded in yellow.
- Each unit of the field is associated to a pixel of the underlying image Y.
- The potential energy ϕ for each clique is defined as a function of the inner product of the image patch corresponding to the clique with a bank of filters of the same size as the clique, $V(\mathbf{x}) = \phi(\mathbf{w}^T\mathbf{x})$.
- This is visualized by the blue filter vector $\mathbf{w}$ that is scanned over the whole image, and the product is computed with each clique.
- In general there will be several filters in a filter matrix W, but for visualization purposes only one is shown.

MODEL

In contrast to ICA, the energy of a MRF is given by a convolution of the image I with potential functions U_k

$$V(\mathbf{I}, \theta) = \sum_{k,x,y} \phi\left(\mathbf{U}_k * \mathbf{I}\right) = \sum_{k,x,y} \phi\left(\sum_{x',y'} \mathbf{U}_{k,x',y'} \mathbf{I}_{x-x',y-y'}\right) \tag{1}$$

where the convolution (denoted by $*$) runs over pixels indices x and y. The elementwise nonlinearity ϕ gives the energy of the cliques of the field. As it is customary in ICA to work on whitened data, we insert a whitening filter $\mathbf{Q}$ in the convolution so it becomes $V(\mathbf{I}, \theta) = \sum_{k,x,y} \phi(\mathbf{U}_k * \mathbf{Q} * \mathbf{I})$.

SCORE MATCHING IS USED FOR ESTIMATING THE ENERGY BASED MODEL EFFICIENTLY

ESTIMATION

- We estimate the model using Score Matching, which works on the non-normalized distribution.
- We rewrite the convolution $\mathbf{I} * \mathbf{U}_k = \mathbf{X}\mathbf{w}_k$ where $\mathbf{X}$ is a matrix containing vectorized patches $\mathbf{x}_i$ from the image, and $\mathbf{w}_k$ is a vectorized filter.
- Similarly we write $\mathbf{X}_j$ as a subset of $\mathbf{X}$ containing only those patches which include the image pixel I_j.

The energy of the model is given by:

$$V(I, \theta) = \sum_{k,i} \phi\left(\mathbf{w}_k^T \mathbf{x}_i\right) \tag{2}$$

Where the sum is over the i patches contained in the matrix $\mathbf{X}$. The Score Function $\Psi_j = \frac{\partial V}{\partial \mathbf{I}_j}$ gives the Score Matching objective J

$$J = \sum_j \left(\frac{1}{2}\Psi_j^2 + \Psi_j'\right), \quad \frac{\partial J}{\partial \mathbf{w}_k} = \sum_j \left(\Psi_j \frac{\partial \Psi_j}{\partial \mathbf{w}_k} + \frac{\partial \Psi_j'}{\partial \mathbf{w}_k}\right) \tag{3}$$

Using this notation we can compute the score function w.r.t. to the image pixels $\mathbf{I}_j$

$$\Psi_j = \frac{\partial V}{\partial \mathbf{I}_j} = \frac{\partial}{\partial \mathbf{I}_j} \sum_{k,x,y} \phi\left(\mathbf{U}_k * I\right) = \sum_k \tilde{\mathbf{w}}_k^T \phi'\left(\mathbf{X}_j \mathbf{w}_k\right) \tag{4}$$

$$\Psi_j' = \frac{\partial^2 V}{\partial \mathbf{I}_j^2} = \sum_k (\mathbf{w}_k \odot \mathbf{w}_k)^T \phi''\left(\mathbf{X}_j \mathbf{w}_k\right) \tag{5}$$

We denote elementwise multiplication of vectors by $\odot$, and $\tilde{\mathbf{w}}$ indicates reversal of the order of elements in a vector.
The gradient of the objective function is now easily obtained from the gradients

$$\frac{\partial \Psi_j}{\partial \mathbf{w}_k} = \tilde{\phi}'\left(\mathbf{X}_j \mathbf{w}_k\right) + \mathbf{X}_j\left[\tilde{\mathbf{w}}_k \odot \phi''\left(\mathbf{X}_j \mathbf{w}_k\right)\right] \tag{6}$$

$$\frac{\partial \Psi_j'}{\partial \mathbf{w}_k} = 2\tilde{\mathbf{w}}_k \phi''\left(\mathbf{X}_j \mathbf{w}_k\right) + \mathbf{X}_j\left[\tilde{\mathbf{w}}_k \odot \tilde{\mathbf{w}}_k \odot \phi'''\left(\mathbf{X}_j \mathbf{w}_k\right)\right] \tag{7}$$

Thus the Score Matching objective can easily be optimized by gradient descent.

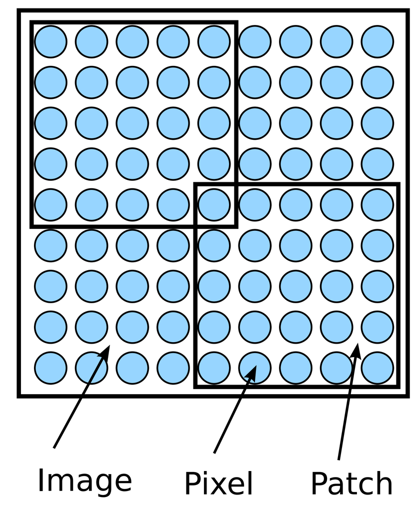

Figure 2: An image size of $2n-1 \times 2n-1$ is sufficient to realize all possible overlaps of the filter with a particular image pixel. By constraining the convolution to include only the "valid" terms, the gradient is exactly the same as for infinitely large images.

RESULTS

- In most **classical MRF** work, the potentials that were used were of rather small size such as 3×3 pixels and typically chosen to be directional derivatives. The larger MRF filters we estimated are very similar to ICA filters in appearance, being localized Gabor-like filters with tuning for spatial frequency, phase and orientation.
- To analyze the similarity between the two models further, we **fit Gabor functions** to the filters so we can analyze their tuning properties. We used a least squares fit to parametrize the filters in terms of length and width, frequency, phase and orientation.
- Histograms of the size and frequency distribution for the three models: The **complete ICA** model produces very localized filters which cover a relatively narrow band of frequencies. Both **overcomplete ICA** and the **MRF** give slightly larger filters with a broadened distribution of frequencies.
- Distributions for the MRF and ICA are somewhat different, but filters for **overcomplete ICA are more similar to the MRF**. This suggests that there are no fundamental differences between the filters obtained from the two models.

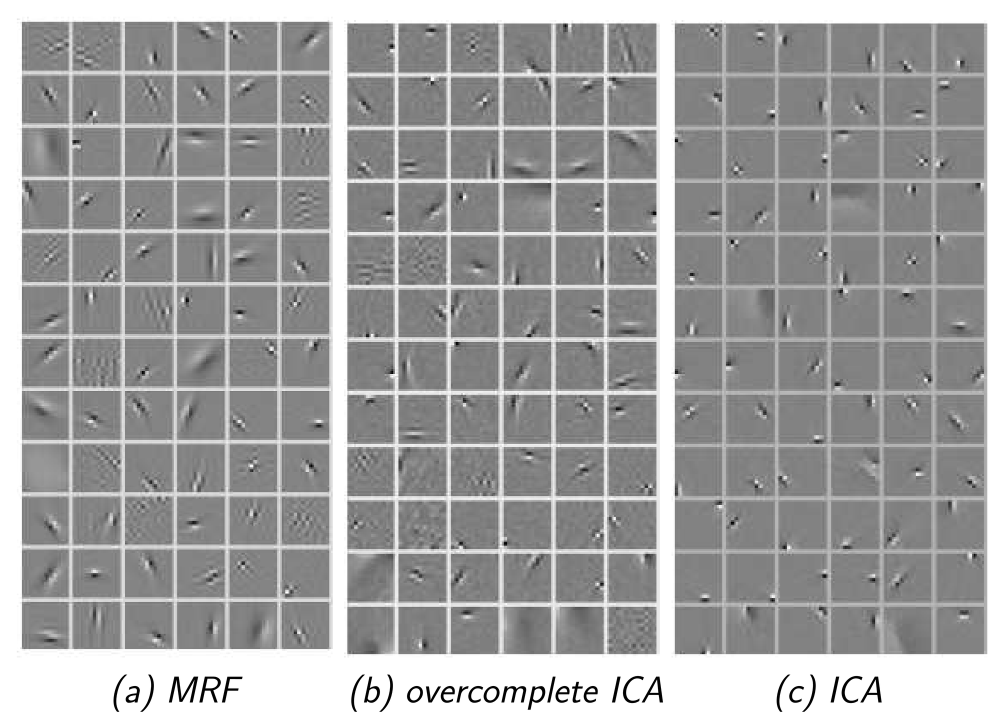

Figure 3: Comparison of the 12×12 filters for the different models. The complete and overcomplete ICA models shows the well-known Gabor like filters. The MRF potentials for 36×36 image patches are very similar, sharing the properties of localization and tuning for spatial frequency, phase and orientation.

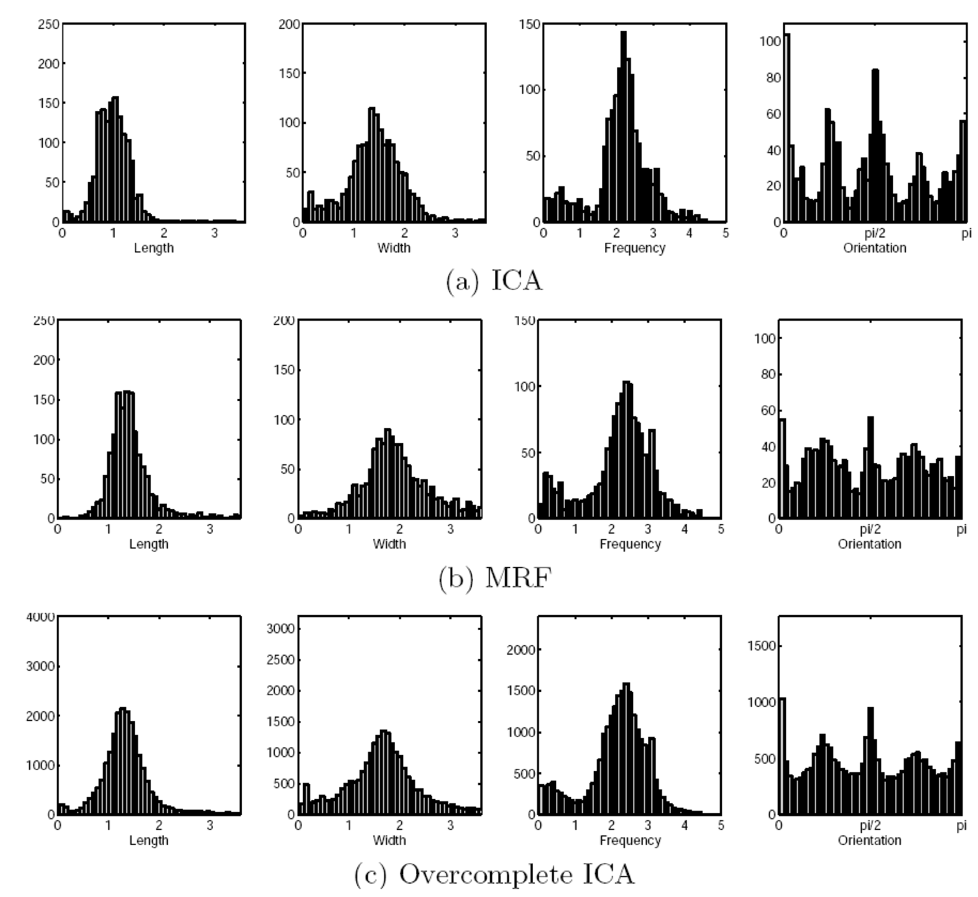

Figure 4: Tuning of ICA (top), MRF and overcomplete ICA (bottom) for 12×12 image patches. We show the size (length and width in pixels) of the Gaussian envelope of the Gabors we fit, and the distribution of frequencies (rad per pixel). Additionally, we show the distribution of orientations, which is clearly not uniform in both cases.

INTERPRETATION

RELATION TO PREVIOUS WORK

- Estimating optimal MRF potentials from natural images has previously been attempted by Roth and Black (2005), making use of Contrastive Divergence (Hinton 2002).
- Filters have a very different appearance, disjoint and distributed over the whole image patch, but small patch size (3×3 and 5×5).

MRF AS A SPECIAL CASE OF OVERCOMPLETE ICA

- It is possible to view the MRF model as a highly overcomplete ICA model with some additional constraints.
- The convolution in (1) can be interpreted as keeping the image fixed, and multiplying it with the filters in different positions.
- The resulting "filters" would be shifted copies of the original filters at different positions in the image and padded with zeros.
- Difference to ICA: The size of the filters is restricted to be much smaller than the image.
- Consequence for filters: While an ICA basis may contain nearly identical filters in different positions, this should not be the case with the MRF model.

CONCLUSION: We have shown that *it is possible to learn the filters used in a non-Gaussian Markov Random Field*. Learning is based on score matching and leads to Gabor-like filters. This gives a well-defined *probabilistic model of whole images* instead of just small patches. Possible applications are in *image denoising and filling-in*.

Urs Köster is supported by a scholarship from the *Alfried Krupp von Bohlen und Halbach-Stiftung* foundation