Seminar on Neuroinformatics 2: Natural image statistics Report

Teppo Niinimäki

May 15, 2010

The subject of the seminar was natural image statistics and the aim was to explore a theory of early human vision and image processing. The focus was on modeling the statistical structure of natural images and learning *features*, properties which can be used to describe images.

A useful tool in the analysis of image processing systems is *frequency-based representation*, that is, representing an image as a weighted sum of phase-shifted sine (or cosine) waves with predefined frequencies. So instead of storing intensities for individual pixels we store the phases and the amplitudes of different frequencies present in the image. Related useful tools are *Gabor functions*, which are sort of spatially localized sinusoids.

Most information processing in the brain is done by cells called *neurons*. They communicate by changins the firing rate of short electrical impulses, which they are continuously sending each others. We mostly want to model cells in the primary visual cortex which is the area performing early stage visual processing. There are two interesting types of neurons in that area: *simple cells* and *complex cells*. Simple cells respond roughly linearly to localized lighting pattern of the image and those can quite acceptably be modeled using Gabor functions. Complex cells are otherwise quite similar but, unlike the simple cells, they are relatively invariant to the spatial phase of the stimulus. Complex cells are typically modeled by energy model, where the response is computed as a square sum of responses of two Gabor functions (or simple cells) with 90 degree phase shift.

When dealing with image processing, it's useful to understand some basic multivariate probabilistics and statistics. We represent the image as a vector of random variables, each of typically corresponds to one pixel. Instead of using whole images we normally use small patches from images. We assume that the image vector has some underlying multivariate probability

distribution, which we would like to model – the more accurately, the better. Typically there are strong non-trivial dependencies between the elements of the image vector. Some of those dependencies can be represented as the covariance matrix of the vector.

Principal component analysis (PCA) is a classical method of analyzing multi-dimensional data. Finding principal components can be seen as selecting variance-maximizing features. The first principal component of a vector $\mathbf{z}$ is the feature $s_1 = \mathbf{w} \cdot \mathbf{z}$ that has maximum variance with the constraint $||\mathbf{w}|| = 1$. Further principal components $(s_2, s_3, \ldots)$ do also maximize variance, but with the additional restriction that they must be orthogonal to the previous ones. The features found by PCA are typically not interesting by themselves. Instead PCA is normally used as a preprocessing tool. Main uses are *dimension reduction* (reducing the number of elements in image vectors), *whitening* (transforming image pixels to uncorrelated variables with unit variance) and *anti-aliasing* (alleviating aliasing problems by removing the highest frequencies).

Models for simple cells

Random variable is said to be *sparse* if it's most of the time very close to zero and gets clearly non-zero values only occasionally. We measure the sparseness of feature s by $\mathrm{E}[h(s^2)]$, where $h(u)$ is some convex function (for instance $-\sqrt{u}$). Sparseness is completely unrelated to variance and thus measures different aspects of the variable. We can learn features maximising their sparseness while restricting them to be uncorrelated and to have unit variance. In the case of natural images, there are two types of sparseness: A single feature is sparse if it's active only in few images and a representation is sparse if given image is represented by only a small number of active features. Luckily, the two give approximately the same function to maximize. More interestingly, learning linear features by maximising sparseness yields results that resemble closely receptive fields of simple cells.

The requirement of sparseness leads to a new data analyzing method, *independent component analysis* (ICA). The ICA is based on a generative model, which says that image patches are generated as a linear superposition of some features A_i, or in matrix notation $\mathbf{I} = \mathbf{A}\mathbf{s}$, where $\mathbf{A}$ is a square matrix containing the features, $\mathbf{s}$ is a vector of feature weights and $\mathbf{I}$ is a resulting image. The idea can be reversed to calculate the feature weights for given image, $\mathbf{s} = \mathbf{W}\mathbf{I}$ where $W = A^{-1}$. It's also assumed that the feature weights are non-Gaussianly distributed and independent. Under these assumptions, finding the most likely features A_i for natural images leads to sparse features which are very similar to those obtained by maximising sparseness.

Regardless of the name and the assumption, the estimated independent components are not really statistically independent. This paradox is a result from the fact that real data from natural images can't be completely described by the simple ICA model. However, a large propotion of this dependency between the components can be explained by the concept of a *global variance variable* which controls the overall activity level of the components on given image patch. Normalization of this variance, and thereby weakening the dependencies between independent components, leads to contrast gain control. An intuitive physical interpretation for this could be canceling the effect of changing lighting conditions. Instead of one global variance one could also have more variance variables, one for each independent component. The resulting nonlinear model is an elementwise product of two simpler submodels, each of which follow an ICA model. This can be seen for example as modeling the process of combining illumination and reflection information.

Models for complex cells and higher levels

Complex cells pool the information from simple cells. Similar to ICA for simple cells, a corresponding method for complex cells is *independent subspace analysis* (ISA). In ISA simple linear features are organized into equally sized groups or subspaces which are treated as second order non-linear features. The strength of the subspace feature is choosed to be the square root of the sum of the squares of the simple features in that subspace which, in essence, is combined energy of those features. This corresponds to the way complex cells pool outputs of simple cells. Features can then be learned by maximizing sparseness of these energy detector features. Unlike in ICA, in ISA model the components in the same subspace are not independent (although restricted to be uncorrelated). For natural images this approach results in subspace features that are phase-invariant but still selective for frequency, orientation and location.

While ISA pools feature into distinct subspaces, another approach to organize simple features is to arrange them topographically. In the visual cortex the cells have indeed a very specific spatial organization, in which cells with similar receptive fields are close to each others. In *topographic ICA* model the linear features are arranged on a two-dimensional grid where features that are far away from each others are independent. On the other hand the features in local neighbourhoods are dependent and correspond to subspaces in ISA. Again these overlapping neighbourhoods are considered as higher order features (corresponding complex cells) which pool local energies from simple features. Estimation of topographic ICA leads to emergence of topography where adjacent features have similar orientation, frequency and

location, just like observed in human brain.

A simple way to try to predict what happens after complex cells, is to fix the model for simple and complex cells and to add a third layer of cells which take the output of complex cells as input. The features for the third layer are then learned by doing a simple ICA of the complex cells outputs. On natural images this leads to higher-order features, which group together collinear complex cells forming a longer contour. In addition to collinearity, they also pool cells for different frequencies. One result from this pooling could be a more realistic edge representation, and thus a better edge detection.

Typically, for instance in ICA, the number of features is limited by the number of pixels in image patch. However, physiologically the number of simple cells is much greater than the number of cells in retina. This problem can be addressed by using either linear generative model with *overcomplete basis* or simpler *energy base model*, both of which are extensions of ICA model where the number of features is not limited.

In addition to feedforward processing where signals travel from lower levers to higher leves, there are top-down *feedback connections* and even *lateral connections* inside single layers in the brain. The Bayesian inference on generative models is a solution to model these phenomena. Updating lower level features according to the feedback leads to *tresholding* which inhibits the activity of lone features not fitting well into the big picture. And using of a certain type of overcomplete basis produces results results similar to *end-stopping*, a phenomenon related to lateral interaction.

Conclusions

The theory of early human visual processing is evolving rapidly. Today, the basic behavior of simple cells and complex cells of the V1 area in the visual cortex is understood to some extent and can be modeled satisfyingly. Many interesting results also exist for some types of interactions and other known slightly more complicated phenomena. However, the simple and complex cells are only the very first step in sophisticated processing occuring in visual cortex. There are a lot of open questions concerning higher levels and multilayer modeling, which would naturally be one of the next steps in understanding more deeply the visual processing taking place in the brain.