# Learning High-level Independent Components of Images through a Spectral Representation

J.T. Lindgren
Department of Computer Science
University of Helsinki, Finland
jtlindgr@cs.helsinki.fi

Aapo Hyvärinen
HIIT Basic Research Unit
University of Helsinki, Finland
aapo.hyvarinen@helsinki.fi

**Abstract**

Statistical methods, such as independent component analysis, have been successful in learning local low-level features from natural image data. Here we extend these methods for learning high-level representations of whole images or scenes. We show empirically that independent component analysis is able to capture some intuitive natural image categories when applied on histograms of outputs of ordinary Gabor-like filters. This can be taken as an indication that maximizing the independence or sparseness of features may be a meaningful strategy even on higher levels of image processing, for such advanced functionality as object recognition or image retrieval from databases.

## 1 Introduction

There seems to be a rather general consensus both in vision research and pattern recognition on the low-level mechanisms for feature extraction in natural images. Linear filters similar to the Gabor family have interesting analytical properties [2] and similar filters have emerged by data-driven statistical methods from natural images [8, 10, 5]. Existence of equivalent mechanisms in biological vision has been known since the classic work of Hubel and Wiesel [3].

Much more controversial is the question of what should be done with the filtered images to facilitate higher level functions such as object recognition, content-based image retrieval and finally, image understanding.

In this paper we show that it is possible to learn some intuitive natural image categories (categorical attributes) by applying data-driven, unsupervised methods on the histograms of outputs of low-level features in high-resolution

natural images. Our method applies independent component analysis (ICA, see [5]) on the histograms of outputs of linear filters — called independent spectral representations in [6].

We present empirical results showing that the method is able to learn features that typically respond highly to a single intuitive natural category and are mostly inactive on images not belonging to the same category.

The rest of this paper is organized as follows. Section 2 quickly outlines ICA, the spectral representations and our framework for learning the categorical attributes. In section 3 we outline our experimental design. Section 4 describes results on natural images. Finally, section 5 concludes with some future directions.

# 2 Learning natural image categories

## 2.1 Independent and sparse components

Maximizing sparsity or independence has turned out to be an useful optimization criteria for learning low-level feature extractors that are similar to those found in natural visual systems [8, 5]. Since sparse coding and independent component analysis are very similar, we will concentrate on ICA in this presentation.

ICA assumes that the observed data matrix $\mathbf{X}$ is generated by a model

$$\mathbf{X} = \mathbf{AS}, \tag{1}$$

where $\mathbf{S}$ denotes a matrix of the hidden sources and $\mathbf{A}$ is some unknown mixing matrix. The goal of ICA is to recover $\mathbf{S}$ by estimating $\mathbf{W} = \mathbf{A}^{-1}$ so that $\mathbf{S} = \mathbf{A}^{-1}\mathbf{X} = \mathbf{WX}$. ICA tries to do this by maximizing the statistical independence of the rows of $\mathbf{S}$. For a more thorough account, see e.g. [5]. When ICA is applied to small natural image patches, Gabor-like filters emerge as rows of $\mathbf{W}$ [8, 5]. In particular, the features are linear filters that are 1) oriented 2) localized in space and 3) bandpass.

A tempting approach would be to try to find sparse or independent representations from a large dataset of full-size natural images, instead of applying the methods to small windows sampled from the dataset. One might hope that this would reveal components of more complicated structure. However, it is likely that the features would only be re-scalings of the original features, since ICA gives quite similar features on different image sizes. Moreover, doing independent component analysis on full size images would be prohibitively expensive computationally and require massive amounts of data.

Another approach that is not fruitful is to do ICA on the independent components of natural image patches. This means simply stacking two linear transformations. Nothing is gained in this way, since the first linear transformation already gave the most independent features.

The approach we take in this paper is to examine how a statistical description

of low-level image features could be used as a starting point for learning higher-order features from images.

### 2.2 Spectral representations

Liu and Cheng [6] argued that if the different marginal statistics derived from an image patch are independent, they provide a low-complexity representation for the full joint distribution of the patch content. Subsequently, they proposed to compute an independent spectral representation (ISR) to characterize a single image.

ISR is basically a set of marginal statistics calculated from outputs of linear filters on an image, with the property that the marginals are assumed to be independent by construction. A "spectral" representation is created by convolving an image separately with filters in a chosen filter bank. For each of the filter outputs, a histogram is estimated as one (multidimensional) marginal statistic. The histograms are then concatenated to form the spectral representation for the image. In ISR, the filters are typically estimated by ICA and their responses are assumed to be independent. Although maximizing the independence is the goal of ICA, true independence is not likely to be obtainable from natural image data. However, this may be a useful first approximation, and even if independence is not attained, the results can often be interpreted as maximally sparse coding [8, 5].

### 2.3 ICA of low-level feature histograms

We propose here that a combination of independent spectral representations and another ICA phase yields interesting attributes that characterize whole images or scenes. This is different from the classic application of ICA which is supposed to yield only very low-level, local features.

Basically, we compute the independent components of the independent spectral representations of each image. ISR of an image consists of histograms of the responses of low-level filters (typically given by ICA). We concatenate the histograms of all filters for each image, thus obtaining a representation vector for each image. Then we perform ICA on these data.

## 3 Experimental design

We applied our method — first ordinary ICA, followed by ICA on the histograms of the outputs of the first ICA — on a well-known set of natural images. Due to space constraints we give only a relatively high-level description of our method[1].

We used the natural image dataset provided by van Hateren and van der Schaaf [10]. The dataset contains approximately 4000 gray-scale images with resolution $1536 \times 1024$. We used the "deblurred" versions. The images show

[1] To ensure replicable results, source code is available at `http://www.cs.helsinki.fi/u/jtlindgr/stuff/`

miscellaneous shots of suburbs and countryside. Shrubs, fields, woods, buildings etc. are featured, photographed at various distances and lightings. For a more in-depth description of the image set and its calibration, see [10].

Although the original ISR construction [6] uses multiple resolutions and window sizes, in this paper we use a simple, single-resolution single-size design that suffices for obtaining the shown results. Thus, for the first phase, we sampled 40000 image patches of size $12 \times 12$ to learn the filter bank $\mathbf{W}$. To perform ICA, we used the FastICA package [4]. We used tanh nonlinearity in symmetric estimation mode. The number of filters was reduced to 100 by considering only the 100 first principal components of the image patches.

Then we computed the output histograms for each image and each filter. First, we randomly selected 100 images to estimate the histogram bin centers. This was done by applying each filter $\mathbf{W}_{i*}$ to all the images, taking note of the minimum $min_i$ and maximum $max_i$ responses of the filter $\mathbf{W}_{i*}$ over the images. Next, we computed a 10-bin binning for each filter $\mathbf{W}_{i*}$ by dividing the interval $[min_i, max_i]$ to 10 equal-width bins. After the center locations were estimated, the independent spectral representation was computed for each of the 4000 images. In this representation, each 10-bin histogram was normalized by the $L_1$-norm.

For each image, the histograms of the different filter responses were concatenated to form a single vector. Thus, each image was characterized by a vector of dimension $10 \times 100 = 1000$. These vectors were normalized by the $L_2$-norm. The 1000-dimensional data was reduced by principal component analysis (PCA) to 50 dimensions.

Finally, we used ICA to learn 50 independent components of the data in this representation. Again, we used tanh nonlinearity in symmetric mode. However, the ICA settings here did not seem to make much difference. The method was robust against a variety of nonlinearities tried, e.g. derivative of gaussian and skewness. Different normalizations of the input data didn't alter the results significantly.

# 4 Results and analysis

It is quite difficult to visualize the obtained 50 independent components. We choose here to show the results in terms of the original images. The following figures represent images ordered by their activation of a single real-valued component (attribute). In ICA, both high (positive) and low (negative) activations of the components can be significant: we selected for each shown component the better tail. The shown figures were power-transformed by 0.6 to enhance visibility.

Figure 1 shows five categorical attributes related to open scenes. The attributes were selected manually from the set of 50. Although the images across the rows are similar, the groups are not identical. Likewise, figure 2 shows components responding to woods. The first three components seem to differ in the amount of light let through the foliage and the viewing distance used. The last

Figure 1: Five attributes related to open scenes. Each row shows the five images for which the attribute was most active.

two rows seem qualitatively different from the first three.

Finally, figure 3 shows some miscellaneous components. The component on the first row responds to clouds, but also to patches having some textural similarity. The following three components display reeds, flora and road surfaces. The last component shows man-made structures.

By looking at the figures shown, it could be suspected that similar "categories" might emerge from presenting the image data as grayscale histograms (simple distributions of pixel values). Figure 4 shows an example from an experiment where we performed ICA on 64-bin grayscale histogram data. Although the method does generate some reasonable attributes, the loss of all spatial information usually makes the results less intuitive, so they will not be considered further here.

We also tried to provide preliminary answers to two other questions raised by our results: 1) How many of the 50 features are meaningful to a human observer? 2) What is the role of the optimization for independence in the second phase, compared to, e.g., PCA? To address these questions, we first computed 50 principal components from the ISR data. Then, for both PCA and ICA, we created a display for both tails of each component: Each display contained $6 \times 6 = 36$ highest responding natural images in the respective tail. This totaled to $2 \times 2 \times 50 = 200$ displays. We presented the displays randomly and without identifiers to five naïve subjects and asked them to vote which of the displays were "reasonable". Then, we grouped the positive votes of each subject per method. On average, the subjects voted 33.6% and 25% of the displays to be reasonable, for ICA and PCA respectively. To compensate for the high variance in the total vote counts, we normalized the votes of each subject to sum to one. The $p$-value, testing for equality of means, for the normalized vote data was 0.047 according to a Kruskal-Wallis test. Thus, we can say that the subjects

Figure 2: Attributes related to woods.

rated the ICA results better with a statistically significant difference.

Perhaps surprisingly, the PCA results we got are similar to those presented by Torralba and Oliva [9]. Their method used PCA on Fourier-transformed low-resolution images. They found that the second principal component was able to order images so that one end of the activation range featured "open" images and the other "closed" (or cluttered) ones. The third principal component ordered natural images against images showing man-made structures. This hints at similarities in the two methods. Of course, the low-level ICA features are bandpass, and therefore the histograms are somewhat related to a Fourier representation. Our method, however, is applied on high-resolution images and benefits from high resolutions due to the histograms being better estimated when the resolution is high.

# 5 Discussion

We have presented a method for unsupervised, data-driven estimation of high-level features that appear to be related to natural image categories. Our construction has several parameters and technical details most of which seem uncritical. Perhaps the most significant of these is the estimation of the histograms and choosing proper binning for them. It is possible that an entirely different density estimator would be better here, as most of the marginal histograms in ISR have a particular, supergaussian shape. We performed some preliminary experiments where we tried to replace the histograms with other statistics. Especially, we tried order statistics and the first four moments of a distribution. These alternatives did not produce significantly better results.

Traditionally, feedforward architectures have been a convenient framework for pattern recognition [7]. However, evidence from biological systems shows

Figure 3: Miscellaneous attributes.

Figure 4: For comparison, an attribute computed from graylevels instead of ISR.

that at least natural pattern recognition uses feedback from higher levels to adjust lower-level functionality [1, 9]. Very relevant to our method are the arguments by Bar [1], suggesting that in biological visual systems the lower level might first supply the higher-level mechanism with a crude approximation of the scene, which is used to generate feedback to make adjustments at the lower level. A crude approximation might be based on the histograms of low-level features as we did here. Furthermore, the categorical features that we found could be useful for higher level mechanisms to aid in decision making as argued, e.g., in [9], who obtained results that are somewhat similar to ours, but using a different method.

An important direction of future work consists of more thoroughly characterizing what kind of properties the current method is able to capture. During our experiments with van Hateren's dataset we noted that many images do not respond highly to any of the estimated components, even though the images are perceptually reasonable. This may also be due to the strong reduction of dimension by PCA, which may lead to considerable loss of information. However, the reduction of dimension was necessary since the number of images must be much larger than the dimension in order for ICA to be possible. Experiments with much larger databases may alleviate this problem.

To conclude, we proposed a method for learning high-level features of whole

images or scenes. This was based on independent component analysis of histograms of low-level features. We showed empirical results that indicate the viability of the method.

# References


[1] M. Bar. A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, 15:600–609, 2003.

[2] J. G. Daugman. Uncertainty relations for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. of the Optical Society of America A*, 2:1160–1179, 1985.

[3] D. Hubel and T. Wiesel. Receptive fields of single neurones in the cat's striate cortex. *J. Physiol. (Lond.)*, 148:574–579, 1959.

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

[5] A. Hyvrinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley, 2001.

[6] X. Liu and L. Cheng. Independent spectral representations of images for recognition. *Journal of the Optical Society of America*, 20(7):1271–1282, 2003.

[7] D. Marr. *Vision*. Freeman, 1982.

[8] 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.

[9] A. Torralba and A. Oliva. Statistics of natural image categories. *Network: Computation in Neural Systems*, 14:391–412, 2003.

[10] 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.