Edge co-occurrences and categorizing images

Edge co-occurrences can account for rapid categorization of natural versus animal images

Laurent U. Perrinet\textsuperscript{1,2} and James A. Bednar\textsuperscript{4}

\textsuperscript{1} - InViBe, Institut de Neurociences de la Timone CNRS / Université de la Méditérrannée, Marseille, France.
\textsuperscript{2} - The Wellcome Trust Centre for Neuroimaging University College London, UK.
\textsuperscript{4} - Institute for Adaptive and Neural Computation, University of Edinburgh

Monday, 22nd of June, 2015

http://invibe.net/LaurentPerrinet/Publications/
PerrinetBednar15

A SIS © I3S seminar.

This work was supported by projects “BrainScaleS” (EU funding, grant number FP7-269921).
(hello) Hi, I am Laurent U. Perrinet and I work at the team "inference in Vision & Behavior" at the INT. My research interest is to put in relation visual functions and the statistics of natural images. I am therefore a computational neuroscientist, discovering the code used to efficiently represent images in the early visual system and the application to novel computational paradigms (sparsity, probabilities, prediction, hierarchical models).

(today) Today I will present a work done during a visit in Karl Friston’s team in London at the UCL: I will talk about the potential role of environment of animals as measured by edge statistics in understanding lateral connectivity in the primary visual cortex. Using these results, I will give some predictions on neurophysiological observations. This is joint work with James Bednar who is an expert in topographical models of cortical areas.

akno Thanks to the BrainScaleS project and the ANR for funding this project.
Oriented edges that constitute images of natural scenes tend to be aligned in co-linear or co-circular arrangements, such as when you follow the contours of these boulders: lines and smooth curves are more common than other possible arrangements of edges. See for example the work of Mariano Sigman on co-circularity in natural images (see Sigman, 2001).

The visual system appears to take advantage of this prior information, and human contour detection and grouping performance is well predicted by what is coined an "association field" (Field et al., 1993)...
Fig. 7. Predicted contour groups. (A) An example stimulus that contains a target contour. (B) The connected line segments show the groups obtained by applying a local grouping function (based directly upon the edge co-occurrence statistics) followed by a simple transitivity rule: if (edge element) binds to (edge element), and (edge element) binds to (edge element), then (edge element) becomes bound to (edge element).

5. Predictions

Is the local grouping function that the visual system uses for contour grouping consistent with the co-occurrence statistics of natural images? To address this question, we generated predictions for the psychophysical experiment.

Consider first the absolute co-occurrence statistics. To generate predictions using the edge co-occurrence data in Fig. 3C, it is necessary to introduce two parameters. The first is a 'tolerance' parameter, which assigns a relative probability to orientation differences around the maximum probabilities shown in Fig. 3C. This parameter produces a small family of local grouping functions, all of which are consistent with the data in Fig. 3C. The second parameter is a binding criterion, which is a threshold on the local grouping function: any pair of edge elements whose co-occurrence probability exceeds this criterion get bound together, otherwise they do not. An example of a thresholded local grouping function is shown in Fig. 3E. The line segments drawn in this diagram show all the specific combinations of distance, orientation difference, and direction that result in grouping to the reference element. Predictions for the contour detection task were obtained by combining a thresholded local grouping function with the simplest possible integration mechanism — a transitivity rule: if (edge element) binds to (edge element), and (edge element) binds to (edge element), then (edge element) becomes bound to (edge element) (Geisler & Super, 2000).

A predicted response ('first' or 'second' interval) was computed for each of the specific stimuli presented in each trial of the experiment, for each subject. The predictions for a given trial were generated as follows. First, we retrieved the specific list of line segments presented in the first and second intervals of the trial. Second, each line segment was compared with every other line segment, and a pair of line segments was bound together if they satisfied one of the geometrical relationships specified in the thresholded local grouping function (i.e. Fig. 3E). Third, the final groups were obtained by applying the transitivity rule over all the two-element groups formed in the second step. Fourth, and finally, we assumed that, in our two-interval forced choice task, observers selected the interval containing the longest group that was consistent with the possible contour locations. The result of this processing is demonstrated in Fig. 7B, where the connected elements show the final groups that are formed for the image in Fig. 7A. As can be seen, the longest group corresponds to the target contour.

The open circles in Fig. 6A–D show the predictions for all the data using the thresholded local grouping function in Fig. 3E, which was obtained from Fig. 3C by setting the tolerance parameter = 20° and the binding criterion = 0.05. As can be seen, the predictions are remarkably good (given only two free parameters). The value of the Pearson correlation coefficient for the predicted accuracy versus the observed accuracy is 0.87.

To generate predictions using the Bayesian co-occurrence statistics only one free parameter is required — the likelihood-ratio criterion. Recall that the rational decision rule is to group a given pair of edge elements together if the likelihood ratio exceeds this criterion. Thus, the optimal local grouping function is obtained by thresholding the likelihood ratio histogram shown in Fig. 3D. The resulting function for the best fitting value of the likelihood-ratio criterion (0.38) is shown in Fig. 3F. The psychophysical predictions are not shown here, but they are slightly better than those obtained using the absolute co-occurrence statistics. The value of the Pearson correlation coefficient for the predicted accuracy versus the observed accuracy is 0.89.
Introduction: linking neural structure to natural scenes

(Gestalt) ... ○this has been measured and quantified by Geisler in 2001 and using a measure of second-order statistics combined with an iterative grouping rule, they could reproduce diverse behavioral results at a global level, for instance here the link that is reported from a display of artificial edges (A) to what is reported as perceived (B). This thus gives a link between this local dependence present in natural images and the emergence in neural computations of some global Gestalt-like rules as implemented in the brain.
One possible candidate substrate for implementing such an association field in mammals is the set of long-range lateral connections between neurons in the primary visual cortex (V1), which could act to facilitate detection of contours matching the association field, and/or inhibit detection of other contours. To fill this role, the lateral connections would need to be orientation specific and aligned along contours, and indeed such an arrangement has been found in tree shrew’s primary visual cortex.

We show here one result of (Bosking et al., J Neurosci 17:2112-27, 1997) which overlay over a map or orientation selectivity the network of lateral connectivity originating from a group of neurons with similar orientations and position. There is a structure in this connectivity towards locality (more pronounced for site B) connecting iso orientations even on long ranges (A). This type of structure tends to wire together those neurons that have similar orientations, indicating a prior to colinearities.

Is there a match of these structures with the statistics of natural images? 1: Hunt & Goodhill have reinterpreted above data and shown that there is more diversity than that - 2) Some authors (Kisvarday, 1997, Chavane and Monier) even say it is weak or inexistent on a the scale of the area...
[Fischer et al., 2007]
This prior for smooth contours can be introduced in models:

(colin) ☯This is a typical assumption that the role of lateral interactions is to enhance the activity of neurons which are collinear: it is the so-called association field formalized in Field 93, as was for instance modeled neurally in the work from P. Series or in this version for computer vision: todo=describe...
Introduction: linking neural structure to natural scenes

Problem statement

Basically, all our knowledge on neuroanatomy come from laboratory-reared animals and...

... it is not yet known how these patterns develop as a result of visual experience, and if they adapt to be efficient wrt the statistics of natural scenes. To investigate this issue, we examined the properties of the visual environment of laboratory animals, to determine whether the observed connection patterns are more similar to the statistics of the rearing environment or of a natural habitat.

Specifically, we analyzed the co-occurrence statistics of edge elements in images of natural scenes, and compared them to corresponding statistics for images taken from within the rearing environment of the animals in the Bosking et al. (1997) study.

Here I show a picture taken from inside a typical cage as it would be seen from the animal. It is qualitatively quite different from natural images and it seems that there are more collinear edges. Our goal here is to quantitatively measure this difference and to give some predictions as to how this may play a role in neurophysiological experiments compared to what would be observed in wild animals.
Outline: Edge co-occurrences and categorizing images

Introduction: linking neural structure to natural scenes
Geisler et al, 2001
Bosking et al, 1997
Problem statement

Method: detection of edges
Geisler et al, 2001
Log Gabor representation
Sparse coding
Some examples of edge extraction

Results: natural vs. laboratory images
Second-order statistics
Quantitative difference using classification
Intermediate Summary

Categorizing animals in images from natural scenes
1. first, we will define a framework adapted to the computation of second-order edge statistics, using the detection of edges in natural images and laboratory images

2. then, we will show the results of extraction on both classes of images and show the observed statistics

3. Finally, we will summarize results and present some predictions and perspectives
Outline: Edge co-occurrences and categorizing images

Introduction: linking neural structure to natural scenes
- Geisler et al, 2001
- Bosking et al, 1997
- Problem statement

Method: detection of edges
- Geisler et al, 2001
- Log Gabor representation
- Sparse coding
- Some examples of edge extraction

Results: natural vs. laboratory images
- Second-order statistics
- Quantitative difference using classification

Intermediate Summary

Categorizing animals in images from natural scenes
So, in the part we will show the method that we used by showing

1. state-of-the-art for extracting second-order statistics in natural images,
2. detailing the dictionary of edges that we used,
3. the edge extraction algorithm
4. we will also define the databases that we used
Fig. 3. Statistical analysis of edge co-occurrence in natural images. For each image, edge elements were extracted, and then each edge element was compared with other edge elements. (A) Edge elements. Each red pixel in this image indicates the location of the center of a significant edge element; the orientations of the elements are not shown. (B) First edge cooccurrence property. The line segments show the most frequently occurring orientation difference for each given distance and direction from the central reference element. The color of a line segment indicates the relative probability. (C) Second edge co-occurrence property. The line segments show the most frequently occurring direction for each given distance and orientation difference from the central reference element. (D) Bayesian likelihood-ratio function. Each line segment shows a possible geometrical relationship between an element and the reference element. The color of a line segment indicates the likelihood that the element and reference belong to the same physical contour divided by the likelihood they belong to different physical contours. (E) A thresholded local grouping function derived from the edge co-occurrence plot in C. (F) A thresholded local grouping function derived from the Bayesian likelihood-ratio function in D.

As can be seen in Fig. 3D, edge elements that are co-circular (i.e. consistent with a smooth continuous contour) are more likely to belong to the same physical contour. These results support our interpretation of the absolute statistics in Fig. 3C, and provide further evidence that the Gestalt principle of good continuation has a physical basis in the statistics of the natural world. Most importantly, these results allow us to determine a maximum likelihood (optimal) local grouping function for contour grouping in natural scenes. Given the fundamental importance of contour grouping for useful vision, it is possible that the human local grouping function is near this optimum.
A successful method to measure the statics of second order was shown by [Geisler et al., 2001, Vision Research] on a set of natural images. In this study, they defined second-order statistics to compare an edge as a function of a central reference edge as a pdf on 3 parameters: the distance $d$ between their centers, the angle $\psi$ between the central edge sand the center of the second and $\theta$ the difference between the orientation of both edges. Probability is represented by this colormap and to represent on the 2D of the screen this 3d function, they represent in (B) the most probable difference of orientation at each distance and angle, showing the tendency of having collinear, parallel structures in natural images and (C) the most probable angle for each difference of angle and distance, showing a prior bias in natural image for cocircular edges.

We will use the same statistical method in this research, but... However the underlying statistics was done using an heuristic on the edge extraction (as in Sigman 01) and I show here the edges extracted by Geisler on a sample image as they were used in this study. That's why we decided to keep the same treatment for the statistics but opted for a different, more principled way to extract edges.
Log Gabor representation

[Fischer et al., 2007, International Journal of Computer Vision]
in order to do that, we first used a linear transformation using a log-gabor representation.

(definition) This representation is a good and generic model of edges as defined by their shape, orientation and scale. It matches what is well described for the response of simple cells’ response in area V1. We show here on the top left that these filters tile evenly the Fourier space, but also that these correspond to a good model of edges at different orientation, scale and phase compared to other dictionaries like the Daubechies wavelet base Db4 in (e) and the steerable pyramid by Eero Simoncelli.

(Fischer) Obviously, this dictionary is over-complete, but their correlation is easy to compute and allow for a relative translation-rotation-scale invariance. We proved that this was better adapted to the extraction of edges than gabors (Fischer, 07).
Sparse coding

(MP) from this linear representation, we searched for the most sparse representation using a $\ell_0$ norm approach for which Matching Pursuit proved to be a good approximation. I refer to this paper that appeared in IEEE TIP for more details (BICV) and that I reviewed in a recent book chapter (Perrinet, 2015 bicv), including a model of complex cells’ response (Fischer, 2007). It is generic and efficient.
Some examples of edge extraction

Natural

Laboratory
Some examples of edge extraction

- Method: detection of edges
- Some examples of edge extraction

We show here the results of the edge extraction on a set of patches extracted from both database. Parameters for each edge are the scalar amplitude, position, phase, orientation and scale. The hue gives the orientation, the length represents the size (scale) of the edge. This shows that edges are qualitatively well extracted.

This method is rather general and proves to be efficiently grabbing most edges. In particular we can reconstruct the image from them and we achieved a performance measured by the RMSE of $\approx 5 - 10\%$

Both images classes appear qualitatively different and we tried to characterize how they differ. First-order statistics (distribution of positions, orientations, scale) showed a typical pattern but no differences between these 2 classes... However, when computing...
Outline: Edge co-occurrences and categorizing images

Introduction: linking neural structure to natural scenes
- Geisler et al, 2001
- Bosking et al, 1997
- Problem statement

Method: detection of edges
- Geisler et al, 2001
- Log Gabor representation
- Sparse coding
- Some examples of edge extraction

Results: natural vs. laboratory images
- Second-order statistics
- Quantitative difference using classification

Intermediate Summary

Categorizing animals in images from natural scenes
let's now see the results of using such a method by

- showing some examples of edge extraction
- computing second-order statistics
- and trying to draw a quantitative measure of the difference between natural and laboratory images
- database collection
Second-order statistics

\[ p(d, \psi, \theta, \sigma | \pi_0) \approx p(d, \sigma | \pi_0)p(\theta, \psi | \pi_0) \]
... second-order stats, we will see a clear difference. The histogram was computed as a 4-dimensional function of distance, azimuth $\psi$, difference of orientation $\theta$ and ratio of scale.

(colin) let’s first replicate the result from Geisler by showing that relative to a given edge (segment in the center), what is the Here I show for each distance and angle the most probable difference of angle, showing that collinear and parallel edges predominate.

(lab) when using the images from the laboratory environment, one finds a different pattern where the colinearity clearly dominates: this quantitatively shows the difference between the edges’s second-order statistics...

(cocir) a similar pattern is observed the cocircular plot. it reproduces the results from Geisler on natural images, but laboratory environment shows a strong bias to colinearity.
Second-order statistics
Probability distribution function of "chevrons" in natural and laboratory images (angles) By computing measures of the independence of the different variables, we found that the probability density function of the second-order statistics of edges factorizes with on one side distance and scale and on the other side the 2 angles. The first component proved to be quite similar across both classes and the greater difference is seen for different angle configuration. As it can be reduced to 2 dimensions, we can plot the full probability as shown here by different contrast values assigned to all possible chevrons configurations, for all possible "azimuth" values $\psi$ on the horizontal axis and difference of orientation $\theta$ on the vertical axis. Such a plot most strikingly shows the difference between these 2 classes.

one issue now that we can show the 2nd order statistics is to know if it would be possible to quantify such difference...
Quantitative difference using classification

- **Database 1**
  - Control

- **Database 2**
  - Man-made

<table>
<thead>
<tr>
<th>Database 1</th>
<th>Database 2</th>
<th>FO</th>
<th>CM</th>
<th>SO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>Man-made</td>
<td>86%</td>
<td>92%</td>
<td>99%</td>
</tr>
</tbody>
</table>
(SVM) ... we did that by using a simple classifier (a SVM classifier using the KL distance between the histogram of one image to the average histogram to each class), so that one gets a simple, translation, orientation and scale invariant classifier which can efficiently differentiate between one natural image and a lab image. It comes as a big surprise as this is only based on some local characteristic, but it sufficient to get good classification. This gives also a quantitative method to rate different methods and databases. The result as computed by f1-score is of 99.3% accuracy.

(animals) as this ROC curve would be kind of boring to show (a right angle "perfect" wedge), we show one more surprising result: applying the same type of procedure to images of animals versus non-animals as used by the team of Simon Thorpe in Toulouse, one can get a pretty good classification of approximately 70% accuracy. This comes as a surprise as we are only using a one-layer, feed-forward computation... still there is much to explore.
Intermediate Summary

\[ p(d, \psi, \theta, \sigma | \pi_0) \approx p(d, \sigma | \pi_0) p(\theta, \psi | \pi_0) \]
To summarize, at this point of this talk I hope I convinced you that

- second-order statistics are efficiently computed by using a the algorithm from Geisler et al. (2001), with a more general edge extraction algorithm that uses sparse coding

- Collinearity and co-circularity results for natural images replicated qualitatively the results from Geisler et al. (2001), confirming that prior information about continuations appeared consistently in natural images. However, we find that the largely man-made environment in which these animals were reared has a significantly higher probability of collinear edge elements.

- We thus predict that if the lateral connection patterns are due to visual experience, the patterns in wild-raised tree shrews would be very different from those measured by Bosking et al. (1997), with shorter-range correlations and less emphasis on collinear continuations. This prediction can be tested in future experiments on matching groups of animals reared in different environments. $p(d, \sigma)$ should be hard-wired while $p(\theta, \psi)$ should be adaptive. cocircularity is weak but existent + anti-cocircularity is surprising= should be present as a connection structure.

- … but, as we are here, may we try this on animals vs natural?…
Categorizing animals in images from natural scenes
• ⬜ so if it works for calssifying natural vs laboratory images, shouldn’t it work for
  the task that Simon Thorpe as extensively studied, that is categorizing natural
  images ⬜ with or without an animal? This seems a bit crazy especially, knowing
  that most model (like the one from Serre 07) would rather consider that it is
  better done in 10 processing layers. But... let’s just do it!

• ⬜ Let’s show the Ratio of probability distribution function of chevron maps in
  artificial and animal images compared to the ’natural’ set. As previously, we
  show the probability of chevron maps for two databases (artificial, animal), but
  here we compare this density to natural images by plotting the log-ratio of
  densities. These plots highlight the differences with natural images and shows
  the deviations in co-linearity for artificial images and in co-circularity for animal
  images (blue: less likely, red: more likely; the opacity is proportional to the
  absolute log-ratio). In this probabilistic framework, both high and low values
  carry information on the category of the image.
Categorizing animals in images from natural scenes

![Bar chart showing F1 scores for different categories of images.](chart.png)
Classification results for different sets of natural images (50% is chance level) and different classification methods. Note that for SVM, we controlled for different feature choice: first-order statistics (SVM 1), full second-order statistics (SVM 2) or the chevron map only (SVM C). Surprisingly, the results are very similar to the results of Serre 07, and we do not lose much performance by choosing the simple chevron map. First, this has direct application in machine learning as this translation- and rotation-independent classifier performs as well as more complicated networks (that is they have more parameters) - and could be imported to architectures such as spikenet.

But what can we conclude from that? It shows that compared to the view of the visual system as a single-track, feed-forward architecture (as shown in the figure showing the architecture of Serre, 07), visual information can follow different, parallel, certainly redundant routes that will allow to optimize the efficiency of the system. These systems may work in synergy and challenges the dichotomy between analytical vs global representations: it rather seems that with the brain, much is possible. This is certainly useful to learn more about the adaptivity and generality of neural computations.
Categorizing animals in images from natural scenes

False alarms vs non-animal

orientation difference θ

azimuth difference ψ

occurrences

closer to non-animals

difference of distance

closer to animals

550 best

50 worst
To see whether the patterns of errors made by humans are consistent with our model, we studied the second-order statistics of the 50 non-animal images that human subjects in Serre et al. (2007) most commonly falsely reported as having an animal. We call this set of images the false-alarm image dataset. (Left) This chevron map plot shows the ratio between the second-order statistics of the false-alarm images and the full non-animal natural image dataset, computed as in Figure 3 (left). Just as for the images that actually do contain animals (Figure ??, left), the images falsely reported as having animals have more co-circular and converging (red chevrons) and fewer collinear and orthogonal configurations (blue chevrons). (Right) To quantify this similarity, we computed the Kullback-Leibler distance between the histogram of each of these images from the false-alarm image dataset, and the average histogram of each class. The difference between these two distances gives a quantitative measure of how close each image is to the average histograms for each class. Consistent with the idea that humans are using edge co-occurences to do rapid image categorization, the 50 non-animal images that were worst classified are biased toward the animal histogram ($d’ = 1.04$), while the 550 best classified non-animal images are closer to the non-animal histogram.
Categorizing animals in images from natural scenes

Target
Distractor
Retina
Edge & scale map
Sparse edge image
Edge co-occurences
Decision layer

Non-animal images

orientation difference $\theta$

azimuth difference $\psi$

probability ratio

$= Ny Q$

$= R kj Q$

$= y Q$

$= y$
To conclude... These results call into question previous claims that a hierarchical analysis of the visual scene is necessary for classification into high-level categories (Serre, 07). We speculate that the observed differences in second-order statistics for images with animals have an underlying basis in the physical constraints governing the shapes of animals. Specifically, animals typically have compact shapes, constrained by their capacity to move, unlike plants rooted in one location that must stretch rather than move towards resources. Thus we expect that the second-order statistics will be useful as a rough but general and fast method for distinguishing a wide range of scene and object categories. Similar observations apply to other sensory systems, where we would expect co-occurrence between primary sensory elements (such as stimulation of a patch of skin, presence of a specific auditory frequency, or activation of a taste or smell receptor) to differ between ecologically important classes of stimuli.

Thank you for your attention.
W. Geisler, J. Perry, B. Super, D. Gallogly
Edge co-occurrence in natural images predicts contour grouping performance.

D. Field and A. Hayes and R. Hess
Contour integration by the human visual system: evidence for a local "association field".

Y. Choe and R. Miikkulainen
Contour integration and segmentation with self-organized lateral connections.

W. Bosking, Y. Zhang, B. Schofield, and D. Fitzpatrick
Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex

E. Callaway and L. Katz
Emergence and refinement of clustered horizontal connections in cat striate cortex.

Self-Invertible 2D Log-Gabor Wavelets.

L. U. Perrinet.
Role of homeostasis in learning sparse representations.

L. U. Perrinet, and J. Bednar
Edge co-occurrences can account for rapid categorization of natural versus animal images.
*Scientific Reports*, 2015.
URL http://invibe.net/LaurentPerrinet/Publications/PerrinetBednar15

P. Seriès, S. Georges, J. Lorenceau, and Y. Frégnac.
Orientation dependent modulation of apparent speed: a model based on the dynamics of feed-forward and horizontal connectivity in V1 cortex.
Neuromorphic implementation

Fig. 1. Cartoon of the VI model, which represents an array of cortical units

[Series et al., 2002]
Matching Pursuit
Matching Pursuit

N=2
Matching Pursuit
Matching Pursuit

N=8
Matching Pursuit
Matching Pursuit

N=32
Matching Pursuit

N=64
Matching Pursuit
Matching Pursuit
Matching Pursuit

N=512
Matching Pursuit

N=1024
Matching Pursuit

N=2048