Learning Association Fields from Natural Images

Francesco Orabona, Giorgio Metta and Giulio Sandini
LIRA-Lab, DIST
University of Genoa, Genoa, ITALY 16145
{bremen,pasa,giulio}@liralab.it

Abstract

Previous studies have shown that it is possible to learn certain properties of the responses of the neurons of the visual cortex, as for example the receptive fields of complex and simple cells, through the analysis of the statistics of natural images and by employing principles of efficient signal encoding from information theory. Here we want to go further and consider how the output signals of 'complex cells' are correlated and which information is likely to be grouped together. We want to learn 'association fields', which are a mechanism to integrate the output of filters with different preferred orientation, in particular to link together and enhance contours. We used static natural images as training set and the tensor notation to express the learned fields. Finally we tested these association fields in a computer model to measure their performance.

1. Introduction

The goal of perceptual grouping in computer vision is to organize visual primitives into higher-level primitives thus explicitly representing the structure contained in the data. The idea of perceptual grouping for computer vision has its roots in the well-known work of the Gestalt psychologists back at the beginning of the last century who described, among other things, the ability of the human visual system to organize parts of the retinal stimulus into "Gestalten", that is, into organized structures. They formulated a number of so-called Gestalt laws (proximity, common fate, good continuation, closure, etc.) that are believed to govern our perception. It is logical to ask if these laws are present in the statistics of the world.

On the other hand it has been long hypothesized that the early visual system is adapted to the input statistics [1]. Such an adaptation is thought to be the result of the joint work of evolution and learning during development. Neurons, acting as coincidence detectors, can discover and use regularities in the incoming flow of sensory information, which eventually represent the Gestalt laws. It has been proposed that, for example, the mechanism that link together the elements of a contour is rooted in our biology, with neurons with lateral and feedback connections implementing these laws.

There is a large body of literature about computational modeling of various parts of the visual cortex, starting from the assumption that certain principles guide the neural code ([21] for a review). In this view it is important to understand why the neural code is as it is. Bell and Sejnowski [2], for example, demonstrated that it is possible to learn receptive fields similar to those of simple cells starting from natural images. In particular they demonstrated that it is possible to reproduce these receptive fields hypothesizing the sparsity and independence of the neural code. In spite of this, there is very little literature on learning an entire hierarchy of features, that is not only the first layer, and possibly starting from these initial receptive fields.

A step in the construction of this hierarchy is the use of 'association fields' [5]. In the literature, these fields are often hand-coded and employed in many different models with the aim to reproduce the human performance in contour integration. These fields are supposed to resemble the pattern of excitatory and inhibitory lateral connection between different orientation detector neurons as found, for instance, by Schmidt *et al*. [19]. In fact, Schmidt has shown that cells with an orientation preference in area 17 of the cat are preferentially linked to iso-oriented cells. Furthermore, the coupling strength decrease with the difference in the preferred orientation of pre- and post-synaptic cell. Models typically consider variations of the co-circular approach [8, 9, 13], that is two oriented elements are part of the same curve if they are tangent to the same circle. Others [22] have considered exponential curves instead of circles obtaining similar results.

Our question is whether it is possible to learn these association fields from the statistics of natural images. Different authors have used different approaches: using a database of tagged images [4, 6], using motion as an implicit tagger [18] or hypothesizing certain coding properties of the cortical layer [10].

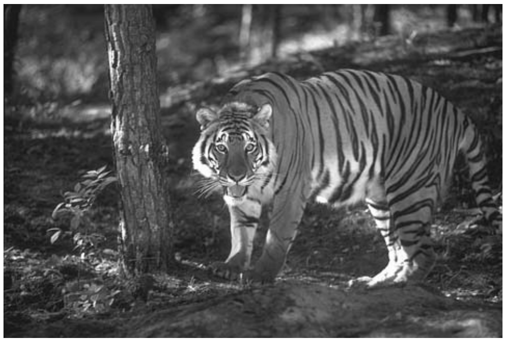

Figure 1. Example image of the dataset.

Our approach is similar to to one of Sigman *et al.* [20], which uses images as the sole input. Further, we aim to obtain precise association fields, useful to link contours in a computer model.

The rest of the paper is organized as follows: section 2 contains a description of the method. Section 3 describes a first set of experimental results and a method to overcome problems due to the non-uniform distribution of the image statistics. In section 4 we show the fields computed with this last modification and finally in section 5 and 6 we show the performance of the fields in edge detection on a database of natural images and we draw some conclusions.

2. Learning from images

We assume the existence of a first layer that simulates the behavior of the complex cells; in this paper we do not address the issue on how to learn them since we are interested in the next level of the hierarchy. Using the output of this layer we want to estimate the mean activity around points with a given orientation. For example it is likely that if a certain image position contains a horizontal orientation, then the adjacent pixels on the same line would be points with an orientation almost horizontal.

To have a precise representation of the orientations and at the same time something mathematically convenient we have chosen to use the tensor notation. Second order symmetric tensors can capture the information about the first order differential geometry of an image. Each tensor describes both the orientation of an edge and its confidence for each point. The tensor can be visualized as an ellipse, whose major axis represents the estimated tangential direction and the difference between the major and minor axis the confidence of this estimate. Hence a point on a line will be associated with a thin ellipse while a corner with a circle. Consequently given the orientation of a reference pixel, we estimate the mean tensor associated with the surrounding pixels. The use of the tensor notation give us the possibility to exactly estimate the preferred orientation in each point of the field and also to quantify its strength and confidence.

We have chosen to learn a separate association field for each possible orientation. This is done for two main reasons:

- It is possible to find differences between the association fields. For example, it is possible to verify that the association field for the orientation of 0 degrees is stronger than that of 45 degrees.

- For applications of computer vision, considering the discrete nature of digital images, it is better to separate the masks for each orientation, instead of combining the data in a single mask that has to be rotated leading to sampling problems. The rotation can be done safely only if there is a mathematical formula that represents the field, while on the other hand we are inferring the field numerically.

We have chosen to learn 8 association fields, one for each discretized orientation. The extension of the fields is chosen of 41x41 pixels taken around each point. It should be noted that even if we quantized the orientation of the (central) reference pixel to classify the fields, the information about the remaining pixels in the neighbor were not quantized, differently to [6, 20]. There is neither a threshold nor a pre-specified number of bins for discretization and thus we obtain a precise representation of the association field.

Images used for the experiments were taken from the publicly available database (Berkeley Segmentation Database [15]) which consists of 300 color images of 321x481 and 481x321 pixels; 200 of them were converted to black and white and used to learn the fields, collecting 41x41 patches; an example image from the dataset is shown in figure 1.

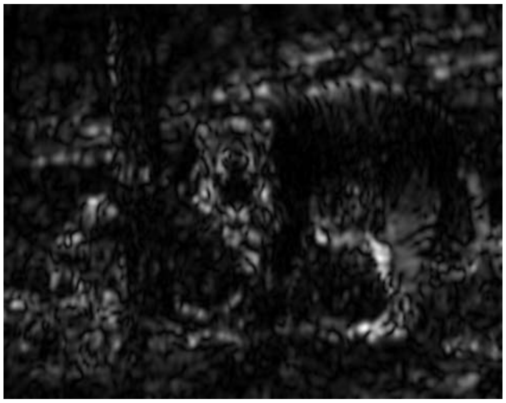

Figure 2. Complex cells output to the image in figure 1 for 0 degrees filter of formula (1).

2.1. Feature extraction stage

There are several models of the complex cells of V1, but we have chosen to use the classic energy model [16] on the intensity channel. The response is calculated as:

$$E_\theta = \sqrt{(I * f_\theta^e)^2 + (I * f_\theta^o)^2} \tag{1}$$

where f_θ^e and f_θ^o are a quadrature pair of even and odd-symmetric filters at orientation θ. Our even-symmetric filter is a Gaussian second-derivative, and the corresponding odd-symmetric is its Hilbert transform. In figure 2 there is an example of the output of the complex cells model for the 0 degrees orientation.

Then the edges are thinned using a standard non-maximum suppression algorithm. This is equivalent to finding edges with a Laplacian of Gaussian and zero crossing. The outputs of these filters are used to construct our local tensor representation.

2.2. Tensors

In practice a second order tensor is denoted by a 2x2 matrix of values:

$$\mathcal{T} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \tag{2}$$

It is constructed by direct summation of three quadrature filter pair output magnitudes as in[11]:

$$T = \sum_{k=1}^{3} E_{\theta_k} \left(\frac{4}{3} \hat{n}_k^T \hat{n}_k - \frac{1}{3} I \right) \tag{3}$$

where E_{θ_k} is the filter output as calculated in (1), I is the 2x2 identity matrix and the filter directions $\hat{n}_k$ are:

$$\begin{aligned} \hat{n}_1 &= (1, 0) \\ \hat{n}_2 &= \left(\tfrac{1}{2}, \tfrac{\sqrt{3}}{2}\right) \\ \hat{n}_3 &= \left(-\tfrac{1}{2}, \tfrac{\sqrt{3}}{2}\right) \end{aligned} \tag{4}$$

The greatest eigenvalue λ_1 and its corresponding eigenvector e_1 of a tensor associated to a pixel represent respectively the strength and the direction of the main orientation. The second eigenvalue λ_1 and its eigenvector e_1 have the same meaning for the orthogonal orientation. The difference $\lambda_1 - \lambda_2$ is proportional to the likelihood that a pixel contains a distinct orientation.

3. Preliminary results

We have run our test only for a single scale, choosing the σ of the Gaussian filters equal to 2, since preliminary tests have shown that a similar version of the fields is obtained with other scales as well. Two of the obtained fields are in figures 3 and 4. It is clear that they are somewhat corrupted by the presence of horizontal and vertical orientations in any of the considered neighbors and by the fact that in each image patch there are edges that are not passing across the central pixel. On the other hand we want to learn association field for curves that do pass through the central pixel. Geisler *et al.* [6] used a human labeled database of images to infer the likelihood of finding edges with a certain orientation relative to the reference point. On the other hand, Sigman *et al.* [20] using only relative orientation and not absolute ones, could not have seen this problem. In our case we want to use unlabeled data to demonstrate that it is possible to learn from raw images and, as mentioned earlier, we do not want to consider only the relative orientations, but rather a different field for each orientation. We believe that this is the same problem that Prodöhl *et al.* [18] experienced using static images: the learned fields supported collinearity in the horizontal and vertical orientations but hardly in the oblique ones. They solved this problem using motion to implicitly tag only the important edges inside each patch.

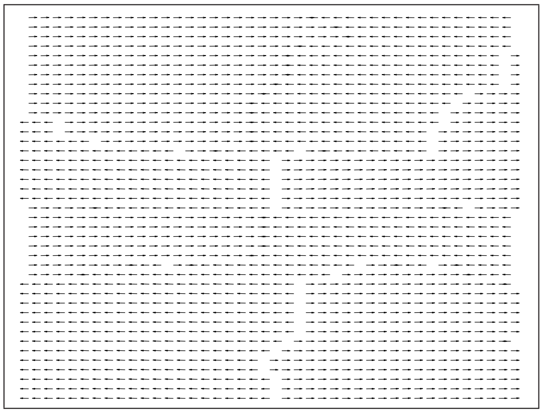

Figure 3. Main directions for the association field for the orientation of 0 degrees in the central pixel.

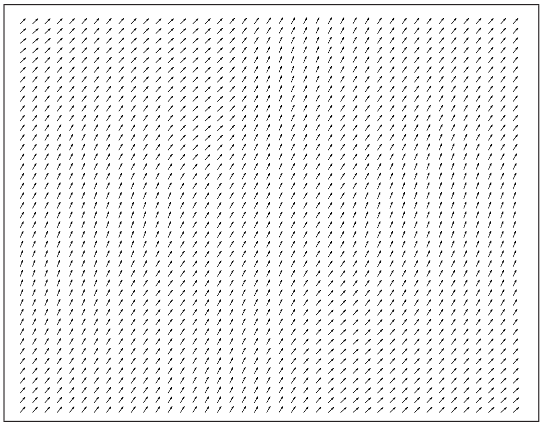

Figure 4. Main directions for the association field for the orientation of 67.5 degrees in the central pixel.

3.1. The path across a pixel

The neural way to solve the problem shown earlier is thought to be the synchrony of the firing between nearby neurons: if stimuli co-occur, then the neurons synchronize [7]. Inspired by this we considered in each patch only pixels that belong to a curve that goes through the central pixel. In this way the gathered data will contain only information about curves connected to the central pixel. Note that we select curves inside each patch, not inside the entire image. The simple algorithm used to select the pixels in each patch is the following:

1. put central pixel of the patch in a list;
2. tag first pixel in the list and remove it from the list. Put surrounding pixels that are active (non-zero) in the list;
3. if the list is empty quit otherwise go to 2.

With this procedure we remove the influence of horizontal and vertical edges that are more present in the images and that are not removed by the process of averaging. On the other hand, we are losing some information, for example about parallel lines, that in any case should not be useful for the enhancement of contours. Note that that this method is completely parameter free; we are not selecting the curves following some specific criterion, instead we are just pruning the training set from some kind of noise. It is important to note that this method will learn the bias present in natural images versus horizontal and vertical edges [3], but it will not be biased to learn *only* these statistics, as in Prodöhl *et al*. [18] when using static images.

4. Results

We tested the modified procedure on the database of natural images and also on random images (results not shown), to verify that the results were not an artifact due to the method.

In figures 5, 6 there are respectively the main orientations, their strengths (eigenvalues) and the strengths in the orthogonal directions of the mean estimated tensors for the orientation of 0 degrees of the central pixel. Same for figures 7 and 8 for 67.5 degrees. The structure of the obtained association field closely resembles the fields proposed by others based on collinearity and co-circularity. We note that the size of the long-range connection far exceeds the size of the classical receptive field. We note also that the noisier regions in the orientation corresponds to very small eigenvalues so they do not influence very much the final result.

While all the fields have the same trend, there is a clear difference in the decay of the strength of the fields. To see this we have considered only the values along the direction of the orientation in the center, normalizing the maximum values to one. Figure 9 shows this decay. It is clear that fields for horizontal and vertical edges have a wider support, confirming the results of Sigman *et al*. [20].

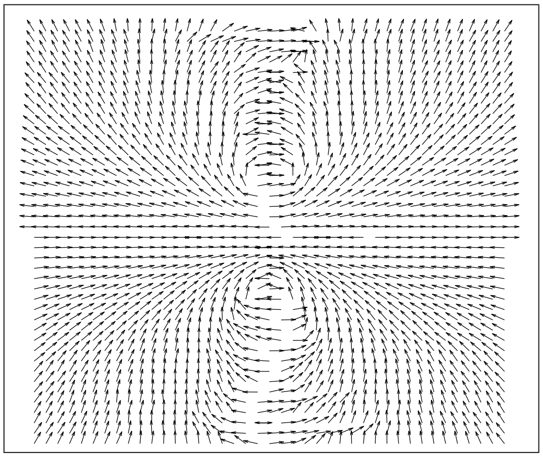

Figure 5. Main directions for the association field for the orientation of 0 degrees in the central pixel, with the modified approach.

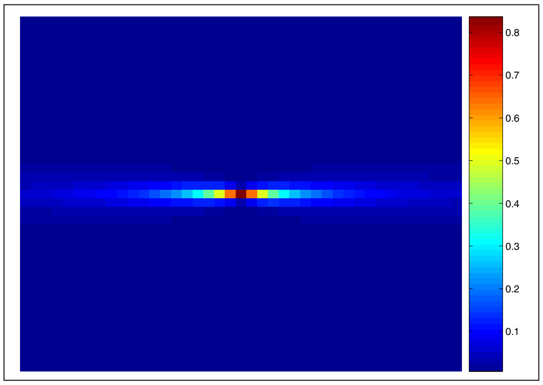

Figure 6. Difference between the two eigenvalues of the association field of figure 5.

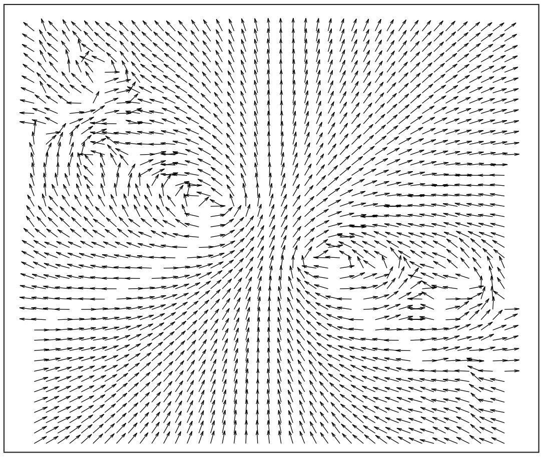

Figure 7. Main directions for the association field for orientation of 67.5 degrees, with the modified approach.

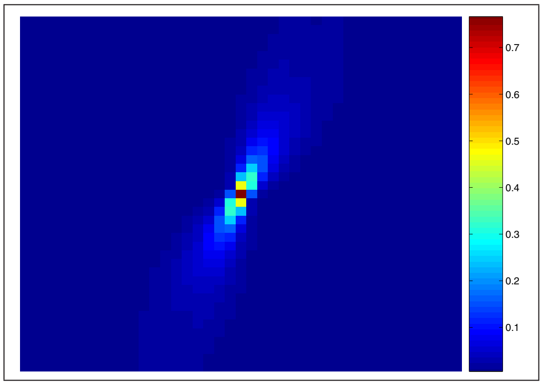

Figure 8. Difference between the two eigenvalues of the association field of figure 7.

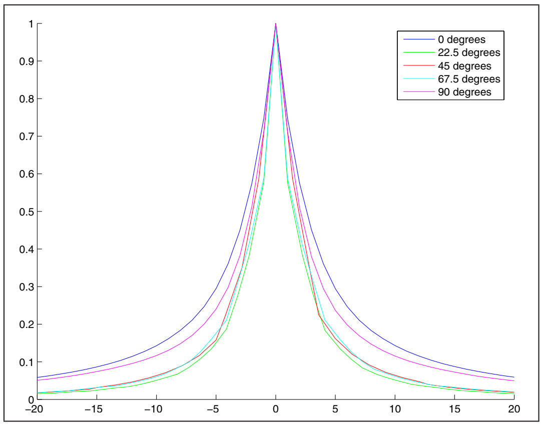

Figure 9. Comparison of the decay for the various orientations. On the y axis there are the first eigenvalues normalized to a maximum of 1, on the x axis is the distance from the reference point along the main field direction.

5. Using the fields

The obtained fields can be used with any existing model of contour enhancement, but to test them we have used the tensor voting scheme proposed by Guy and Medioni *et al.* [9]. The choice is somewhat logical considering to the fact that the obtained fields are already tensors. In the tensor voting framework points communicate with each other in order to refine and derive the most preferred orientation information. Differently to the original tensor voting algorithm we don't have to choose the right scale of the fields [12] since it is implicitly in the learnt fields. We compared the performances of the tensor voting algorithm using the learned fields versus the simple output of the complex cell layer, using the Berkeley Segmentation Database and the methodology proposed by Martin *et al.* [14, 15]. We can see the results in figure 10: there is a clear improvement using the tensor voting and the learned association fields instead of just using the simulated outputs of the complex cells alone.

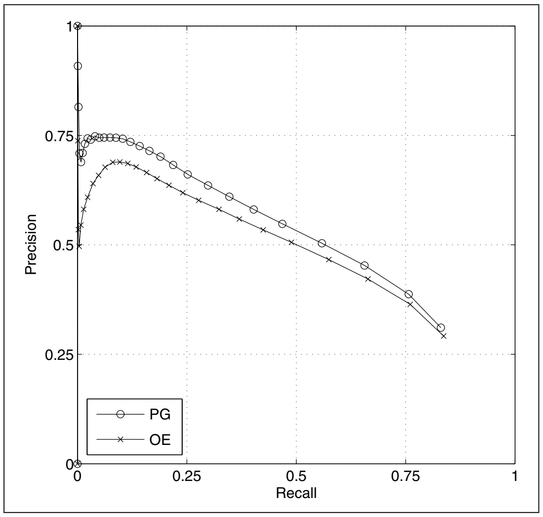

Figure 10. Comparison between tensor voting with learned fields (PG label) and the complex cell layer alone (OE label).

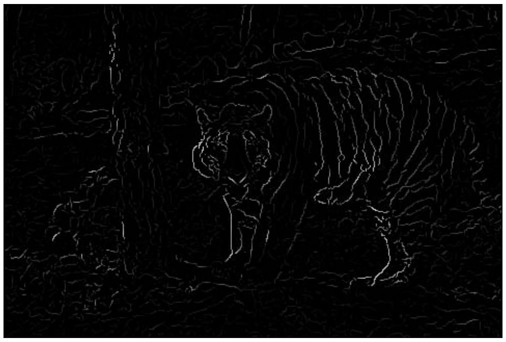

Figure 11. Test image contours using the complex cell layer alone.

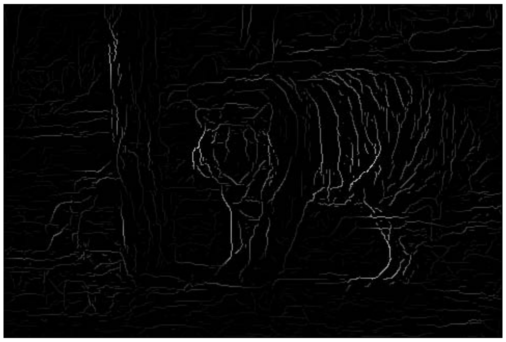

Figure 12. Test image contours using tensor voting with the learned fields.

An example of the results on the test image in 1, after the non-maximum suppression procedure, are shown in figures 11 and 12.

6. Conclusion

Several authors have studied the mutual dependencies of simulated complex cells responses to natural images. The main result from these studies is that these responses are not independent and they are highly correlated when they are arranged collinearly or on a common circle. In the present paper we have presented a method to learn precise association field from natural images. A bio-inspired procedure to get rid of the non-uniform distribution of orientations is used, without the need of a tagged database of images [4, 6], the use of motion [18] or supposing the cortical signals sparse and independent [10]. The learned fields were used in a computer model, using the tensor voting method, and the results were compared using a database of human tagged images which helps in providing clear numerical results.

However the problem of learning useful complex features from natural images could in any case find a limit beyond these contour enhancement networks. In fact the *usefulness* of a feature is not directly related to image statistics but supposes the existence of an embodied agent *acting* in the natural environment, not just perceiving it. In this sense in the future we would like to link strategies like the one used by Natale *et al*. [17] and the approach in the current paper, to link the first stages of unsupervised learning, to reduce the dimensionality of the inputs, to other stages of supervised learning for the definition of the extraction of useful features for a given task.

Acknowledgment

This work was supported by EU project RobotCub (IST-2004-004370) and CONTACT (NEST-5010). The authors would like to thank Harm Aarts for the discussion on the earlier version of the manuscript.

References

[1] H. B. Barlow. Possible principles underlying the trasformations of sensory messages. In W. A. Rosenblith, editor, *Sensory Communication*, pages 217–234. MIT Press, 1961. 1

[2] A. J. Bell and T. J. Seinowski. The ’indipendent components’ of natural scenes are edge filters. *Vision Research*, 37:3327–3338, 1997. 1

[3] D. M. Coppola, H. R. Purves, A. N. McCoy, and D. Purves. The distribution of oriented contours in the real world. *PNAS*, 95:4002–4006, 1998. 4

[4] J. H. Elder and R. M. Goldberg. Ecological statistics of gestalt laws for the perceptual organization of contours. *Journal of Vision*, 2:324–353, 2002. 1, 6

[5] D. J. Field, A. Hayes, and R. F. Hess. Contour integration by the human visual system: evidence for local ”association field”. *Vision Research*, 33(2):173–193, 1993. 1

[6] W. S. Geisler, J. S. Perry, B. J. Super, and D. P. Gallogly. Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, 41:711–724, 2001. 1, 2, 3, 6

[7] C. M. Gray, P. König, A. K. Engel, and W. Singer. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, 338:334–336, 1989. 4

[8] S. Grossberg and E. Mingolla. Neural dynamics of perceptual grouping: textures, boundaries, and emergent segmentations. *Perceptual Psychophysics*, 38:141–171, 1985. 1

[9] G. Guy and G. Medioni. Inferring global perceptual contours from local features. *Int. J. of Computer Vision*, 20:113–133, 1996. 1, 5

[10] P. O. Hoyer and A. Hyvärinen. A multilayer sparse coding network learns contour coding from natural images. *Vision Research*, 42(12):1593–1605, 2002. 1, 6

[11] H. Knutsson. Representing local structure using tensors. In *Proc. 6th Scandinavian Conference on Image Analysis*, pages 244–251, Oulu, Finland, 1989. 3

[12] M.-S. Lee and G. Medioni. Grouping ., -, →, θ, into regions, curves and junctions. *Journal of Computer Vision and Image Understanding*, 76(1):54–69, 1999. 5

[13] Z. Li. A neural model of contour integration in the primary visual cortex. *Neural Computation*, 10:903–940, 1998. 1

[14] D. Martin, C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color and texture cues. *IEEE Transactions on Pattern Analysis and Machine intelligence*, 26(5):530–549, 2004. 5

[15] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. 8th Int’l Conf. Computer Vision*, volume 2, pages 416–423, July 2001. 2, 5

[16] M. Morrone and D. Burr. Feature detection in human vision: A phase dependent energy model. *Proc. Royal Soc. of London B*, 235:221–245, 1988. 3

[17] L. Natale, F. Orabona, G. Metta, and G. Sandini. Exploring the world through grasping: a developmental approach. In *Proc. 6th CIRA Symposium*, pages 27–30, June 2005. 6

[18] C. Prodöhl, R. P. Würtz, and C. von der Malsburg. Learning the gestalt rule of collinearity from object motion. *Neural Computation*, 15:1865–1896, 2003. 1, 3, 4, 6

[19] K. Schmidt, R. Goebel, S. Löwel, and W. Singer. The perceptual grouping criterion of collinearity is reflected by anisotropies of connections in the primary visual cortex. *European Journal of Neuroscience*, 5(9):1083–1084, 1997. 1

[20] M. Sigman, G. A. Cecchi, C. D. Gilbert, and M. O. Magnasco. On a common circle: Natural scenes and gestalt rules. *PNAS*, 98(4):1935–1940, 2001. 2, 3, 4

[21] E. Simoncelli and B. Olshausen. Natural images statistics and neural representation. *Annual Review of Neuroscience*, 24:1193–1216, 2001. 1

[22] V. Vonikakis, A. Gasteratos, and I. Andreandis. Enhancement of perceptually salient contours using a parallel artificial cortical network. *Biological Cybernetics*, 94:194–214, 2006. 1