Karls-Eberhard-Universität Tübingen
Mathematisch-Naturwissenschaftliche Fakultät
Institut für Neurobiologie
Lehrstuhl für Kognitive Neurowissenschaften

Bachelorarbeit Kognitionswissenschaft

Sparse Coding of Natural Binocular Images

Freschta Mohammadzada

28. September 2015

Gutachter
Prof. Dr. Hanspeter A. Mallot
Lehrstuhl für Kognitive Neurowissenschaften
Universität Tübingen

Zweitgutachterin
Prof. Dr. Barbara Kaup
Lehrstuhl für Kognitionspsychologie
Universität Tübingen

Betreuer
Gerrit Ecke
Lehrstuhl für Kognitive Neurowissenschaften
Universität Tübingen

Content

1. Introduction

We live in a three-dimensional, complex world and throughout the day we are constantly exposed to information about our versatile environment. How is it possible for us to live in and adapt to a consistently changing environment?

How is the brain able to process large amounts of information in a systematic, logical and at the same time non-linear way?

The visual cortex plays a vast role in our conscious perception of the visual world. Its underlying mechanisms help us to get a vivid image of the surrounding world.

Because our two eyes are located at different lateral positions in the head, they send slightly different two-dimensional images to our brain. These positional differences are called binocular disparities. Our brain measures these disparities to reconstruct depth perception, thus a three-dimensional structure of the visual environment.

The mechanism in our brain that measures and uses binocular disparity is called *stereopsis* (Poggio & Poggio, 1984).

Disparity-selective neurons, both binocular simple cells and complex cells, have been found in the primary visual cortex (V1) of many mammals (Barlow, Blakemore & Pettigrew, 1967; Hubel & Wiesel, 1962).

Simple cells are characterized as cells in V1, that respond to orientated edges and gratings. Their receptive fields are defined as confined regions of the visual field that contain clearly defined excitatory and inhibitory regions (Wolfe et al., 2009). They are active when light falls either in the center of the field (gratings) or in the outer region (edge); the position of the light stimulus within the field is very important for the activity of the neuron.

Contrary to simple cells, complex cells in V1 and V2 (secondary visual cortex), that also respond to orientated edges and gratings, are not position-selective, meaning that they are also active when a stimulus is positioned elsewhere, as long as it falls into their respective fields with a preferred orientation. Thus their receptive fields don't have clearly defined excitatory and inhibitory regions.

Hubel and Wiesel (1962) proposed that complex cell fields are of higher-order and may be built up by simpler fields.

As Information processing begins at early stages of the visual cortex, it gains greater specificity at later stages, meaning that at higher levels, fewer cells are activated, but each is more selective and

specific to stimuli (Barlow, 1972).

It is thus reasonable to assume that the visual cortex has employed an efficient strategy to make use of the massive flow of information that is contained in natural images; such images show consistent statistical properties that crucially differentiate them from random ones (Barlow, 1989).

In the following it will be explained why receptive fields of simple cells in V1 are specifically well suited to describe these properties and especially how they probably do it:

The spatial receptive fields of simple cells in mammalian primary visual cortex are characterized as being localized, oriented and bandpass (Hubel & Wiesel, 1962; Olshausen & Field, 1996) and Field (1987) has shown that these properties are well described as a result of producing *sparse representations*.

An input is represented sparsely when only a small number of cells, that convey meaningful information (of the input), is active. The probability of producing a response is the same and equally distributed across all cells, however it is low for any given cell, thus sparse (Field, 1994).

In order to achieve a better understanding on why the visual system might work that way, Olshausen and Field have developed an unsupervised learning algorithm that produces receptive fields that resemble those of simple cells.

The underlying principle is called *sparse coding*: "information is represented by a relatively small number of simultaneously active neurons out of a large population" (Figure 1; Olshausen & Field, 2004, S. 481).

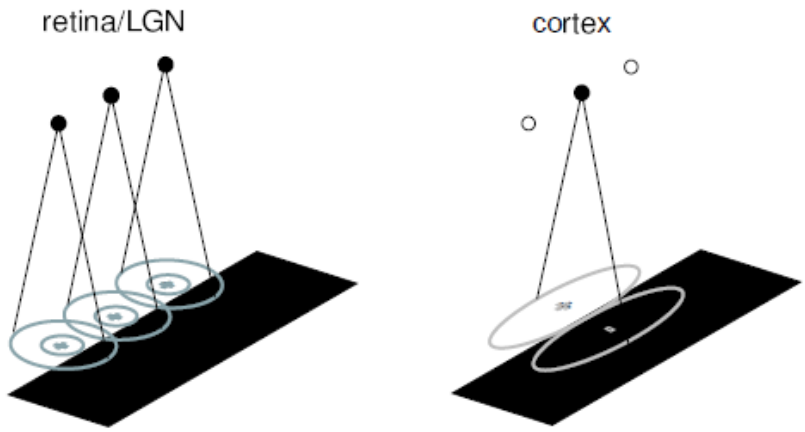

Figure 1: *Example of sparse coding. Cortical neurons use fewer active units to represent information in images than neurons in the retina or LGN (Filled circles denote active units; unfilled circles denote inactive neurons.) (From Olshausen, 2003).*

Thus, Olshausen and Field's learning algorithm is focusing on remodeling the structure of natural images, while efficiently using the redundancy in these images to minimize the dependence among the outputs. Sparse coding groups redundant information together and allows only a few events to meaningfully describe the image, i.e. the sparse code itself is highly redundant (Olshausen, 2003).

The algorithm that finds a sparse code can be formulated as an optimization problem that minimizes the following cost function:

$$E = -[\textit{reconstruction error}] - \lambda[\textit{sparseness of } a_i]$$

λ is a positive factor, it determines the importance of the second term relative to the first.

The first term describes how to generate data from given images: let us assume that an image $I(x, y)$ can be described in terms of a linear superposition of its spatial features:

$$I(x) = \sum_i a_i \cdot \varphi_i(x) + \nu(x)$$

φ_i is a set of basis functions describing spatial features (in this model they are the equivalent to neurons, that are activated for a given image). The coefficients a_i tell us how much of each feature is contained in the image (the coefficients can be seen as an equivalent to the firing rate of neurons). The variable ν denotes additional Gaussian noise. To make sure that the output provides a complete representation of the visual scene, the number of outputs, i.e. the set of basis functions φ_i should be larger than the dimensionality of the image, which is also often referred to as overcompleteness (Olshausen, 2003). First, this allows the image to be described in various ways; second, it allows the basis functions (neurons) to only become active when the input pattern is close to their preferred pattern (Olshausen & Field, 2004). Because, as mentioned above, what we want is a meaningful representation of the input rather than actually reducing redundancy.

After we have a set of all given features (and coefficients), the reconstruction error of a given image is set by $[\textit{reconstruction error}] = -\sum_{x,y} [I(x,y) - \sum_i a_i \cdot \varphi_i(x,y)]^2$

where $I(x, y)$ is the actual image and $\sum_i a_i \cdot \varphi_i(x, y)$ is the reconstructed image.

The second term determines the sparseness of the code for a given image. If activity is distributed over many coefficients, the representation incurs a higher cost, compared to those representations in which activity is carried by only a few coefficients. The cost function is defined by the sum of each

coefficient's activity passed through a non-linear function $S(x)$:

$$[sparseness\,of\ a_i] = -\sum_i S\left(\frac{a_i}{\sigma}\right) \quad .$$

σ is a scaling constant. For the non-linear function $S(x)$, functions such as $|x|$, $-e^{-x^2}$ or $\log(1+x^2)$ are good choices, in that they are most active among those coefficients, that are not zero; the further the coefficients are away from zero, the higher is the cost function. Learning is achieved when the total cost function $E = -[reconstruction\,error] - \lambda[sparseness\,of\ a_i]$ is minimized, with regard to a_i . φ_i then derives via gradient descent on E averaged over many image presentations. Optimization is done by two alternating steps. First, in order to determine the coefficients a for a given image, a is formulated by the equilibrium solution to the differential equation:

$$\dot{a}_i = b_i - \sum_j C_{ij} a_j - \frac{\lambda}{\sigma} S'(a_i) \quad \text{with} \quad b_i = \sum_{x,y} \varphi_i(x,y) I(x,y) \quad \text{and} \quad C_{ij} = \sum_{x,y} \varphi_i(x,y) \varphi_j(x,y) \quad .$$

Figure 2 gives an intuitive impression of how the equation is assembled and how each term contributes to finding those coefficients that maximally respond to a given image.

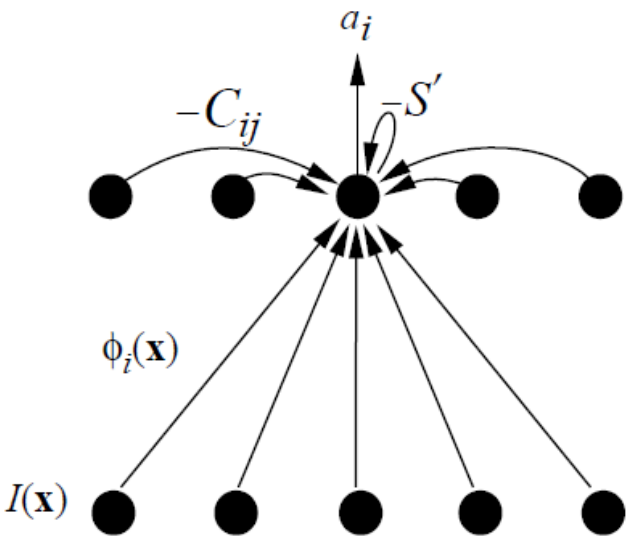

Figure 2: *Neuronal implementation of sparse optimization. represents the input image that is multiplied with the basis function of each unit.* $-C_{i,j}$ *is an inhibitory recurrent term. It is computed by the activities of other coefficients that are weighted by the overlap of their basis functions.* $-S'$ *is a self-inhibitory non-linear term, which contributes to sparse representation in that it restrains coefficients to become active. (From Olshausen, 2003).*

Second, with the computed coefficients we have a learning rule that updates φ :

$$\Delta\varphi_i(x_m, y_n) = \eta\langle a_i[I(x_m, y_n) - \hat{I}(x_m, y_n)]\rangle \quad .$$

$\hat{I}$ is the reconstructed image: $\hat{I}(x) = \sum_i a_i \varphi_i(x_m, y_n)$; η is the learning rate.

With Olshausen and Field's learning algorithm it is indeed possible to find a sparse code for monocular natural images that develops receptive fields that resemble those of simple cells in V1 (Olshausen & Field, 1996), in that the basis functions are spatially localized, oriented and bandpass.

In this study, we wanted to know what the output of the learning algorithm is when having binocular images as an input. We slightly modified parts of the algorithm in order to work with binocular images. A patch from the left image and the corresponding patch from the right image was extracted and concatenated horizontally.

To analyze the resulting receptive fields and how the disparity selectivity is generated, either by a position-shift model or by a phase-shift model, we fitted the resulting basis functions with Gabor functions. Simple-cells are generally well modeled by Gabor functions (Jones & Palmer, 1987).

A two-dimensional Gabor function is a product of a Gaussian envelope and a sine wave. The position-shift model can be represented in terms of a shift in the center location of the Gaussian (Figure 3 (a), the phase-shift model is modeled by the phases of the sine wave (Figure 3 (b)) (Tsao, Conway & Livingstone, 2003).

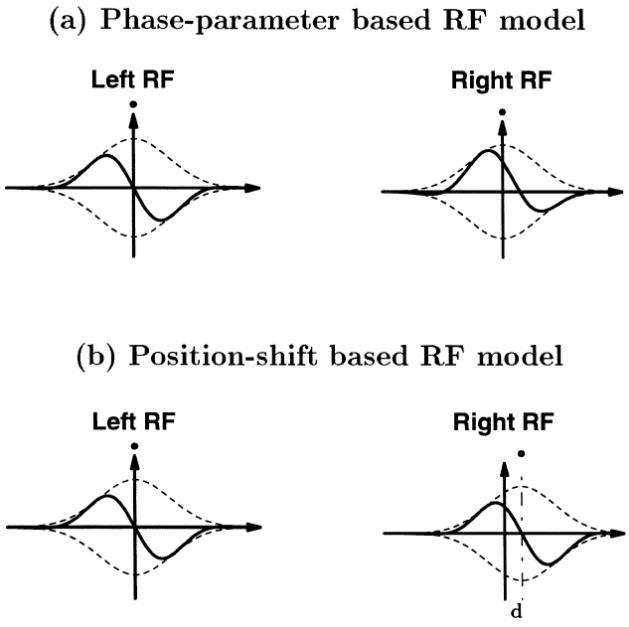

Figure 3: *Two binocular receptive fields models. (a) Phase-parameter based model (b) Position-parameter based model. The dotted lines indicate the peak position of the gaussian envelope, the solid lines indicate the left and right receptive fields profiles. (from Zhu & Qian, 1996).*

2. Methods

We employed MATLAB R2014b for all programs and figures. The computers we used had 64-Bit versions of Windows 7 or 8 installed. We faced problems with sparse coding and fitting the resulting basis functions with Gabor functions when using 32-bit versions, due to memory limitations.

The program for sparse coding was provided in the internet (https://github.com/viirya/Sparse-coding-with-GPU). This program is a combination of the improved and further developed versions by Olshausen and Field (1996) and Honglak Lee (2007).

In order to get a working version of the given software, in the file 'l2ls_learn_basis_dual.m', line 34, we had to change ***'options = optimset('GradObj','on', 'Hessian','on', 'TolFun', 1e-7)'*** to ***'options = optimset('GradObj','on', 'Hessian','on','Algorithm','trust-region-reflective')'***, due to compatibility problems in MATLAB versions (for further information see release notes MATLAB R2014a: New default algorithms in fmincon and quadprog http://de.mathworks.com/help/optim/release-notes.html).

For some individual Gabor fits we needed to tweak the number of iterations in the file 'autoGaborSurf.m' in line 48, when MATLAB would freeze or take up too much memory.

Some values for each training sections remained the same: dimensionality of the patches (16 x 16 pixels), number of patches (5000), number of iterations (4000) and the batch size (1000).

2.1 Test run: Sparse Coding of monocular images

In order to confirm that the original program with the altered settings works, we ran a test with original pictures from the download page, which are the same ones Olshausen and Field (1996) used in their experiment (ten natural monocular images, 512 x 512 pixels each). To have comparable results, the input was learned for 192 basis functions and we set the sparsity value equal to the one of Olshausen and Field, that is beta = 0.14.

2.2 Sparse Coding of stereo images with constant and natural disparities

The program for sparse coding needed to be slightly adjusted for our purposes. The part of the program (getdata_imagearray.m) that extracts the patches of the images was changed in that it extracts two corresponding patches of a binocular image. One patch of the left images and the

corresponding patch of the right image were merged together by simply concatenating both horizontally and delivered to the sparse coding implementation in order to learn the bases.

2.2.1 Natural binocular images with constant disparities

We downloaded seven natural monocular images (PNG-files) from arbitrary locations on the internet. They served as a basis to create binocular images with constant disparities in eight directions. Altogether we had a data set consisting of 56 binocular images based on 7 images (The data set can be found on the supplementary CD). The binocular images were 2x 512 x 512 pixels.
The input was learned once for 49 basis function with sparsity value 50 and once for 196 basis functions with sparsity value 15.

x	-2	-1	1	2	0	2	4	0
y	2	1	1	0	2	2	0	4

Table 1: The eight directions disparities in x- and y-direction applied to the seven pictures.

Figure 4: *Natural binocular image with constant disparities x = 0 and y = 4, created from monocular images, downloaded from the Internet.*

2.2.2 Natural stereo images with natural disparities

We downloaded 16 natural stereo images (binocular images) with natural disparities from the internet. Some of the stereo images were GIF-files originally, therefore, we needed to tailor them. In addition to that, we scaled the stereo images down to 256 x 128 pixels to avoid overlarge disparities (see *Discussion*) (The data set can be found on the supplementary CD).
The input was learned once for 49 basis functions with sparsitiy value 50 and once for 196 basis functions with sparsity value 15.

Figure 5: *Natural stereo image with natural disparities, downloaded from the Internet*

2.3 Analysis

In order to analyze the data, we fitted the resulting receptive fields with Gabor functions to examine position-shifts and phase-shifts.
The program that fitted the Gabor functions was downloaded from the internet (http://www.mathworks.com/matlabcentral/fileexchange/31485-auto-gaussian---gabor-fits).
There is a new link for opening the DRAM-Code in the file 'fetchMcmcPackages.m', line 11 has been changed to '***http://helios.fmi.fi/~lainema/dram/dramcode.zip***'.
We wrote a MATLAB script (gaborFunction.m), which takes a matrix with the resulting basis functions (or receptive fields) and fits them with the Gabor function ('autoGaborSurf.m').

Additionally, we calculated the cross correlation of the resulting basis functions to determine the location of the optimal disparity at which the left and right image is best correlated.

3. Results

3.1 Test run: Sparse Coding of monocular images

As Figure 6 demonstrates, the resulting 192 basis functions match with the results Olshausen and Field (1996) obtained. Different spatial frequency could be localized, as well as different orientations, which are well described by Gabors functions.

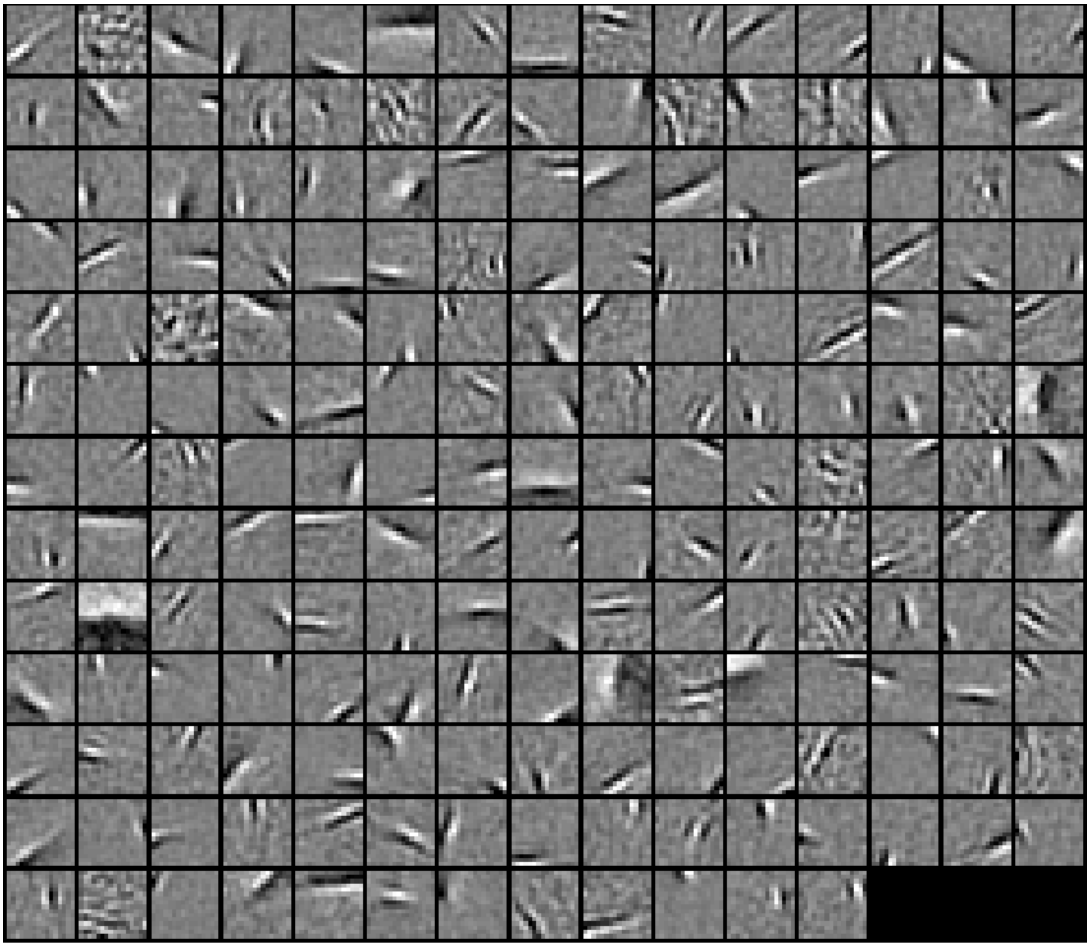

Figure 6: *The 192 learned basis functions with sparsity 0.14*

3.2 Sparse Coding of binocular images with constant disparities with 49 and 196 basis functions

3.2.1 Natural binocular images with constant disparities and 49 basis functions

Figure 7 shows the 49 learned basis functions.

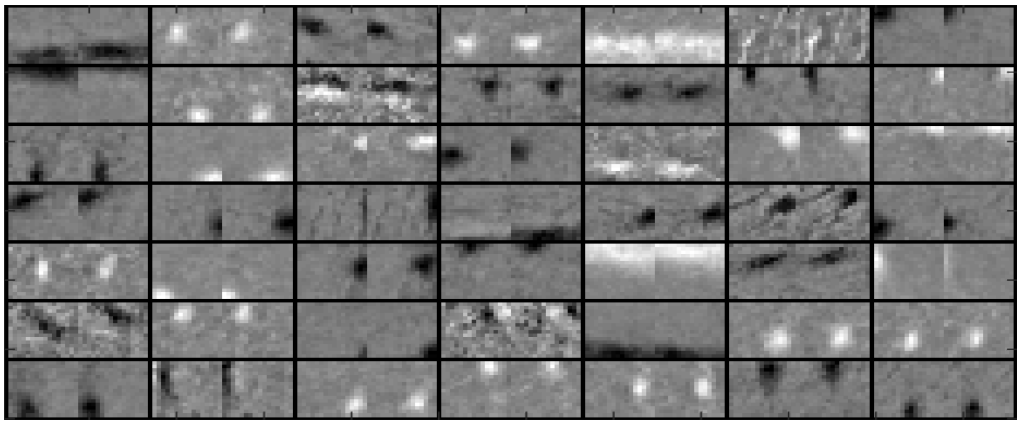

Figure 7: *The 49 learned basis functions for natural binocular images with constant disparities.*

As can be seen, the resulting basis functions are rather dotted than Gabor-like.

After fitting the Gabor functions to our data (Figure 8), we could see both, phase-shifts and position-shifts. Figure 9 shows the phase-shift measured in pixel. We assume that the sinus functions of the left and right basis functions have the same frequency. This is never accurately the case, so in order to calculate the phase-shift in pixels we applied the mean of both fitted frequencies.

The position-shift is depicted in Figure 10. The mapped cross-correlation of the basis functions is shown in Figure 11 and Figure 12 demonstrates in which location the corresponding patches are best correlated. The darker the dots, the more dots are superimposed, indicating that at those locations a lot of bases are sensitive for the same disparity.

Figures 13 and 14 show the correlations of the cross-correlations and position-shifts in x- and y-directions. The correlation for both is weak: r = 0.38.

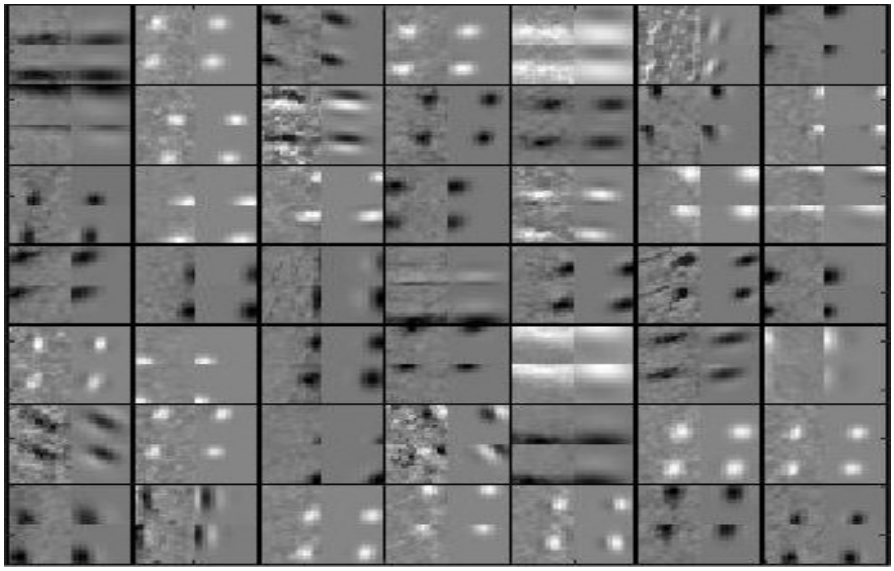

Figure 8: *The 49 basis functions fitted with Gabor functions*

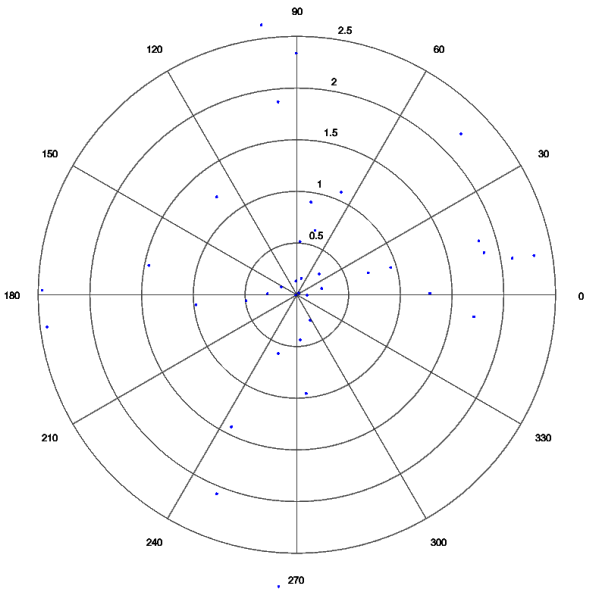

Figure 9: *Phase-shift of 49 Basis functions of natural binocular images with constant disparities in pixels.*

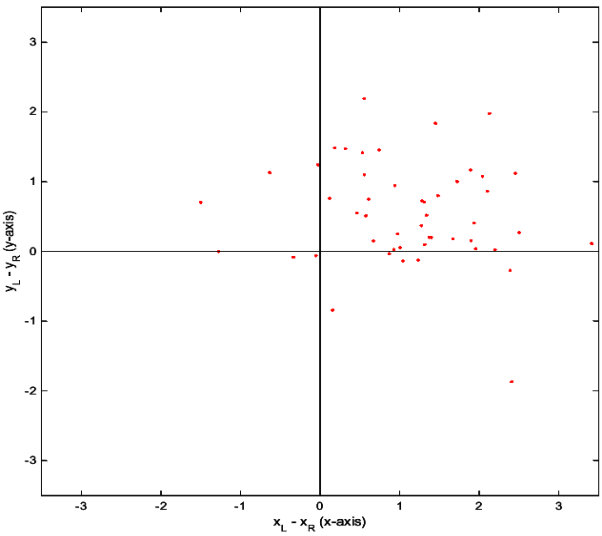

Figure 10: *Position-shift of 49 basis functions of natural binocular images with constant disparities*

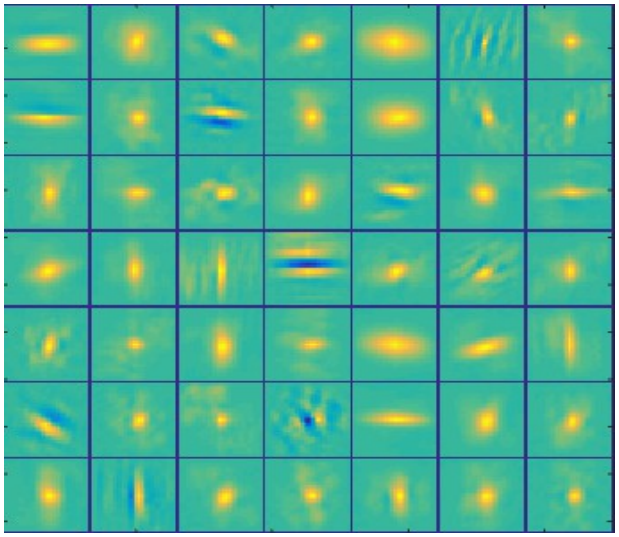

Figure 11: *Heatmap of the maximum Cross-Correlation of the 49 basis functions (constant disparities)*

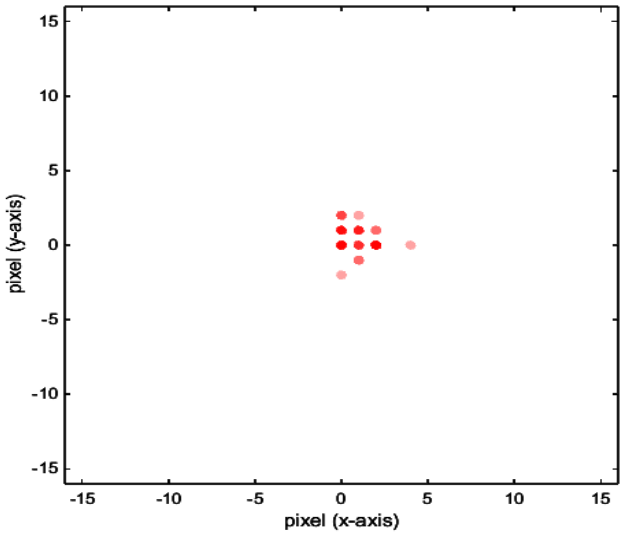

Figure 12: *Maximum Cross-Correlation of the 49 basis functions. Please note that, as mentioned above, the disparities were manipulated.*

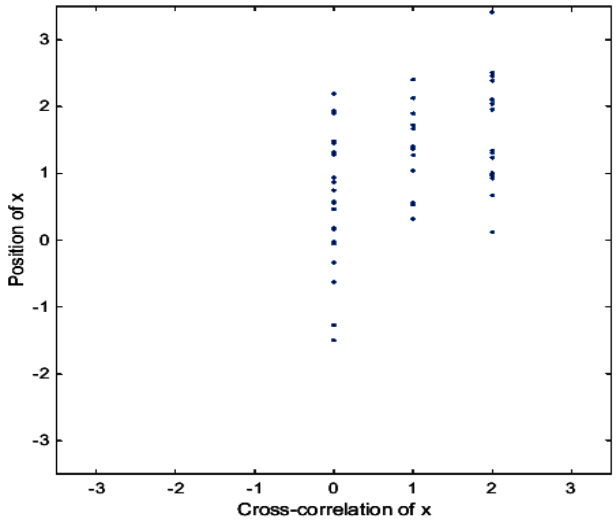

Figure 13: *Correlation of the position-shift in x-axis and cross-correlation of the x-axis (49 basis functions, constant disparities)*

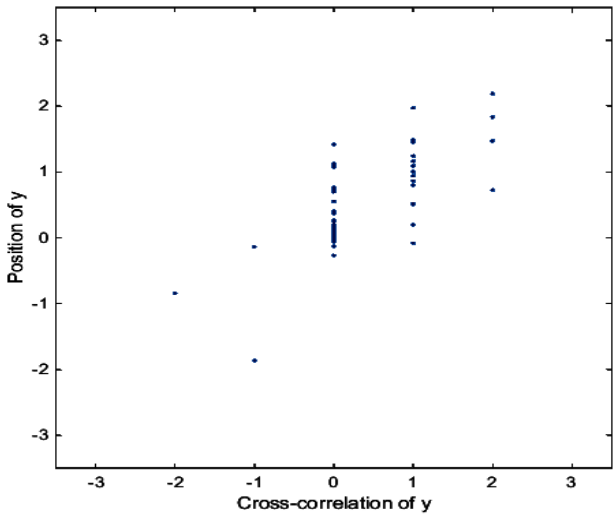

Figure 14: *Correlation of the position-shift in y-axis and cross-correlation of the y-axis (49 basis functions, constant disparities)*

3.2.2 Natural binocular images with constant disparities and 196 basis functions

Figure 15 demonstrates the 196 learned basis functions, the sparsity value was set to 15.
Dotted basis functions still remain here, however more Gabor-like basis functions can be found.

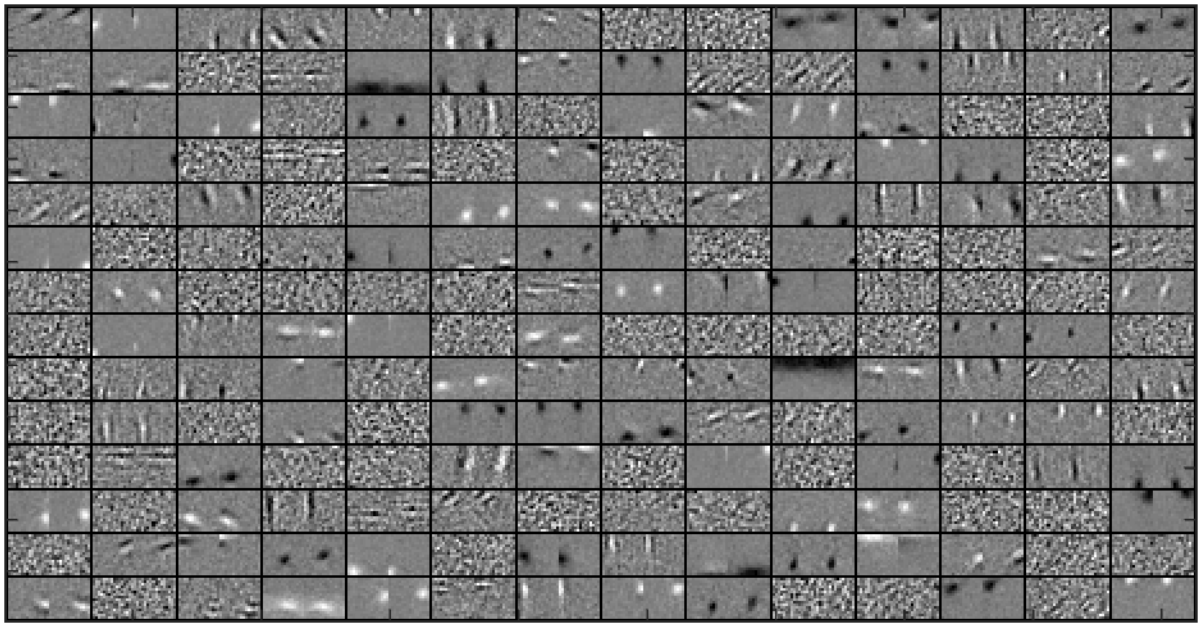

Figure 15: *The learned 196 basis functions for natural binocular images with constant disparities*

3.3 Sparse Coding of stereo images with natural disparities with 49 and 196 basis functions

3.3.1 Natural stereo images with natural disparities and 49 basis functions

Figure 16 shows the 49 learned basis functions. The sparsity value was set to 50. Some of the resulting basis functions show a Gabor-like form while most of them are rather elongated.

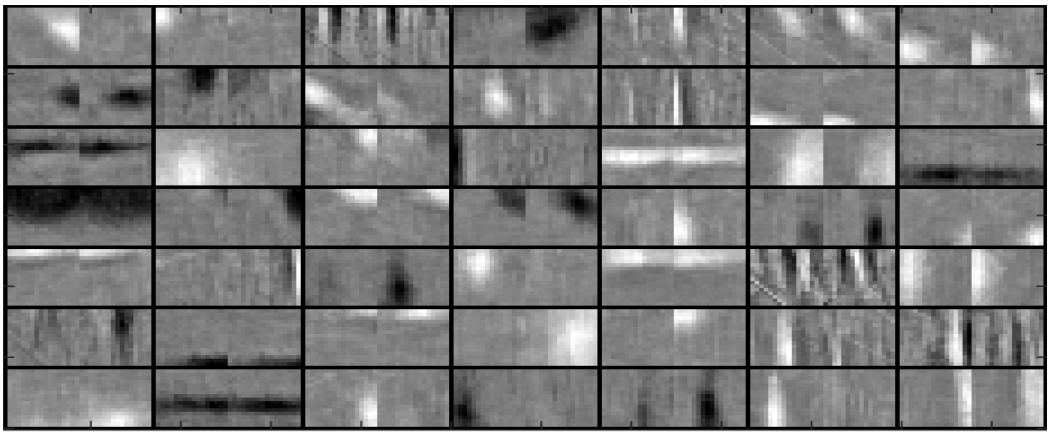

Figure 16: *The 49 learned basis functions for natural binocular images with natural disparities.*

After fitting the Gabor functions to our data (Figure 17), we could see both, phase-shifts and position-shifts. Figure 18 shows the phase-shift measured in pixel. We assume that the sinus functions of the left and right basis functions have the same frequency. This is never accurately the case, so in order to calculate the phase-shift in pixels we applied the mean of both fitted frequencies.

The position-shift is depicted in Figure 19. The mapped cross-correlation of the basis functions is shown in Figure 20 and Figure 21 demonstrates in which location the corresponding patches are best correlated. The darker the dots, the more dots are superimposed, indicating that at those locations a lot of bases are sensitive for the same disparity.

Figures 22 and 23 show the correlations of the cross-correlations and position-shifts in x- and y-directions. The correlation for both is relatively high: r = 0.7906.

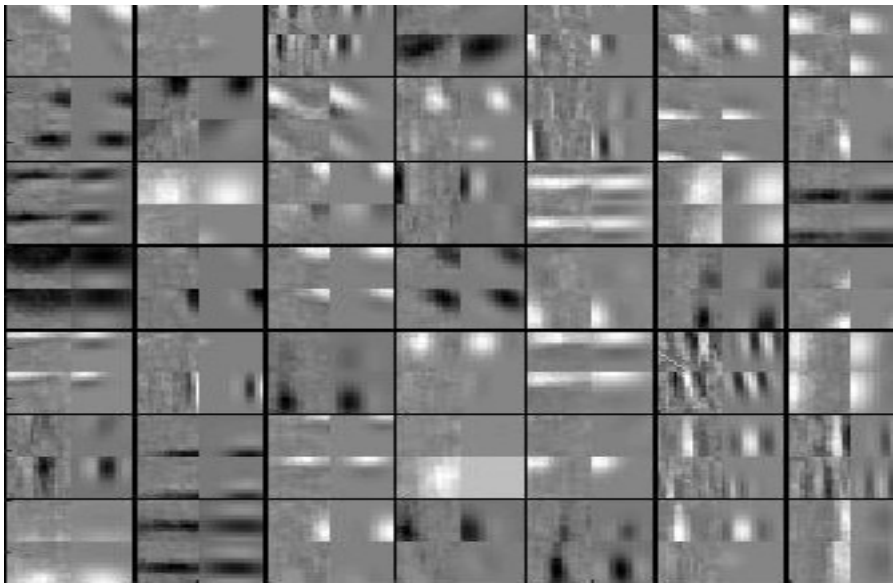

Figure 17: *The 49 basis functions fitted with Gabor functions*

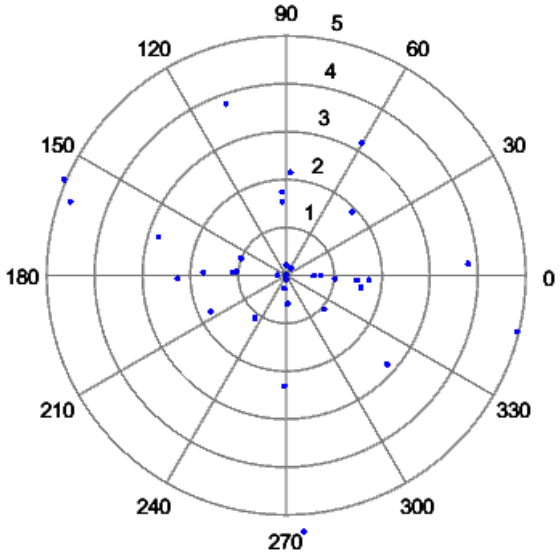

Figure 18: *Phase-shift of 49 Basis functions of stereo images with natural disparities in pixels.*

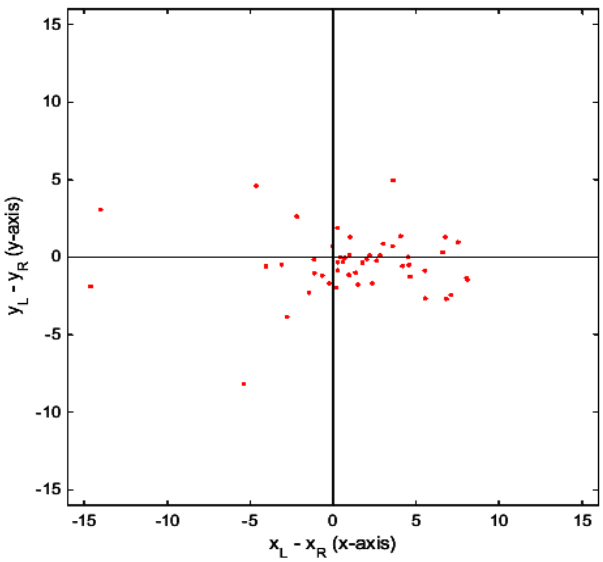

Figure 19: *Position-shift of 49 Basis functions of stereo images with natural disparities*

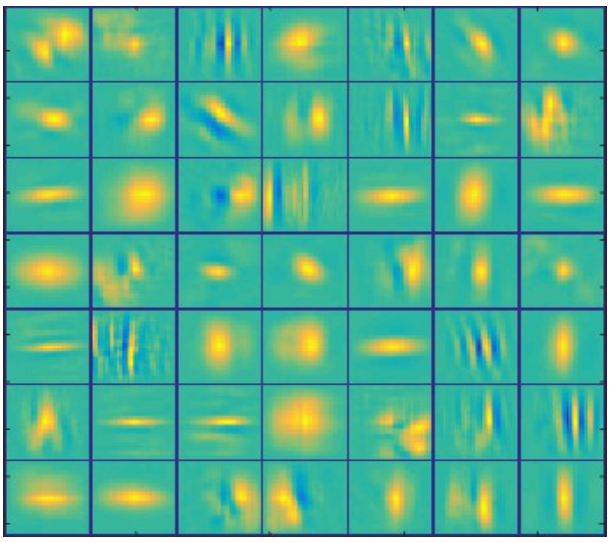

Figure 20: *Heatmap of the maximum Cross-Correlation of the 49 basis functions (natural disparities)*

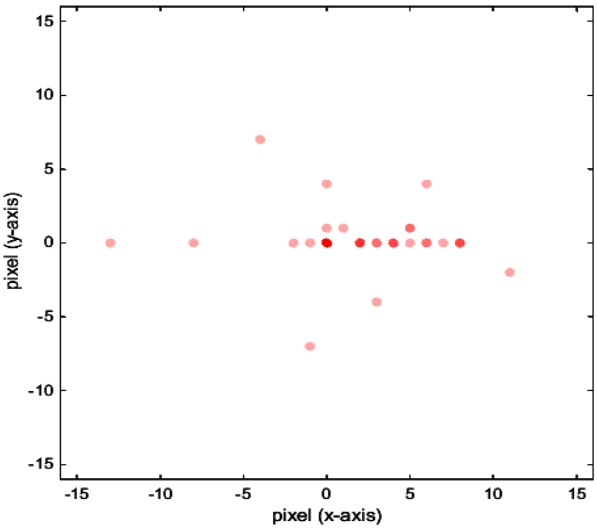

Figure 21: *Maximum Cross-Correlation of the 49 basis functions (natural disparities).*

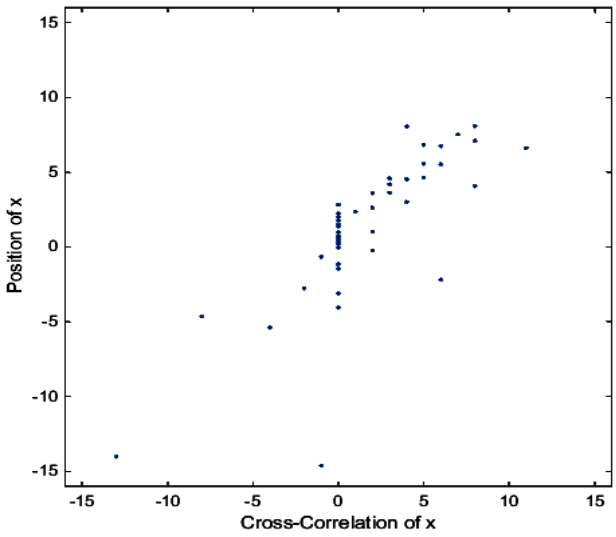

Figure 22: *Correlation of the position-shift in x-axis and cross-correlation of the x-axis (49 basis functions, natural disparities)*

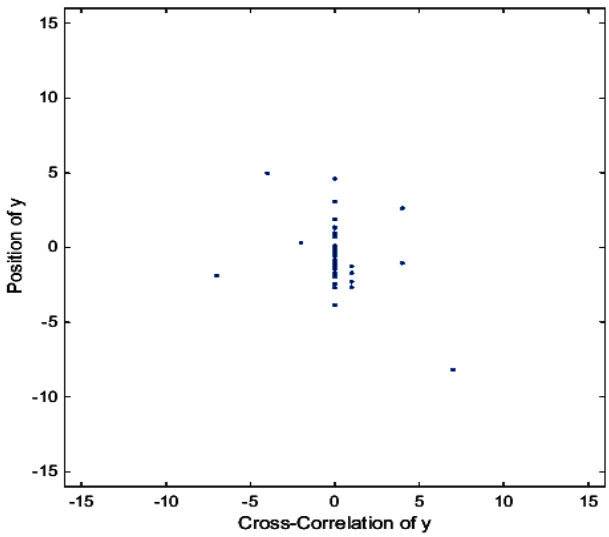

Figure 23: *Correlation of the position-shift in y-axis and cross-correlation of the y-axis (49 basis functions, natural disparities)*

3.3.2 Natural stereo images with natural disparities and 196 basis functions

Figure 24 shows the 196 learned basis functions, the sparsity value was set to 15.
Here we can see a lot of Gabor-like patches. However, in a lot of cases, while left bases are responding in an either Gabor-like or elongated fashion, it looks like the corresponding (right) bases do not respond at all, the bases remain gray or noisy.

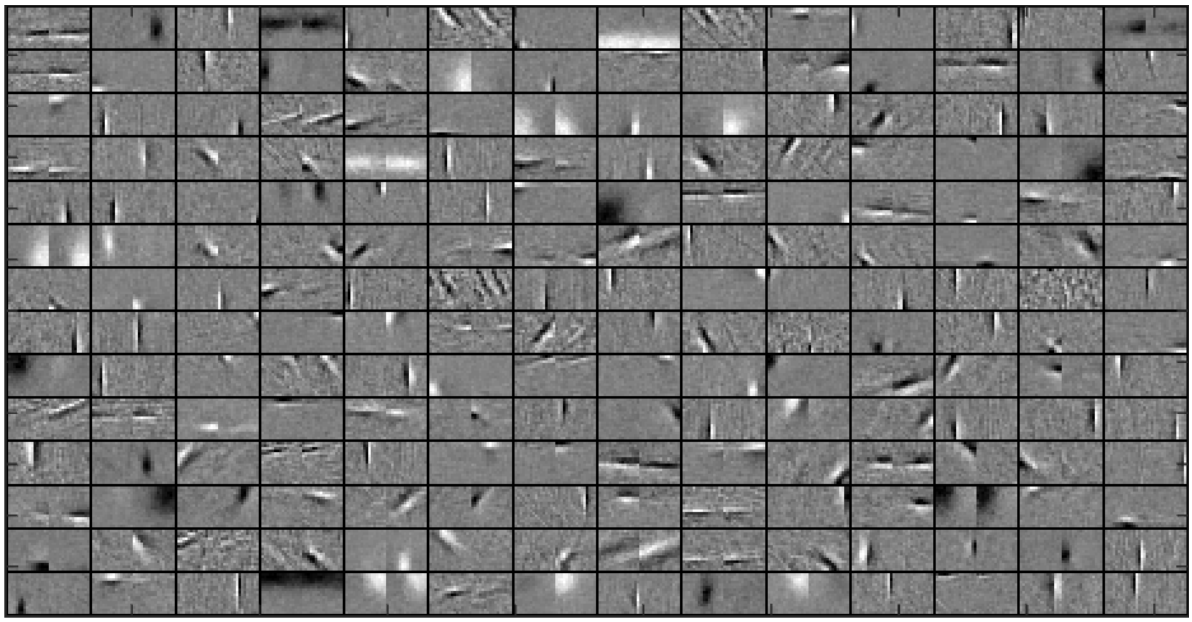

Figure 24: *The learned 196 basis functions for natural stereo images with natural disparities*

After fitting the Gabor functions to our data (Figure 25), we can see both, phase-shifts and position-shifts. Figure 26 shows the phase-shift measured in pixel. We assume that the sinus functions of the left and right basis functions have the same frequency. This is never accurately the case, so in order to calculate the phase-shift in pixels we applied the mean of both fitted frequencies.
Figure 27 shows the position-shift. The mapped cross-correlation of the basis functions is demonstrated in Figure 28 and Figure 29 displays in which location the corresponding patches are best correlated. As mentioned above, darker (superimposed) dots indicate that at those locations a lot of bases are sensitive for the same disparity.
Figure 30 and 31 show the correlations of the cross-correlations and position-shifts in x- and y-directions. The correlation for both is mid-strong: r = 0.5343.

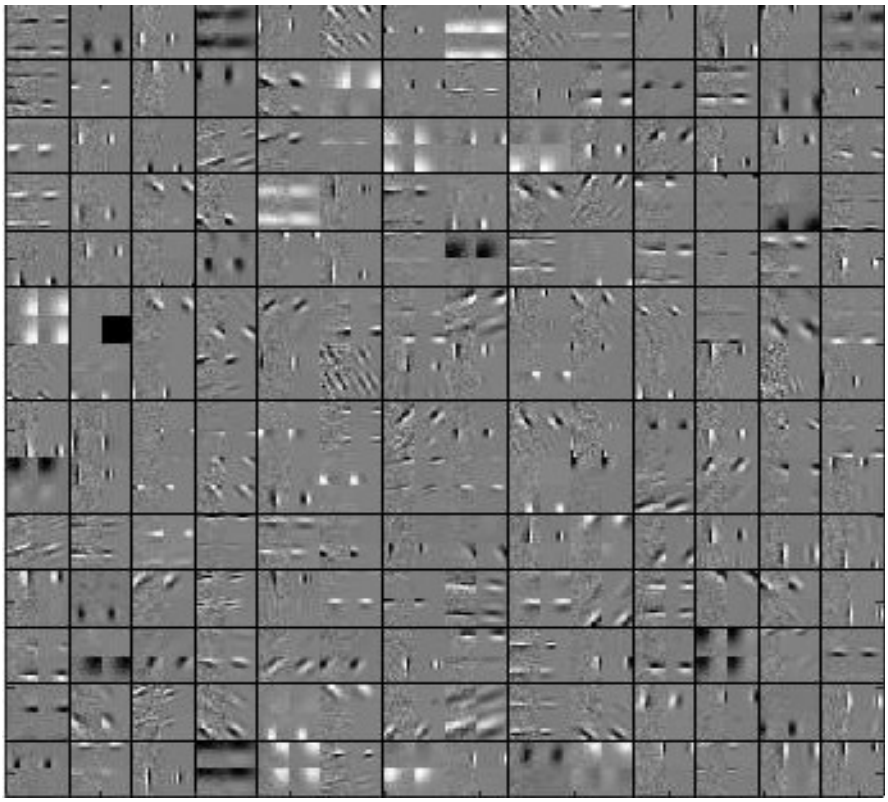

Figure 25: *The 196 basis functions fitted with Gabor functions*

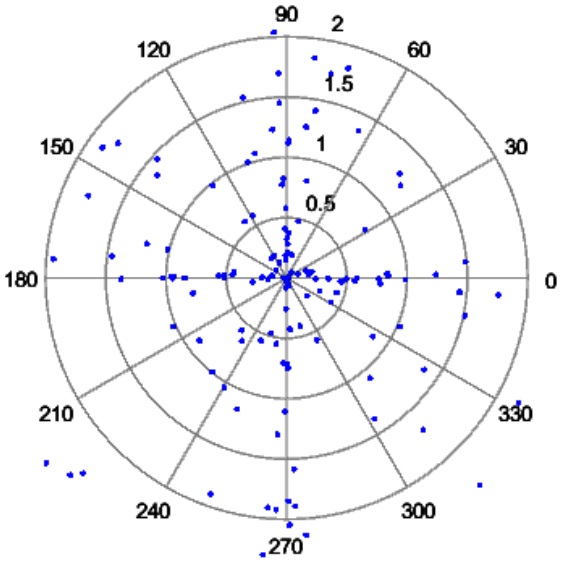

Figure 26: *Phase-shift of 196 basis functions of stereo images with natural disparities in pixels.*

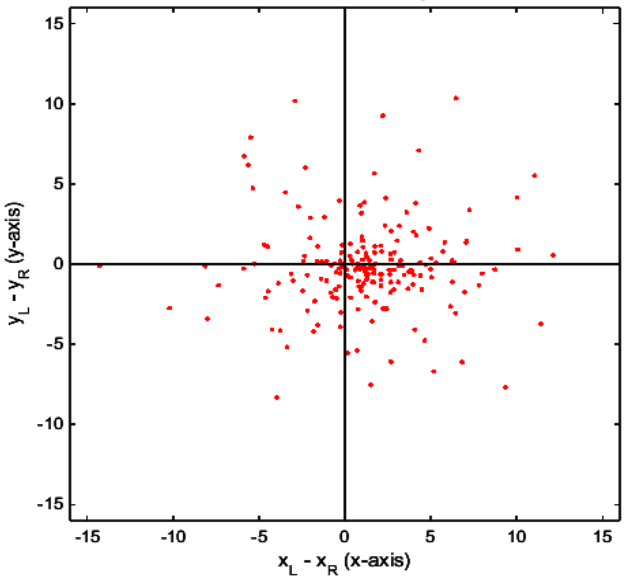

Figure 27: *Position-shift of 196 basis functions of stereo images with natural disparities*

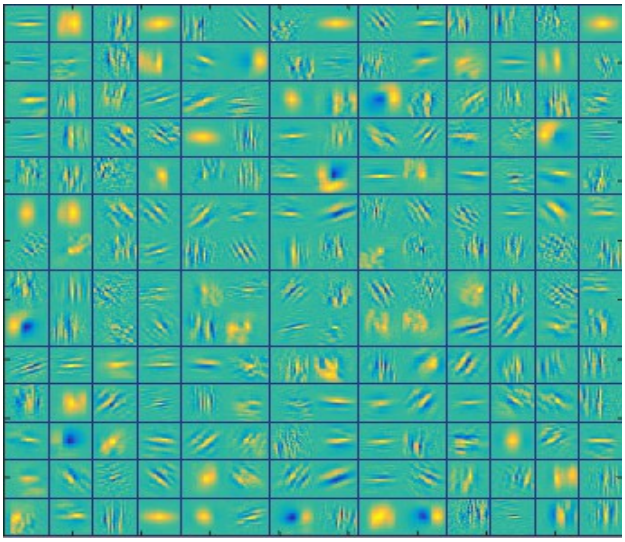

Figure 28: *Heatmap of the maximum Cross-Correlation of the 196 basis functions (natural disparities)*

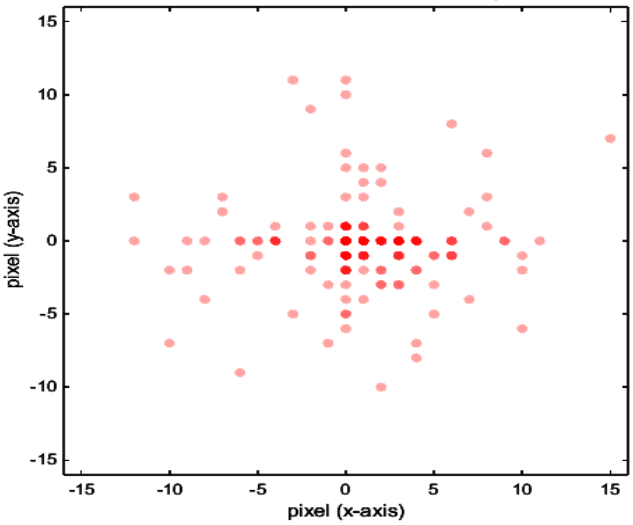

Figure 29: *Maximum Cross-Correlation of the 196 basis functions (natural disparities).*

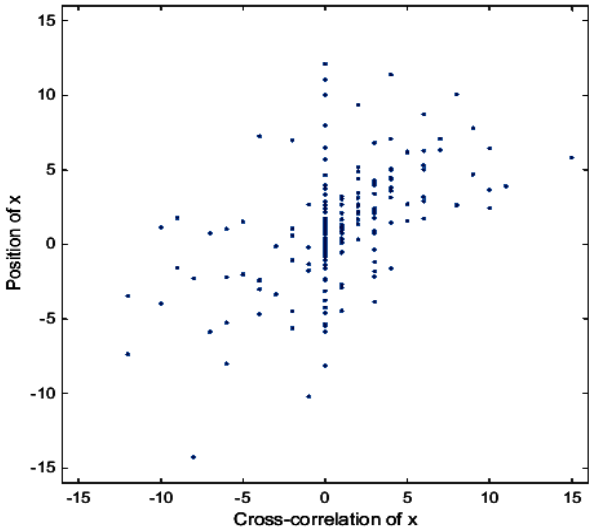

Figure 30: *Correlation of the position-shift in x-axis and cross-correlation of the x-axis (196 basis functions, natural disparities)*

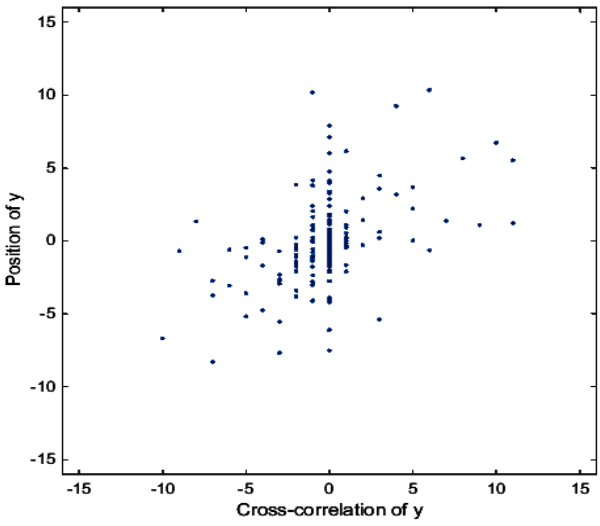

Figure 31: *Correlation of the position-shift in y-axis and cross-correlation of the y-axis (196 basis functions, natural disparities)*

4. Discussion

In the present study a learning algorithm was used that finds a sparse code for natural binocular images to develop binocular receptive fields (represented by basis functions). To analyze whether the disparity selectivity is generated by a position-shift model or by a phase-shift model, the resulting basis functions were fitted with Gabor functions. As an additional and simpler test for the location of the optimally tuned disparity of the basis functions, we further analyzed the cross correlations of the left vs. the right base. In order to get a better understanding of how disparities are coded we then correlated the maximum position of the cross correlation with the position-shift in x- and y-direction.

Hubel and Wiesel (1962) conjectured that disparity emerges through a position-shift of the monocular left and right receptive fields. On the contrary, Ohzawa, DeAngelis and Freeman (1990), have suggested a phase-shift model to explain how disparity might be measured.

Nonetheless a number of scientists (e.g. Fleet, Wagner & Heeger, 1995; Prince, Cumming & Parker, 2002; Tsao et al., 2003) have shown that the position-shift-model and the phase-shift-model are both common in binocular receptive fields. Fleet et al. (1995) presented a formal description and analysis on the contribution of both, phase-shift model and position-shift model, in a binocular energy model. Prince et al. (2002) also concluded that both models are commonly represented in binocular receptive fields. They found a large number of disparity-sensitive neurons in V1 of the awake macaque by recording their responses to dynamic random dot stereograms and fitting the neurons with one-dimensional Gabor functions. Tsao et al. (2003) tested disparity-tuned simple-cells in alert macaque, that were fixating a random dot stereogram. They extracted phase and position disparities by fitting a Gabor function to a single curve (an intercolar cross-correlogram) (A correlogram is an image of a correlation). The cross-correlation of the left and right receptive fields showed phase-shifts, position-shifts, as well as a hybrid of both.
Therefore, we also expected to obtain phase- and position-shifts and even hybrids of both in binocular receptive fields by applying sparse coding.

The 49 basis functions that learned randomly extracted corresponding patches from binocular images developed mostly dotted receptive fields when the disparities were set (constant disparities) (Figure 8), but developed rather elongated receptive fields when disparities were naturally given in the stereo images (natural disparities) (Figure 16).

Whereas, the 196 basis functions that learned patches from binocular images with constant and natural disparities developed mostly Gabor-like receptive fields (Figure 15,24).
This leads to the assumption that more basis functions ensure a more thorough and complete description of the input that is being learned (Overcompleteness of the representation) (Olshausen, 2003, Field, 1994) (see *Introduction* and *Limitations* for further explanations).

By looking closely at the learned corresponding bases of stereo images with *constant disparities* (196 basis functions), one can differentiate between basis functions that indicate a phase-shift (Figure 32, left) and ones that emerge from a position-shift (Figure 32, middle), as well as basis functions that contain both shifts (Figure B, right).

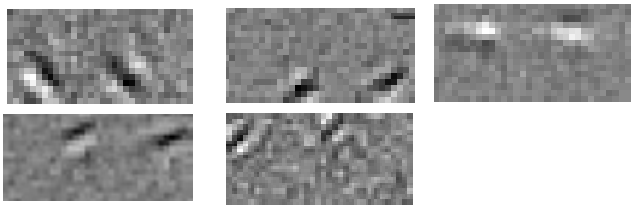

Figure 32: *Examples of basis functions with phase-shifts (left), position-shifts (middle) and a hybrid of both (right).*

Despite of seeing these shifts simply by looking at them, the fitted parameters of the Gabor functions of the 49 learned basis functions clearly show that phase-shifts and position-shifts exist (Figure 9, 10).
As for the position-shift, the cross-correlation, as well as for the correlations we measured disparities in a range of ca. -3.5 to 3.5 pixels, which was expected because they resemble the disparities we created for the input images that were learned (Figure 10, 11, 12).
The correlations of the cross-correlations and the position-shifts in x- and y-direction were quite weak. One explanation could be that a combination of both, phase and position, is important to ensure a high correlation with the cross-correlation. The Gaussian coarsely determines disparities, whereas phase does it more subtly.

In the case of the 196 basis functions that learned stereo images with *natural disparities*, one basis often showed response, while the corresponding basis remarkably often showed nothing (Figure 24). Overlarge disparities could be the reason for such activity pattern (see *Limitations*); as mentioned in the methods, we already shrunk the images down to get results. Another explanation is that the applied algorithm might develop a set of monocular Gabor-like receptive fields and at the same time might find disparities.

Analyzed fitted Gabor functions showed occurring phase- and position-shifts (Figures 26, 27), which can directly be seen by plainly looking at the resulting 196 basis functions.
The correlations of the cross-correlation of the 49 basis functions and the position-shifts in x- and y-direction were relatively strong, indicating that the impact of a phase-shift might have been relatively low, especially compared to the 196 basis functions, where the correlations were only mid-strong.

The given results account for the sufficiency of an algorithm that finds a sparse code for natural binocular images with either constant or natural disparities.

Limitations

Despite of these pleasing results, this study faced some limitations, we want to address here:

1. We needed to vary the sparsity value because for some inputs it was either too small or too big, which led the algorithm to not find the optimal sparse code. A not suitable sparsity value could explain the noise obtained for a lot of patches. Even more iterations might help to get rid of the noise. Aside from the sparsity value, the learning rate may also crucially contribute to issues in finding the optimal sparse code, i.e. if it is too large.
2. The number of images may not have been sufficient as an input to exploit the full spectrum of disparities existing in natural images and that can be detected by disparity-tuned cells. In a small number of images, the cells might not find a large amount of similar disparities, especially contemplating the fact that 5000 corresponding patches are extracted randomly.
3. When dealt with the natural binocular images with natural disparities that we downloaded from the internet, we faced some issues regarding their size. Large images often contain overlarge disparities that cannot be detected anymore, therefore they needed to be shrunken down.
4. An overcomplete representation was not given (dimensionality of the output (basis functions) is greater than the dimensionality of the input). 49 (and 196) basis functions are by far not sufficient to meaningfully represent all important and descriptive properties consisting in 5000 randomly extracted 16 x 16 patches. Therefore a complete representation would here consist of at least 512 basis functions.
5. We analyzed the sine wave and the Gaussian of the Gabor function separately and obtained the results for the phase- and position-shifts. Analyzing the phase-shift relative to the position-shift might give a better understanding on their relationship and contribution to

selectivity to disparity in disparity-tuned cells.

6. In the analysis we shouldn't have included those bases, that only showed a fitted Gaussian, because the sinus term didn't take any part in those fits.

The unsupervised learning algorithm was created by Olshausen and Field (1996) to find a sparse code for natural images in order to develop receptive fields, which resemble those of simple cells in V1. One might, thus, argue that sparse coding may not be ideally suitable for stereo images, contemplating the fact that stereoptic depth perception is processed at later stages of the visual cortex, i.e. by complex cells which are built up by simple cells (Hubel & Wiesel, 1962). Additionally, Ohzawa et al. (1990) found that complex cells are well suited as disparity detectors because they are not sensitive to polarity and position as simple cells are. Even though Olshausen and Field (1996) successfully showed how sparse coding develops monocular receptive fields, the algorithm may not be sufficient for binocular images. However, we must not forget that information about stereoptic depth vision, as well as any visual information, is perceived firstly by simple and complex cells at early stages of the visual cortex. We know that the brain processes and filters information that is perceived but what strategies does it employ and why? Therefore, an unsupervised learning algorithm with the principle of sparse coding, proposed by Olshausen and Field (1994) is a great approach toward these questions. Especially because it takes into account the statistical structure of natural images.

Barlow (1989) excellently explained that *unsupervised* or *self-organized learning* algorithms optimally extract characteristic features of the input data by learning its statistical regularities. The brain might do this by effectively exploiting redundant information that is contained in visual perception. For a system that works independently and in a self-organizing manner, it is rather important to be given an input of redundant information to learn the most characteristic and commonly occurring features in the input. Such systems are highly efficient because they intrinsically and dynamically regulate themselves to produce the most optimal output.

Learning, regulating and adapting in a self-organizing learning system, such as the Hebb learning rule, contributes to synaptic modification and therefore to the principle of synaptic plasticity in the brain.

Indeed, the brain can be seen as a self-organizing system itself. Especially contemplating the fact that *supervised learning* in the brain or in *biological systems* is rather uncommon (Dayan & Abbott, 2001).

It may also fascinatingly explain how the brain's underlying mechanisms are not isolated systems,

but rather, they influence and support each other, merging to one system.

Understanding why these systems act as they do could help us to have a better and thorough understanding of the brain itself. In order to understand the brain we cannot just focus on one system, for instance, the visual system. Exploring the underlying functions helps us to gain important insights on how the system works and reacts. But the question on *why* they work that way remains still unclear. Computational methods, as the ones we discussed above, allow us to dig deeper into this matter. They are based on mathematical and statistical approaches, tying the relationship between neuroscience and mathematics, statistics.

Nevertheless we must not forget that computers are not by far like the brain and vice versa. For instance, the computer works linearly and the brain non-linearly. The Non-linearity of the brain might make it difficult for scientists to develop algorithms that work in that manner.

However, in the computer vision field (and other fields), a lot of scientists have managed to construct algorithms that work in a self-organizing manner, thus not under supervision. Their approach is fundamentally important for further investigations on the human brain and specifically on the human visual system.

References

Barlow, H. B., Blakemore, C., & Pettigrew, J. D. (1967). The neural mechanism of binocular depth discrimination. *The Journal of physiology*, *193*(2), 327-342.

Barlow, H. B. (1972). Single units and sensation: a neuron doctrine for perceptual psychology? *Perception* 1(4):371-94

Barlow, H. B. (1989). Unsupervised learning. *Neural computation*, *1*(3), 295-311.

Dayan, P., & Abbott, L. F. (2001). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems.* London: The MIT Press.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *JOSA A*, *4*(12), 2379-2394.

Field, D. J. (1994). What is the goal of sensory coding?. *Neural computation*, *6*(4), 559-601.

Fleet, D. J., Wagner, H., & Heeger, D. J. (1996). Neural encoding of binocular disparity: energy models, position shifts and phase shifts. *Vision research*, *36*(12), 1839-1857.

Honglak L., Battle A., Raina, R., & Ng, A. Y. (2007). Efficient sparse coding algorithms. *Advances in Neural Information Processing Systems, (NIPS) 19*.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, *160*(1), 106.

Jones, J. P., & Palmer, L. A. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of neurophysiology*, *58*(6), 1233-1258.

Olshausen, B. A., Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*(6583), 607-609.

Olshausen, B. A. (2003). Principles of image representation in visual cortex. *The visual neurosciences*, *2*, 1603-1615.

Olshausen, B. A., & Field, D. J. (2004). Sparse coding of sensory inputs. *Current opinion in neurobiology*, *14*(4), 481-487.

Ohzawa, I., DeAngelis, G. C., & Freeman, R. D. (1990). Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science*, *249*(4972), 1037-1041.

Poggio, G. F., & Poggio, T. (1984). The analysis of stereopsis. *Annual review of neuroscience*, *7*(1), 379-412.

Prince, S. J. D., Cumming, B. G., & Parker, A. J. (2002). Range and mechanism of encoding of h orizontal disparity in macaque V1. *Journal of Neurophysiology*, *87*(1), 209-221.

Tsao, D. Y., Conway, B. R., & Livingstone, M. S. (2003). Receptive fields of disparity-tuned simple cells in macaque V1. *Neuron*, *38*(1), 103-114.

Wolfe, J. M., Kluender, K., R., Levi, D. M., Bartoshuk, L. M., Herz, R., S., Klatzky, R., L., ...Merfeld, D. M. (2009). *Sensation & Perception* (2nd ed.). Sunderland: Sinauer Associates, Inc.

Zhu, Y. D., & Qian, N. (1996). Binocular receptive field models, disparity tuning, and characteristic disparity. *Neural Computation*, *8*(8), 1611-1641.

Selbstständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Bachelorarbeit selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt habe. Alle Stellen, die dem Wortlaut oder dem Sinne nach anderen Werken entnommen sind, sind durch Angaben von Quellen als Entlehnung kenntlich gemacht. Diese Bachelorarbeit wurde in gleicher oder ähnlicher Form in keinem anderen Studiengang als Prüfungsleistung vorgelegt.

Ort, Datum Unterschrift