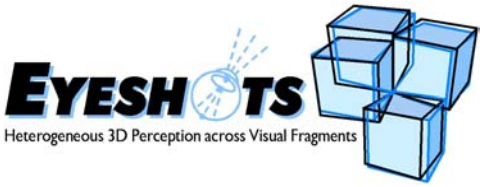


Project no.: FP7-ICT-217077
Project full title: Heterogeneous 3-D Perception across Visual Fragments
Project Acronym: EYESHOTS
Deliverable no: D3.1b
Title of the deliverable: Demonstration of object selective cells at intermediate complexity showing properties of disparity

Date of Delivery:	10 March 2010	
Organization name of lead contractor for this deliverable:	WWU (F.H. Hamker)	
Author(s):	M.A. Voss, J. Wiltschut, F. Beuth and F.H. Hamker	
Participant(s):	WWU	
Workpackage contributing to the deliverable:	WP3	
Nature:	Report	
Version:	1.1	
Total number of pages:	27	
Responsible person:	Fred H. Hamker	
Revised by:	Marc Van Hulle	
Start date of project:	1 March 2008	**Duration:** 36 months

Project Co-funded by the European Commission within the Seventh Framework Programme		
Dissemination Level		
PU	Public	**X**
PP	Restricted to other program participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Abstract:

The report investigates the question of how the human brain can utilize binocular information for object recognition. The brain uses an efficient coding to represent disparity and object selectivity. We will present the results of our computational models of the early (V1) and mid level (V2, V4) stages of the visual stream. The first part focuses on self-organizing and Hebbian learning of receptive field in V1 which results in localized, oriented, disparity tuned and band-pass filtering receptive fields. The second part extends the learning process to higher cortical areas and aims more on object selectivity as on disparity. It demonstrates the learning of receptive fields at an intermediate complexity level using edge and disparity information.

Contents

1 Executive summary

This document contains the Technical Report for Dilverable D.1b. "Demonstration of object selective cells at intermediate complexity showing properties of disparity". We want to investigate the open question of how the brain could use the information of both eyes for object detection. This issue includes the role of disparity to detect depth.

The European project "Eyeshots" focuses on the research of a visuo-motor system which is based on the concept of "active and fragmented vision". The primate brain actively generates a cognitive interpretation of a perceived scene. It doesn't encode the scene as pure 2D images or reconstruct real 3D data. Instead, it creates an efficient code in which a scene consists of distributed and loose features.

The learning in the primate brain is not limited to the early visual areas and we want to extend our learning approach to higher areas. As the complexity of this task is very extensive we split it into two subtasks. The first one has the goal to learn simple cells in V1 and compare them with biological data. For this purpose we use a selforganising learning algorithm [38] with post-synaptic inhibition and Hebbian learning. It allows the cells to tune themselves to an efficient coding of disparities and features in scenes. The previous Report (Software module) D.1a. "Demonstration of learning disparity-tuned feature selective cells" already demonstrated (as a preliminary result) simple cells which learn receptive fields tuned to localized, oriented, disparity tuned and band passed filters, comparable to those in area V1 of the primate brain. We will now extend the Deliverable D.1a to complete evaluation.

The second subtask focuses on the extension of the learning process to higher cortical areas (in this case the areas V2 / V4). We use specialized filters realized by a disparity energy model. The filter bank is used by the partners in WP2 and was provided within the Eyeshots consortium. As a side goal, our usage will strengthen the cooperation within the project "Eyeshots". The learning algorithm uses the Hebbian principle and apply a trace learning paradigm to pool stimuli that belong to the same object. We will show that our learnt cells can discriminate different objects and display at the same time disparity tunings. The task split allows us to work in parallel on the subtasks.

Therefore, the document will consist of two main parts. In the first one (Section 3), we will evaluate the learning of simple cells like in V1.

In the second part (Section 4) we present the extension to higher visual areas to complex disparity and object tuned cells. We found these different types of complex cells: 1) pure object selective, 2) disparity tuned cells to create object selective and 3) mixed ones. After presenting issues concerning the complexity of the cells, we will discuss the disparity and object selectivity of the learnt cells.

The results will help to develop the approach of the software modul D3.2a "Object-based top-down selection using learned bi-directional connections between feature detectors to localize the object of interest in a cluttered 3D scene" which will be able to detect an object in cluttered scenes.

2 Introduction

The human brain actively generates a cognitive interpretation of a perceived scene. It doesn't save the scene as a pure 2D image or reconstruct a 3D model. Rather, it creates a very efficient code in which a scene consists of distributed and loose features. Primates use information of both eyes to improve object recognition in a scene.

We here describe the research background of learning disparity and object selective cells in the visual pathway. In the history, models of primary visual cortex (V1) encoding disparity have primarily been constructed by hand to fit specific data. Little work has been done on developing learning algorithms that lead to generalized receptive fields similar as observed in the brain. For this reason, we developed a general learning algorithm[38] which is motivated by biological research to capture the basic principles of primate 3D perception. It uses post-synaptic inhibition and Hebbian learning and is able to learn simultaneously feed forward (stimuli driven) and feedback (attention driven) connections.

Since the early studies of receptive field properties in the primary visual cortex (see Hubel & Wiesel[11] or [35, 7]), one of the major question in neural coding has been why neurons have a particular receptive field structure. Because V1 neurons respond well to edges, edge detection has been considered as a useful operation of early vision which should represent the important structural features of a visual scene[16]. Information from the left and right eye are first combined in V1, where many cells are tuned to binocular disparity. Complex cells in V1 often show a tuning to a preferred disparity. But V1 is not the source of stereoscopic depth perception, it only provides local estimates of absolute disparity. The neurons of the next higher area, V2, begin to show sensitivity to relative disparity and for disparity-defined edges ([37, 30]) and V4 neurons show selectivity for disparity defined planes. Section 3 will focus on simple disparity cells which can be compared with simple cells in V1 and section 4 will focus on cells of intermediate complexity level which are object selective and can be compared to complex cells in V2, or V4.

3 Analysis of learned binocular disparity-tuned feature selective cells

The neural process of stereoscopic depth discrimination is thought to be initiated in the primary visual cortex. We have presented a model of primary visual cortex with nonlinear dynamics and Hebbian learning for monocular [38] and binocular image stimuli [D3.1a]. Here, we want to present the analysis of the disparity-tuned feature selective cells that were learned by presenting natural stereo image scenes to our model.

3.1 Methods

3.1.1 Stereo input

Information from the left and right eye are first combined in V1 and many cells are tuned to binocular disparity. To learn our simple, V1 like model cells we used a binocular image set of 40 image pairs. From these images we randomly cut out 400.000 12x12 pixel image patches at corresponding positions in the two views in a small, central region around the point of fixation.

3.1.2 Probing disparity tuning with displaced random dot images

For a comparison with physiological data and an evaluation of the characteristics with respect to disparities of our learned binocular simple cells we have probed their horizontal and vertical disparity tuning with random dot images (RDI). We used 100 different RDIs as a basis to displace them horizontally and vertically and let our model with fixed, learned V1 like simple cells respond to these inputs. We took the mean of the responses over all 100 random dot inputs with the same vertical and horizontal displacement to obtain a response that is independent from the structure of the learned receptive fields but that represents its disparity tuning for a specific combination of vertical and horizontal disparity.

3.1.3 Fitting receptive fields

Corresponding to physiological data many of the learned receptive fields exhibit a structure that can be fitted with a Gabor function (G) [10] quite well with a Mean Square Error (MSE) of about $MSE < 0.1$. Basically a Gabor function is the product of a Gaussian envelope and a sinusoid. We used a Levenberg-Marquardt method [27], a two-dimensional (2D) Gabor function and minimize the MSE. The function has 10 free parameters, namely the center coordinate of the Gaussian on the x-axis (X_0) and y-axis (Y_0), the rotation angle of the Gaussian (γ), the width of the Gaussian along the minor axis (W_p) and the major axis (W_q), the frequency (f), the phase (ϕ) and the rotation angle (θ) of the sinusoid, the amplitude (A_m) and the amplitude offset (A_o) to fit our learned receptive fields. With these parameters the Gabor function can be formulated as

$$G = E \times S + A_o \tag{1}$$

$$E = A_m \times \exp\left(\frac{-p^2}{2W_p^2}\right) \times \exp\left(\frac{-q^2}{2W_q^2}\right) \tag{2}$$

$$p = (X - X_0)\cos\gamma + (Y - Y_0)\sin\gamma \tag{3}$$

$$q = -(X - X_0)\sin\gamma + (Y - Y_0)\cos\gamma \tag{4}$$

$$S = \cos\left(2\pi f u + \phi\right) \tag{5}$$

$$u = (X - X_0)\cos\theta + (Y - Y_0)\sin\theta. \tag{6}$$

For the analysis of learned receptive fields with the help of the parameters of the best fits one has to take the ambiguity of the Gabor function into account as $G(..., \theta, \phi) = G(..., \theta + \pi, \phi + \pi)$. For the fitting procedure the range of θ has been restricted to values between 0 and π. The ambiguity can cause large differences between parameters of the Gabor fits of the left receptive field and the right one even when they are very similar in their structure.

3.2 Results

3.2.1 Disparity tuning curves for exemplary cells

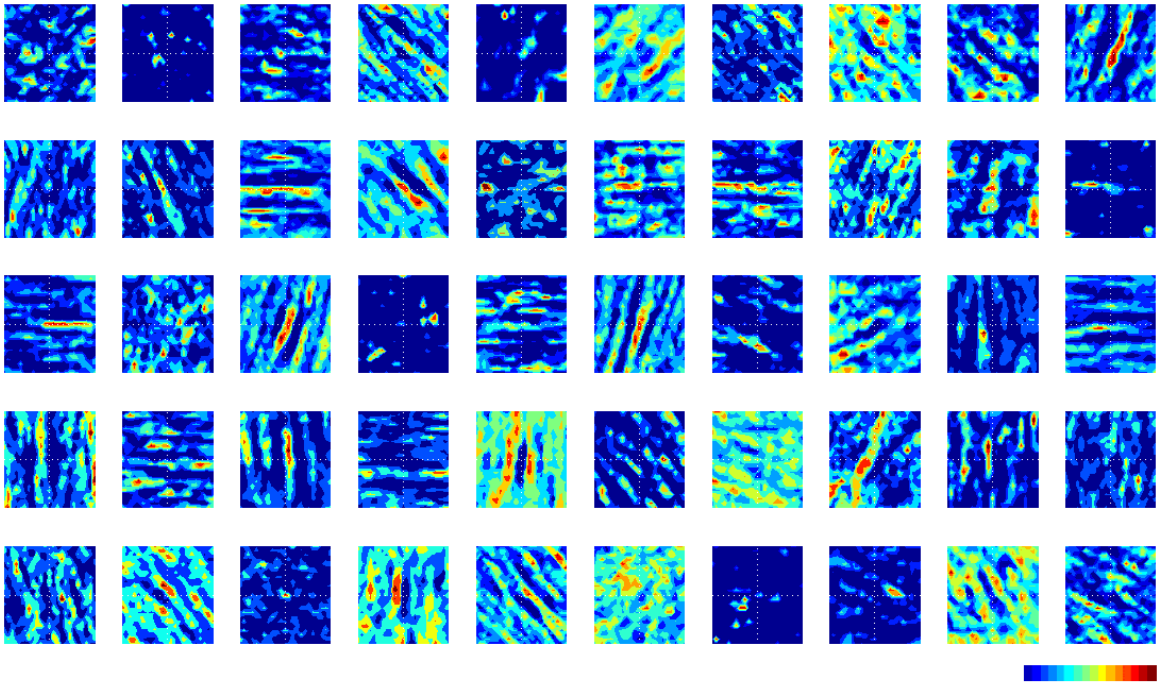


Figure 1: Mean responses of exemplary learned simple cells to vertically and horizontally displaced random dot stimuli. The dotted lines indicate zero horizontal (x-axis) and vertical disparity (y-axis).

Some exemplary results of probing our learned V1 like cells with RDIs 3.1.2 are shown in Figure 1.

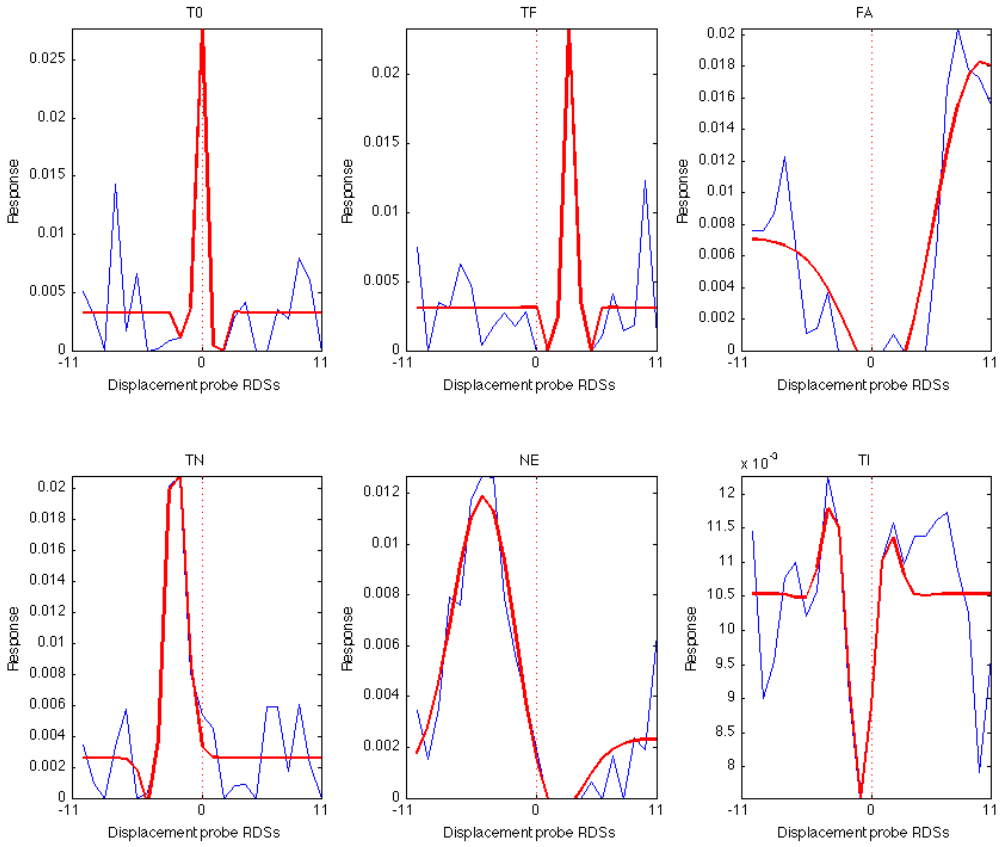


Figure 2: Exemplary cells for the categorization introduced by Poggio and collaborators. From top left to bottom right T0= tuned-excitatory, TF=tuned-far, FA=far, TN=tuned-near, NE=near and TI=tuned-inhibitory cells.

Our results show no indications of a classification of tuning characteristics into distinct classes as introduced by Poggio and collaborators [22, 23, 26, 24]. Examples of receptive fields that show a tuning characteristic exemplary for each response category were found as shown in Figure 2 but overall there was no clustering that would support such a distinct categorization of disparity tuning. This is also consistent with recent studies [28, 2, 20]. The exemplary cells presented in Figure 2 show the horizontal disparity tuning with no vertical disparity. For a comparison with other data we also show the 1D Gabor fit in red. The disparity-tuning response for the full range of horizontal and vertical disparities could not be fitted with a 2D Gabor as the structure of the responses was significantly different and more complex. This could indicate that Gabors are in either case not the best choice to describe the disparity-tuning characteristics of these binocular simple cells.

3.2.2 Ocular dominance

Using the amplitude A_m of the Gabor functions fitted to the receptive fields (RF) visualized in Figure 3 as an indicator for ocular dominance, the learned cells do not exhibit a preference for monocular or binocular tuning characteristics. The whole range between monocular cell pairs with a high amplitude A_m in one of the eyes and a very low one in the other - to cells with an identical amplitude in both eyes and all variations in between get represented pretty evenly. Most cells seem to be clustered around combinations with a constant sum of the left and right amplitude.

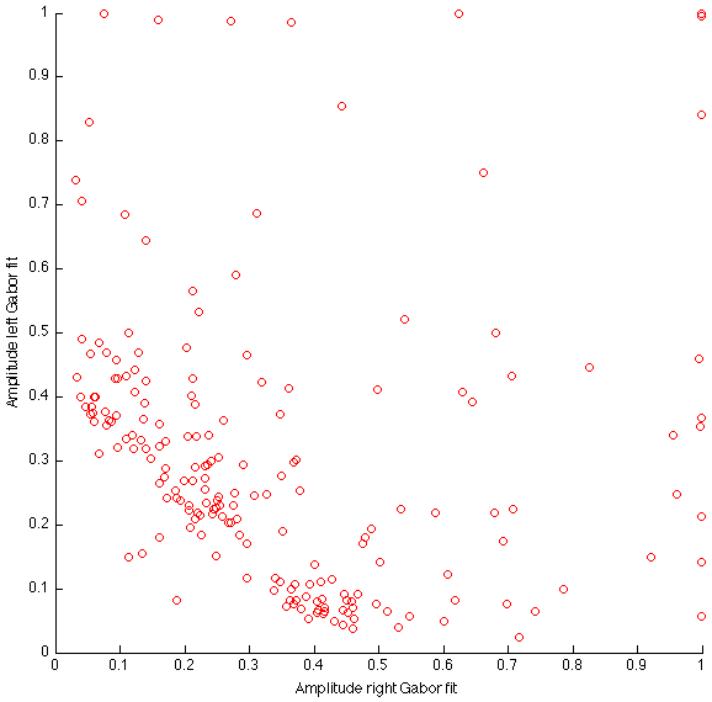


Figure 3: Amplitudes A_m of the fitted Gabor functions of the left (y-axis) and right eye (x-axis).

3.2.3 Orientation similarity left - right

To compare the preferred orientation in the left and right eye estimated by the parameter θ of G one has to be aware of the ambiguity described in 3.1.3 and its consequences. Here we have chosen to adjust the orientation and phase of one of the Gabor fits if the absolute value of the orientation-difference between the Gabor fits of the binocular receptive fields in the two eyes is higher than 0.5π. Most of these pairs show an orientation angle close to zero in one eye and an orientation angle close to π in the other one. Minimizing the error introduced through this adjustment the angle with the smallest absolute difference to

zero or π gets set to the opposite edge of the range. Even though this practice introduces a small error in one of the orientations we have chosen this way opposed to discard these fits or use the fitted parameters with no adjustment at all. From 194 neurons used in the subsequent analysis we adjusted the phase and orientation of 23 in this way.

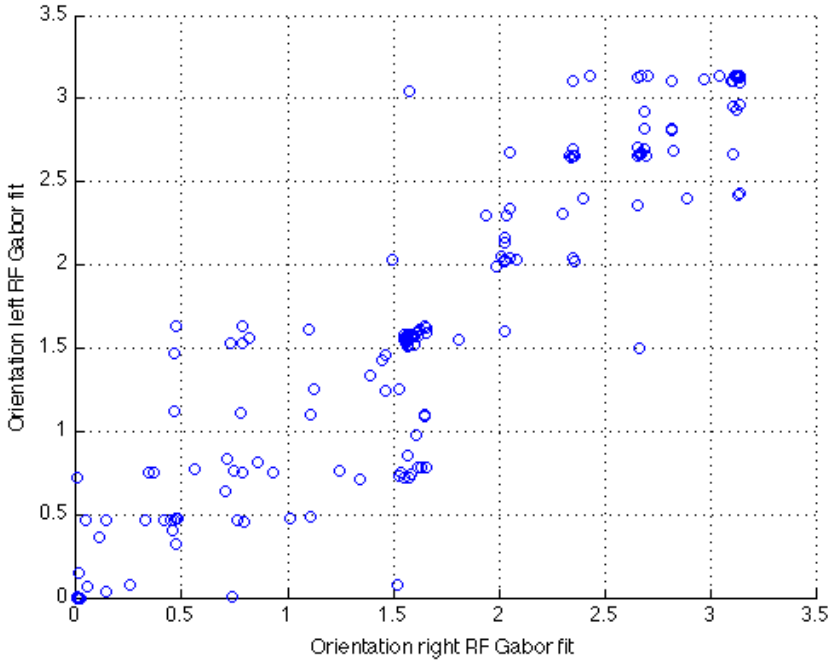


Figure 4: Orientations θ (radian) of the sinusoid being part of the 2D Gabor functions fitted to the learnt receptive fields in the left (y-axis) and right eye (x-axis).

The orientation similarity in the receptive fields of our learned cells in the two eyes is shown in Figure 4. There is a strong correlation between the orientations θ of the learned receptive field fits of both eyes to that effect that the orientations in both eyes only differ in a small magnitude. The similarity of the orientations of the learned receptive fields can be illustrated even better with a histogram of the absolute difference of the angles of the two Gabor fits as shown in Figure 5. Another advantage of this illustration is that one can calculate the absolute difference without introducing an error from the fitted parameters in contrast to the procedure described before to visualize the orientations. The results are in line with data from the cat [4, 17] and also with data from macaque V1 neurons [5] and all of these suggest that binocular neurons in the early visual system have similar orientations in both receptive fields and therefore have no separate encoding scheme for orientation disparities.

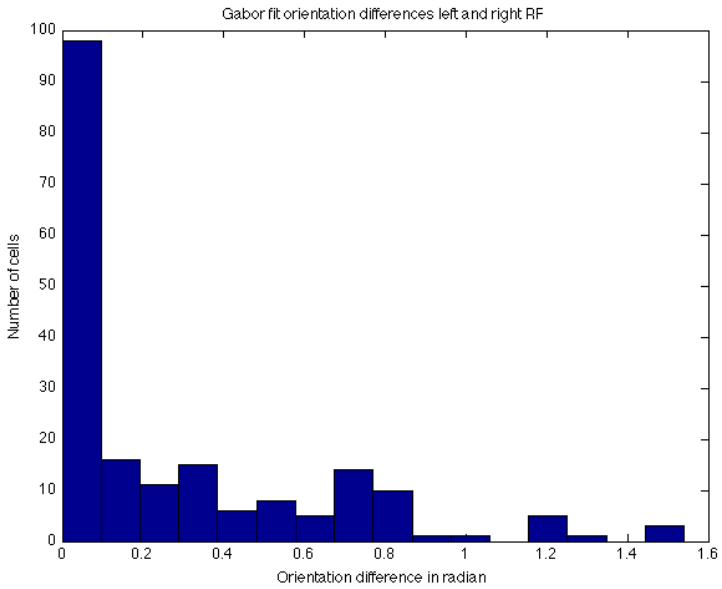


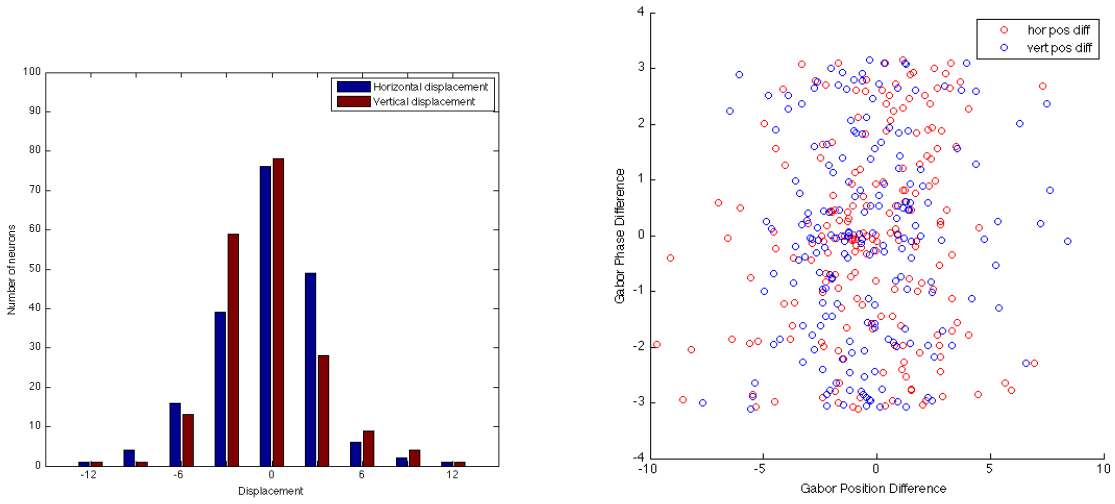
Figure 5: Histogram of the orientation differences between the Gabor fit for the left and the right receptive field.

3.2.4 Orientation anisotropy horizontal - vertical

Barlow et al. [3] have found an orientation anisotropy for horizontal and vertical disparities as they measured a larger range for horizontal ($\pm 3.3°$) than for vertical disparities ($\pm 1.1°$). In contrast to that subsequent studies [8, 13, 15, 18, 36] did not show such an anisotropy for small eccentricities. The results of our learned receptive fields also exhibit little differences of the encoding range for binocular disparities between horizontal and vertical disparities as Figure 6(a) indicates.

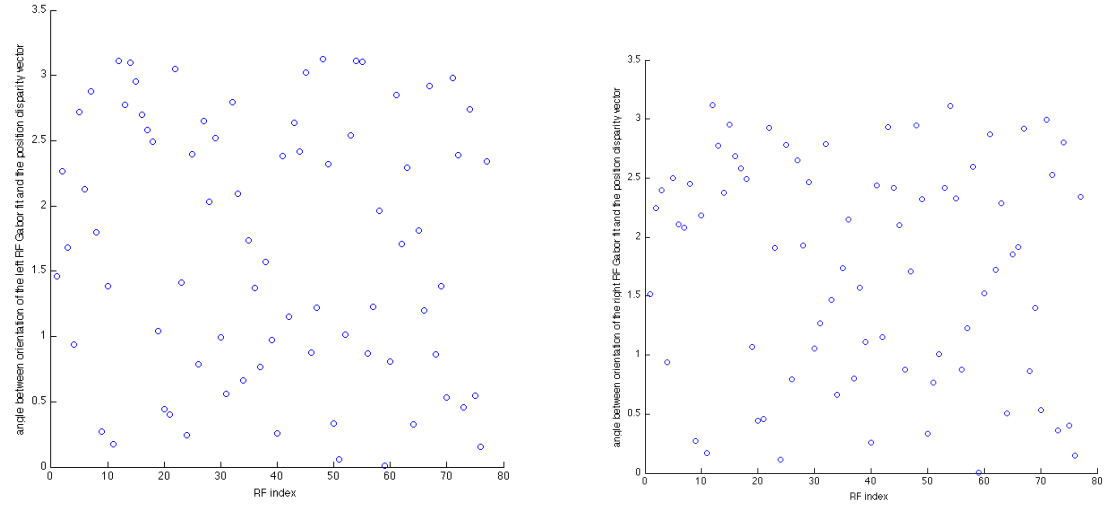
3.2.5 Phase and position

The learned receptive fields in our model show both phase and position-based mechanisms to encode binocular disparity. Our results suggest that there is no categorization into distinct classes of disparity encoding neurons into phase or position encoding neurons but that these two mechanisms to encode disparity might be used in single V1 neurons at the same time with smoothly varying influence over the whole population as visualized in Figure 6(b). Even though it is not clear so far what is the contribution of these two mechanisms to binocular vision other reports [28, 1, 2] agree with the view that both phase- and position-based mechanisms are used to encode disparities and their contribution to binocular vision was found to be very similar.



(a) (b)

Figure 6: (a) Horizontal and vertical displacement of the center of the Gaussians of the Gabor functions fitted to the left and right receptive fields. (b) Distribution of phase differences and position differences of the fitted, learned receptive fields.



(a) (b)

Figure 7: Angle in radian between the orientation of the left (a) and right (b) receptive field and the position difference vector.

3.2.6 Relation of RF orientation and disparity direction

As positional differences parallel to the orientation of the receptive fields cannot be detected, also known as the aperture problem, it would have advantages in many sit-

uations if neurons encoded disparities in the direction orthogonal to the orientation of their receptive fields. This assumption is also the basis for many artificially constructed disparity detectors. To compare our results with this model we have calculated the angle between the orientation of the sinusoid of the Gabor fits and the corresponding positional difference vector and the results are shown in Figure 7(a) for the left and Figure 7(b) for the right receptive fields. It is obvious that our results are not in compliance with the perpendicular model of the relation between disparity direction and receptive field orientationas there seems to be no correlation between these two parameters.

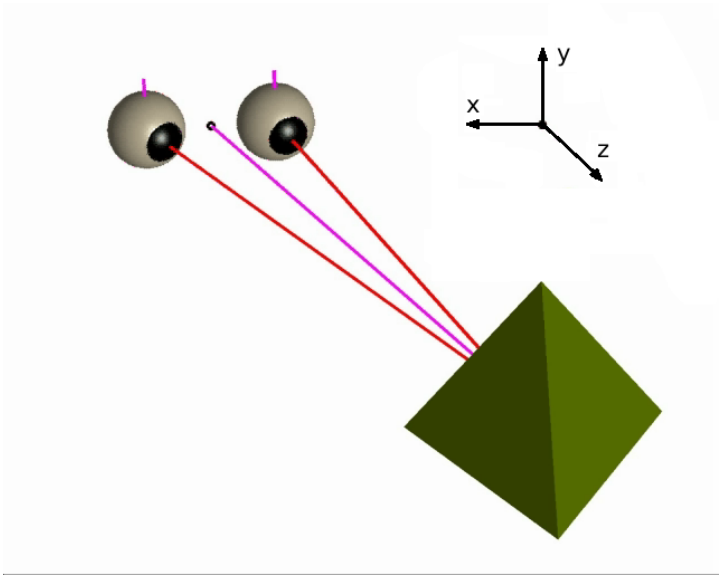


Figure 8: The coordinate system and the arrangement of the eyes, object and fixation position in the simulator.

4 Complex cells tuned to disparity and objects

The goal is to successfully use biological motivated learning mechanism to get objective selective cells with intermediate complexity that show disparity tuning characteristics. Disparity in a stereoscopic stimulus is changing if the object position is shifted (with constant fixation) or if the fixation point is shifted (with constant object position). We will test the cross combination of both effects. Each object will be presented for a probabilistic time period (about 10000 ms) and the network responses determine the adaption of the weights via a trace learning paradigm. We did not use a biologically implausible supervised learning paradigm.

4.1 Methods

We used two different ways to create disparity. (i) We moved the fixation position in the depth (z-axis) while maintaining a fixed object position and (ii) we held the fixation constant and moved the position of the object. We varied both, fixation as well as object positions and tested every combination of them.

4.1.1 Stimuli

The stimulus is the left and right eye view of an 3D object. We chose a raytracer engine[6] to produce the images and compiled 3D models of cubes to create 10 different objects (see Figure 9). Figure 8 shows the arrangement of the eyes, the object position and the fixation.

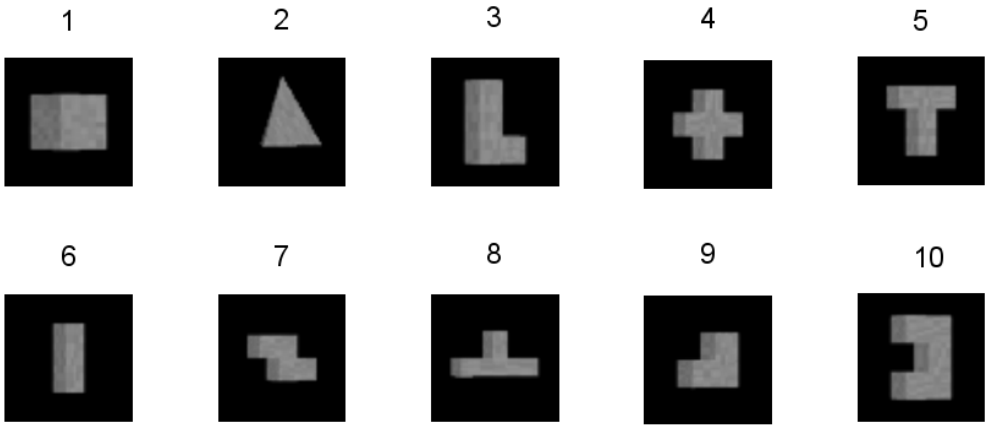


Figure 9: The stimuli consist of 10 different 3D objects. Here we show the image of the left eye.

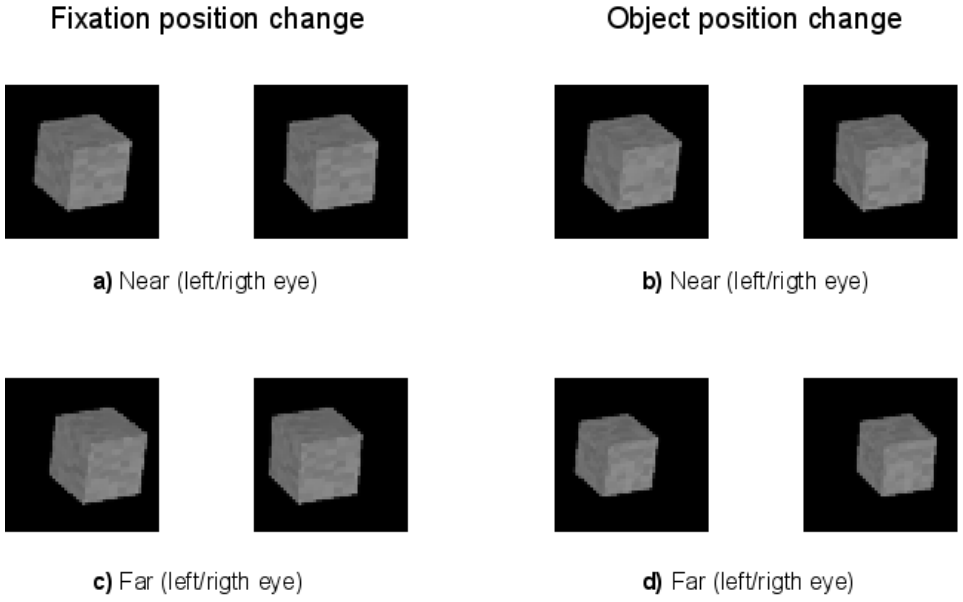


Figure 10: Shift/change of the object in the left/right eye stimuli due to different fixation or object positions. Both fixation and object positions are at 600mm for **a)** and **b)**. From **a)** to **c)**, the fixation position changes from 600 mm to 700 mm. From **b)** to **d)**, the object position changes from 600 mm to 700 mm. The figures are produced with different shift as in the model to get more clearer illustrations.

The eyes were simulated as virtual cameras with a distance of 66mm (like the distance of the human eyes) to each other. They are placed at the positions $(x, y, z) = (-33, 0, 0)$ mm for the left eye and $(+33, 0, 0)$ for the right eye (all positions are described as (x, y, z) vectors and all measurements are in millimeter). The filter bank[34] implements an energy model (see [19, 29, 31]) using 56 Gabors with 8 orientations (with a $\frac{\pi}{8}$ step size) and 7 different phase disparity shifts ($\pi \cdot \{-0.75, -0.5, \ldots, 0.75\}$). The envelope size of each Gabor filter is 11x11 pixel. The system can detect disparities in a range of about $[-1.5, +1.5]$ pixel and the 3D world arrangement must meet these constrains. The receptive field of the object selective cells should have a comparable size like a complex cell in V2 (see [12]) and we must select an appropriate aperture angle of the virtual cameras. We have choosen an aperture angle of 3° which encloses the whole object resp. 4.4° for the object with background (resulting in a stimulus like in Figure 9). Border effects are avoided with the additional background. With the mentioned constrains, we choose 20mm as the size of the objects. Object 1, the cube has exactly a side length of 20mm, the other objects have comparable sizes. The object resp. fixation positions are in a range of $[(0, 0, 610), (0, 0, 635)]$mm. The stimuli images were rendered at a resolution of 52×52 pixel.

These decisions fit the absolute disparities into the range of $[-1.5, +1.5]$ pixel and the object stays in the middle of the images. After processing the stimuli with the energy model, the resolution shrinks to 42×42 nodes due to border effects. From this we pool with a maximum rule (see [33, 32]) over 3×3 nodes and the resolution diminishes again to 14×14 nodes per features (orientation and phase of the Gabors). Thus, the next layer uses $14 \times 14 \times 56 \Rightarrow 10976$ input cells.

4.1.2 Complexity of the cells

Our model consists of three hierarchical layers, the responses of the Gabor filters, the pooled energy model responses and object selective cells. The degree of invariances is a core idea of the distinction between simple and complex cells in the brain. A receptive field of a simple cell is formed of spatial distinct excitatory and inhibitory regions where a complex cell combines these regions [35, 11]. The complexity of simple and of complex cells increases hierarchically in the brain. The complexity of the object selective cells will be tested by how far the cells respond invariant to location (see the x/y-plane in the 3D coordinate system), to features (which Gabor filter) or to disparity. The property disparity follows from the fixation and object position (see the z-axis in the 3D coordinate system). Therefore, we can loosely map the cells of our model to biological ones. The object recognition cells in the third layer loosely correspondent to V2 or V4 cells. They have the ability for spatial, feature and disparity invariance. In the section 4.2 we evaluate which properties were developed by means of the learning algorithm.

4.1.3 Fixation versus object position shifts

This paragraph describes the different image changes between a fixation position shift or an object position shift. The former one, the fixation shift results in a pure vertical shift of the object demonstrated by Figure 10a and 10c. The second case, the object position shift results in a vertical shift in the opposite direction and a shrinking in size (see Figure 10b and 10d). Therefore, a stimulus created via fixation change contains edge and disparity information and one from object position change contains edge, disparity and resized edge information.

4.1.4 Dynamic cell population

We use a rate coded neuron model where we describe input cells with firing rate r_i^{Input} and output cells with firing rate r_j. Each output cell gains excitation from all input cells (as a weighted sum) and is inhibited by all other output cells via Anti-Hebbian inhibition (similar as in [38]). The firing rate r_j of an output cell is computed over time using the following equation:

$$\tau \frac{\partial r_j}{\partial t} = \sum_i w_{ij} r_i^{\text{Input}} - \sum_{y, y \neq j} f\left(c_{jy} r_y\right) - r_j \tag{7}$$

where

$$f(x) = d_{nl} \cdot \log\left(\frac{1+x}{1-x}\right) \tag{8}$$

gives the non-linear processing. $\tau = 10$ is the time constant of the cells. The connection w_{ij} denotes the strength of the feed forward weight from input cell i to the output cell j. Lateral inhibition is given by the connection weight c_{jk} and can differ across the cells due to the Anti-Hebbian learning.

4.1.5 Learning

The connection weights w_{ij} are changed over time via the Hebbian principle using the equation that were successfully used to learn biological plausible receptive fields from natural scenes ([38]):

$$\tau_l \frac{dw_{ij}}{dt} = \left(r_j - \tilde{r}\right)^+ \left(\left(r_i^{\text{Input}} - \tilde{r}^{\text{Input}}\right) - \alpha_w \left(r_j - \tilde{r}\right)^+ w_{ij}\right) \tag{9}$$

$\tilde{r}$ is the mean of the activation of the particular cell population $\left(e.g., \tilde{r} = \frac{1}{N}\sum_{j=1}^{N} r_j\right)$ and τ_l the time constant for learning. α_w constrains the weights analogous to the Oja learning rule [21] and it holds $(x)^+ = \max(x, 0)$.

A simple trace learning paradigm was applied to pool stimuli that belong to the same object. During learning of the connection weights w_{ij} the input stimuli of the present

trial ($r_i^{\text{Input}}(t)$) is combined with the firing rate from the corresponding output cell of the previous trial (stimuli presentation) ($r_j(t-1)$).

Lateral connections between cells were learned by Anti-Hebbian learning. The name Anti-Hebbian implies that this strategy is the opposite of the Hebbian learning rule. Similar to the learning of the synaptic connection weights, where the connection between two cells is increased when both fire simultaneously, in the Anti-Hebbian case the inhibition between two cells is strengthened. The more frequent two cells are activated at the same time, the stronger they inhibit each other, increasing the competition among those two cells:

$$\tau_c \frac{\partial c_{ij}}{\partial t} = r_j \cdot r_i - \alpha_c r_j \cdot c_{ij} \tag{10}$$

where τ_c is the learning rate of the Anti-Hebbian weights. Anti-Hebbian learning leads to decorrelated responses and to a sparse code [9].

4.1.6 Discrimination value $\mathbf{d_{TM}}$

To determine the similarity of two population codes $(\mathbf{r}, \mathbf{s})$ the angle between those two vectors is regarded. Therefore, to compare two cell populations the following equation is used:

$$d_{TM}(\mathbf{r}, \mathbf{s}) = 1 - \frac{\langle \mathbf{r}, \mathbf{s} \rangle}{|\mathbf{r}|\,|\mathbf{s}|} \tag{11}$$

with $dim(\mathbf{r}) = \dim(\mathbf{s})$. The lower the value of $d_{TM} \in [0;1]$ the more the two vectors show similar cell distributions.

4.2 Results

Two questions should be solved with this model. The first one is the learning and recognition of objects using a binocular energy-based filter bank and the second one is to show disparity selectivity in the cell responses.

4.2.1 Object recognition

Our results show that regardless of the number of different objects and independently of the number of cells (as long that there is at least one cell per object) the model is able of learning and discriminating all objects. Figure 11 shows the average response of each cell to the different objects. It can be seen that each object is learnt by several cells and thus an object is characterized by a specific population code.

An analysis of whether the model is able to discriminate among the objects is shown in Figure 12. Using the discrimination value (d_{TM}) (see 4.1.6) the similarity of the average population responses between two objects is shown. Low values (indicated by darker areas) give clue to similar population codes which would indicate that discrimination

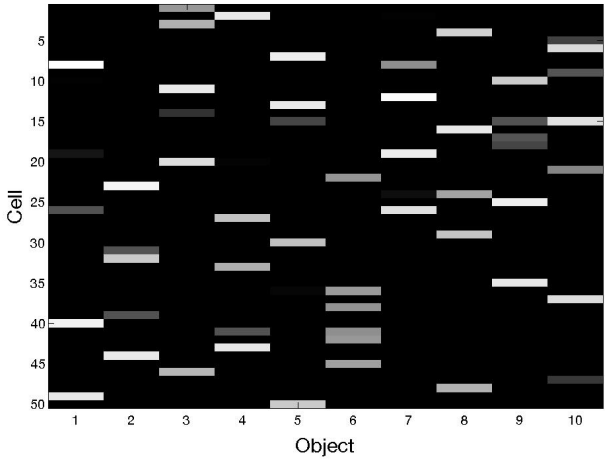


Figure 11: The selectivity of each cell for the objects. For each object (x-axis) the average firing response of each cell (y-axis) is plotted. The average firing response is calculated over all input stimuli that contain the same object. The strength of the average firing pattern to an object is characterized by the brightness (0 dark, 1 bright). Every object is represented by a combination of different cells with nearly no overlap to other objects.

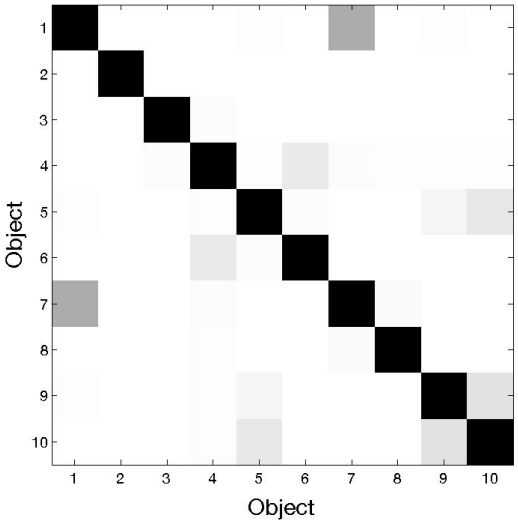
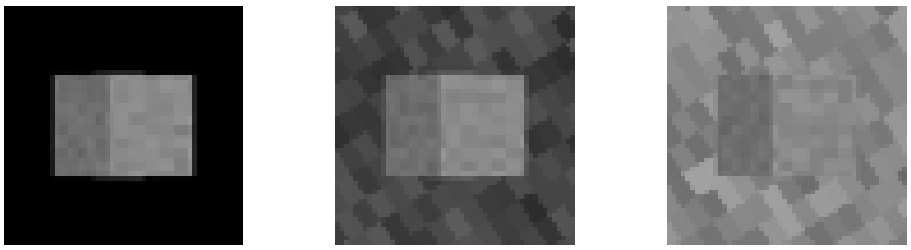


Figure 12: Discrimination between the objects. Using the discrimination value d_{TM} (see 4.1.6), the similarity of the average response (shown in Figure 11) to an object are shown here. The brightness correlates with the dissimilarity of the population code responding to the objects.



(a) No noise (b) Exp. Noise 1 (c) Exp. Noise 2

Figure 13: The stimuli without noise and the both experiments with noisy stimuli.

between those two objects is impaired. The results show that all objects are very dissimilar in their population code and thus are very easy to discriminate. Only object 1 and object 7 show slightly overlapping population codes ($d_{TM} = 0.68$) but the objects can easily be discriminated (compare Figure 11). Although some cells tend to code more than one object the results show that all objects can be discriminated perfectly.

The sensitivity of the object recognition ability was tested with noise embedded into the input stimuli. Figure 13 shows typical input stimuli for two experiments that use the same views of the objects but with a cluttered background. With such cluttered background the edge information becomes more difficult to use for the object recognition task and thus other sources like disparity have to be used in order to successfully discriminate the objects. Figure 14 shows the similarity of the responses to the different objects compared to the stimuli set without noise. The population response to an object in the normal condition (compare Figure 11) is compared to the cell responses of the other experiments for the same objects. All objects show a nearly perfect match in their population code compared to the experiment without noise, and thus also show the same good object discrimination ability. Only object 2 and 9 in the second noise experiment show slightly different population codes but their discrimination values are at least about 0.7 and a look into the raw data confirms that even in their case a good object discrimination is maintained.

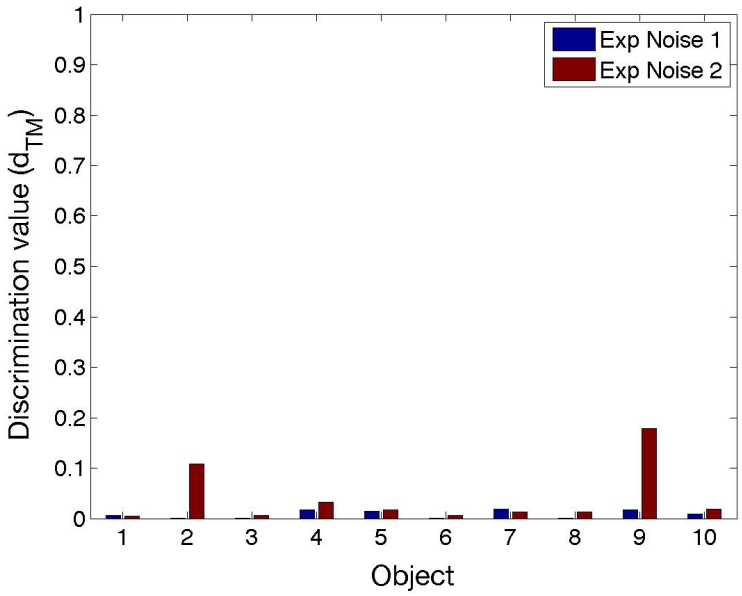


Figure 14: Discrimination ability despite noise. For each object the cell firing pattern response of the noise experiment is compared with the response to the same object without noise. Again, a low value indicates a good match between those to cell populations.

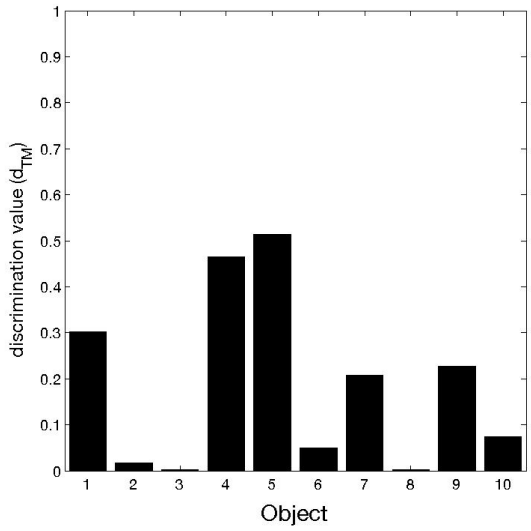
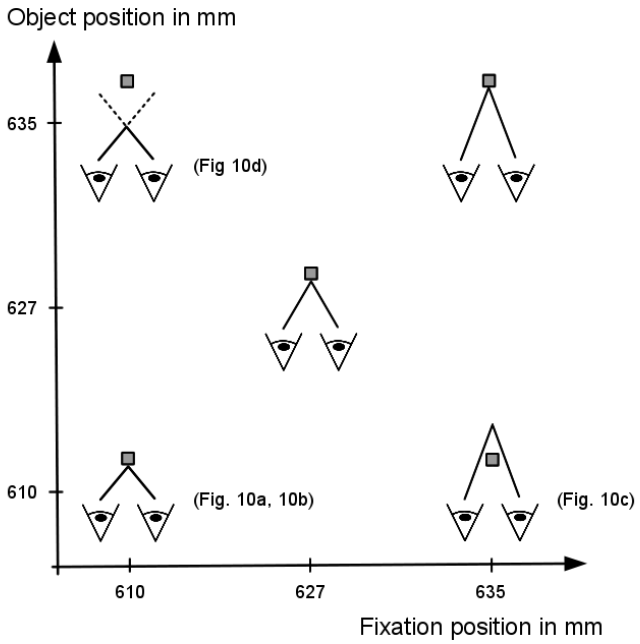
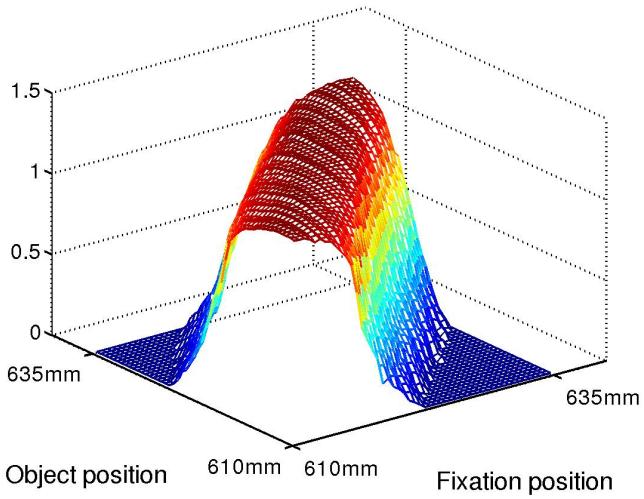


Figure 15: Similarity of the cells encoding the same object. The firing pattern for each cell is compared with the firing patterns of cells that encode the same object using the discrimination value d_{TM}. For some objects, the cells respond all with the same firing pattern (Object 2, 3, 6, 8 and 10) but for the other objects the cell responses show dissimilarity (Object 1, 4, 5, 7 and 9).

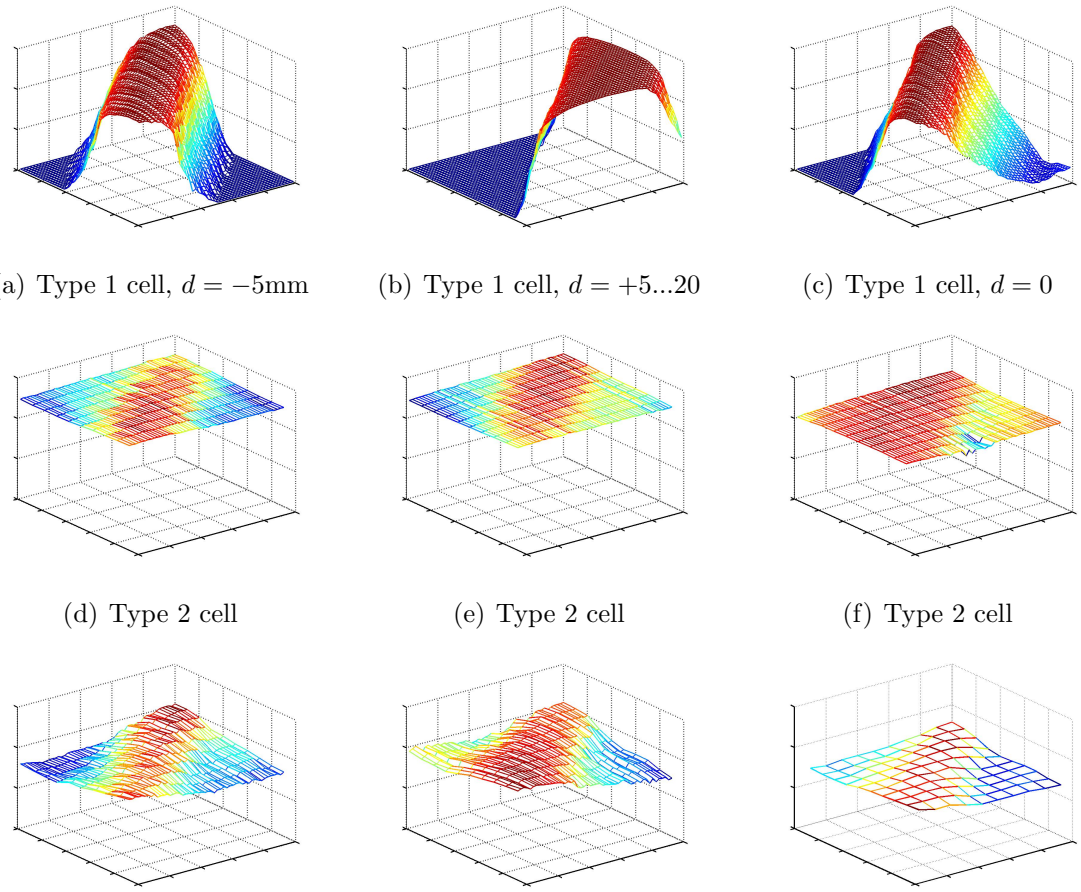


(a) Legend



(b) Example

Figure 16: **a)** The plot displays different combinations of fixation/object positions for b). and show the relative object positions for the examples in Figure 10. **b)** A tuning curve of a cell to a given object in relation to object position and fixation position. The Figure 17 will use this description again.



(a) Type 1 cell, $d = -5$mm (b) Type 1 cell, $d = +5...20$ (c) Type 1 cell, $d = 0$

(d) Type 2 cell (e) Type 2 cell (f) Type 2 cell

(g) Type 3 cell (h) Type 3 cell (i) Type 3 cell

Figure 17: Tuning characteristics of different cells: **a-c)** Type 1 cells: Each cell is tuned to a specific disparity in relation to the fixation point, for example the cells fire maximal for the distance d (object position minus fixation) of a) -5mm, b) between $+5$ to 20mm and c) 0mm. **d-f)** Type 2 cells: The cell is tuned to an object and is completely invariant to object position or fixation point position. **g-i)** Type 3 cells: The cell is a mix of the first two types. See Figure 16 for the legend.

4.2.2 Object and disparity selectivity

Although each object is coded by several cells, Figure 15 indicates that the population code of cells that code the same objects differ sometimes. The discrimination values between all cells that are tuned to the same object show that for half of the objects the cells do not respond with the same firing pattern. This is a first indicator that different cells have developed different tuning characteristics for the same object. A detailed look on the disparity tuning characteristic (Figure 17) reveals that there are three kinds of typical tuning behaviours:

- Type 1 cell: The cell is tuned to a specific disparity in relation to the fixation point (Figure 17 a - c). Thus it responds best to specific combinations of object distance and fixation point. This also includes a certain degree of pooling such as the cell responds only to a specific absolute disparity to the fixation point invariant in object position (along the z-axis).
- Type 2 cell: The cell is tuned to an object and is completely invariant to object position or fixation point position (Figure 17 d - f). This cell is a perfect object recognition cell, because its firing rate is independent of all but the information provided by the object itself. This might indicate that the cell is tuned to the edge information with the contribution of the object depth information.
- Type 3 cell: The cell is a mix of the first two types (Figure 17 g - i). The cell shows good object recognition (as Type 2) but with a slight preference to a specific absolute disparity to the fixation point (as Type 1).

All cell types have the capability of spatial (x/y-plane) and feature (Gabor type) invariance, because they fire not depending of the edge position (and also orientation) in the binocular stimulus. Thus, all types meet the complexity criteria in terms of spatial invariance as well as Gabor features and differ in the degree of disparity invariance.

5 Discussion

Object recognition in the brain is mostly driven by the features from the object itself. The most dominant features are given by the form (edge information) and by the texture (colour information) of the object. But other features might also play an important part in object recognition. One of those features is the disparity of the object. Even in a complete random dot image, where neither texture nor form information is given, the human brain is capable of perceiving an object if the image contains disparity information [14, 25].

We used an energy model (provided by our collaborators from EyeShots) whose output consists of edge and disparity cell responses comparable to V1. We have shown that given such information our model is capable of learning, discriminating and recognizing different objects. Each object is learnt and represented by a distribution of activated cells (Figure 11), the objects are all well discriminable (Figure 12) and put into a different environment (Figure 13) the objects can be recognized as good as the objects without background (Figure 14).

The different tuning characteristics of the cells (Figure 17) shows the complexity of the learnt cells. Type 1 cells are similar to the input simple cells but have a spatial (z-axis) invariance and so they pool over specific absolute disparity to the fixation position, while Type 2 cells show object recognition completely invariant of object position or fixation. Type 2 cells might indicate that the model accomplishes object recognition with the help of the depth information of the object. The performance in the noise experiments (where edge information is less usable for object recognition) strengthens the assumption that features like depth of the object might be used for the successful discrimination of the objects. Type 3 cells indicate that the cells are invariant to fixation and object position but also show a small preference for a specific absolute distance to the fixation point. The degree of the influence of object depth in the object recognition will be focus of future research.

To summarize, we have developed a model that learns with self-organized cells and Hebbian learning from simple edge and disparity cell responses an object recognition representation of intermediate complexity that shows tuning characteristics in absolute disparity.

References

[1] A. Anzai, I. Ohzawa, and R. D. Freeman. Neural mechanisms underlying binocular fusion and stereopsis: position vs. phase. *Proc Natl Acad Sci U S A*, 94(10):5438–5443, May 1997.

[2] A. Anzai, I. Ohzawa, and R. D. Freeman. Neural mechanisms for encoding binocular disparity: receptive field position versus phase. *J Neurophysiol*, 82(2):874–890, Aug 1999.

[3] H. B. Barlow, C. Blakemore, and J. D. Pettigrew. The neural mechanism of binocular depth discrimination. *J Physiol*, 193(2):327–342, Nov 1967.

[4] C. Blakemore, A. Fiorentini, and L. Maffei. A second neural mechanism of binocular depth discrimination. *J Physiol*, 226(3):725–749, Nov 1972.

[5] H. Bridge and B. G. Cumming. Responses of macaque v1 neurons to binocular orientation differences. *J Neurosci*, 21(18):7293–7302, Sep 2001.

[6] Nickolay Chumerin. Nikolay Chumerin's myRaytracer. Online Resource: http://sites.google.com/site/chumerin/projects/myraytracer, 2009.

[7] G. C. DeAngelis, I. Ohzawa, and R. D. Freeman. Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. i. general characteristics and postnatal development. *J Neurophysiol*, 69(4):1091–1117, Apr 1993.

[8] D. Ferster. A comparison of binocular depth mechanisms in areas 17 and 18 of the cat visual cortex. *J Physiol*, 311:623–655, Feb 1981.

[9] P. Földiàk. Forming sparse representations by local anti-hebbian learning. *Biol Cybern*, 64:165–170, 1990.

[10] D. Gabor. Theory of communication. *J. Inst. Elect. Eng.*, 93:429–457, 1946.

[11] D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol*, 160:106–154, 1962.

[12] D. H. Hubel and T. N. Wiesel. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *J Neurophysiol*, 28:229–89, March 1965.

[13] D. E. Joshua and P. O. Bishop. Binocular single vision and depth discrimination. receptive field disparities for central and peripheral vision and binocular interaction on peripheral single units in cat striate cortex. *Exp Brain Res*, 10(4):389–416, 1970.

[14] B. Julesz. *Foundations of cyclopean perception.* University of Chicago Press, Chicago, 1971.

[15] S. LeVay and T. Voigt. Ocular dominance and disparity coding in cat visual cortex. *Vis Neurosci*, 1(4):395–414, 1988.

[16] D. Marr and E. Hildreth. Theory of edge detection. *Proc R Soc Lond B Biol Sci*, 207(1167):187–217, Feb 1980.

[17] J. I. Nelson, H. Kato, and P. O. Bishop. Discrimination of orientation and position disparities by binocularly activated neurons in cat straite cortex. *J Neurophysiol*, 40(2):260–283, Mar 1977.

[18] T. Nikara, P. O. Bishop, and J. D. Pettigrew. Analysis of retinal correspondence by studying receptive fields of binocular single units in cat striate cortex. *Exp Brain Res*, 6(4):353–372, 1968.

[19] I. Ohzawa, G. C. DeAngelis, and R. D.Freeman. Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science*, 249(4972):1037–41, August 1990.

[20] I. Ohzawa, G. C. DeAngelis, and R. D. Freeman. Encoding of binocular disparity by complex cells in the cat's visual cortex. *J Neurophysiol*, 77(6):2879–2909, Jun 1997.

[21] E. Oja. A simplified neuron model as a principal component analyzer. *J Math Biol*, 15(3):267–273, 1982.

[22] G. E. Poggio. Mechanisms of stereopsis in monkey visual cortex. *Cereb Cortex*, 5(3):193–204, 1995.

[23] G. F. Poggio and B. Fischer. Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. *J Neurophysiol*, 40(6):1392–1405, Nov 1977.

[24] G. F. Poggio, F. Gonzalez, and F. Krause. Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity. *J Neurosci*, 8(12):4531–4550, Dec 1988.

[25] G. F. Poggio and T. Poggio. The analysis of stereopsis. *Annu Rev Neurosci*, 7:379–412, 1984.

[26] G. F. Poggio and W. H. Talbot. Mechanisms of static and dynamic stereopsis in foveal cortex of the rhesus monkey. *J Physiol*, 315:469–492, Jun 1981.

[27] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C (2nd ed.)*. New York: Cambridge, 1992.

[28] S. J D Prince, B. G. Cumming, and A. J. Parker. Range and mechanism of encoding of horizontal disparity in macaque v1. *J Neurophysiol*, 87(1):209–221, Jan 2002.

[29] N. Qian. Computing stereo disparity and motion with known binocular cell properties. *Neural Computation*, 6(3):390–404, 1994.

[30] Fangtu T Qiu and Rüdiger von der Heydt. Figure and ground in the visual cortex: v2 combines stereoscopic cues with gestalt rules. *Neuron*, 47(1):155–66, July 2005.

[31] J C A Read and B G Cumming. Sensors for impossible stimuli may solve the stereo correspondence problem. *Nat Neurosci*, 10(10):1322–8, October 2007.

[32] M. Riesenhuber and T. Poggio. Are cortical models really bound by the "binding problem"? *Neuron*, 24(1):87–93, 111–25, September 1999.

[33] M. Riesenhuber and T. Poggio. Hierarchical models of object recognition in cortex. *Nat Neurosci*, 2(11):1019–25, November 1999.

[34] Sabatini S.P., Gastaldi G., Solari F., Diaz J., Ros E., Pauwels K., Van Hulle M.M., Pugeault N., and Krüger N. Compact and accurate early vision processing in the harmonic space. In *International Conference on Computer Vision Theory and Applications (VISAPP), Barcelona*, 2007.

[35] R. L. De Valois, D. G. Albrecht, and L. G. Thorell. Spatial frequency selectivity of cells in macaque visual cortex. *Vision Res*, 22(5):545–559, 1982.

[36] R. von der Heydt, C. Adorjani, P. Hänny, and G. Baumgartner. Disparity sensitivity and receptive field incongruity of units in the cat striate cortex. *Exp Brain Res*, 31(4):523–545, Apr 1978.

[37] R. von der Heydt, H Zhou, and H S Friedman. Representation of stereoscopic edges in monkey visual cortex. *Vision Res*, 40(15):1955–67, 2000.

[38] Jan Wiltschut and Fred H Hamker. Efficient coding correlates with spatial frequency tuning in a model of v1 receptive field organization. *Visual Neuroscience*, 26:21–34, 2009.