



**Project no.:** FP7-ICT-217077  
**Project full title:** Heterogeneous 3-D Perception across Visual Fragments  
**Project Acronym:** EYESHOTS  
**Deliverable no:** D3.1a  
**Title of the deliverable:** Demonstration of learning disparity-tuned feature selective cells

<b>Date of Delivery:</b>	03 March 2009	
<b>Organization name of lead contractor for this deliverable:</b>	WWU	
<b>Author(s):</b>	F.H. Hamker and M.A. Voss	
<b>Participant(s):</b>	WWU	
<b>Workpackage contributing to the deliverable:</b>	WP3	
<b>Nature:</b>	Other (software module)	
<b>Version:</b>	1.0	
<b>Total number of pages:</b>	9	
<b>Responsible person:</b>	Fred H. Hamker	
<b>Revised by:</b>	Marc Van Hulle	
<b>Start date of project:</b>	1 March 2008	<b>Duration:</b> 36 months

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

**Abstract:**

This deliverable provides information and first obtained results of the developed software module called PVCStereo that allows for learning binocular receptive fields based on Hebbian learning. The software module is written in C++, using an artificial neural-network architecture (ANNarchy) library, developed in our lab. The learning algorithm has been demonstrated to learn from stereo images of natural scenes. The set of stereo images was taken from a two-camera system mounted in our lab. We tested if the model learns disparity-tuned and feature selective cells from this set of stereo images. As expected, the cells learn localized, oriented, disparity tuned and bandpass filtering receptive fields, comparable to those in area V1 of the primate brain. The detailed comparison with physiological data and the analysis of the determination of the disparity tuning curves is ongoing and will be reported in D 3.1b.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Approach and methodology</b>	<b>2</b>
<b>3</b>	<b>Brief description of results</b>	<b>4</b>
<b>4</b>	<b>Concluding remarks</b>	<b>4</b>
<b>5</b>	<b>Appendix: PVCStereo Documentation</b>	<b>7</b>
5.1	Requirements . . . . .	7
5.2	Classes . . . . .	7
5.3	File List . . . . .	8

## 1 Introduction

The EU project "Eyeshots" aims to establish a visuo-motor system grounded in the concept of "active fragmented vision". According to this concept, perception generates dynamic cognitive interpretations of the scene, which does not imply a real metrical 3-D reconstruction of the observed space, but instead a loose representation of objects that are actively bound on time for the task at hand (in terms of affordance, salience, and planning of actions). We here describe the research background of the software developed for learning the receptive fields of neurons in the visual pathway. Our algorithm is motivated by research to capture the basic principles of primate 3-D perception. Inputs from left and right eyes are combined in primary visual cortex (V1), where many cells are tuned for binocular disparity. Complex cells in V1 often show a tuning for a preferred disparity. However, V1 is not the source of stereoscopic depth perception. V1 rather provides local estimates of absolute disparity.

So far, models of primary visual cortex encoding disparity have primarily be constructed by hand based on data, but little work has been done on developing learning algorithms that lead to general purpose receptive fields similar as observed in the brain. The deliverable of work package 3.1 provides an alternative branch to work package 2.1 where specialized filters for vergence control are developed. The output of the model from work package 3.1 will therefore be used in WP2, where the pros and cons of general purpose compared to specialized disparity detectors are investigated, and within WP3 to learn more complex feature detectors by combining V1 detectors.

## 2 Approach and methodology

Since the early studies of receptive field properties in primary visual cortex (Hubel & Wiesel, 1962; Valois et al., 1982; DeAngelis et al., 1993), a major issue in neural coding has emerged, dealing with the question of why neurons have a particular receptive field structure. Since V1 neurons respond well to edges, edge detection has been considered as a useful operation of early vision emphasizing the important structural properties of a visual scene (Marr & Hildreth, 1980). However, this does not answer the questions about optimal edge detectors and particularly why edge detectors should emerge and not any other potentially useful detector. Important progress has arisen from the efficient coding hypothesis (Attneave, 1954; Barlow, 1961; Laughlin, 1981; Atick & Redlich, 1990; Hateren, 1993; Field, 1994). Particularly, recent contributions in this respect have shown that algorithms seeking for a statistical independence of the neural responses converge to localized, oriented, band-pass filters (Olshausen & Field, 1996; Bell & Sejnowski, 1997; van Hateren & van der Schaaf, 1998). However, despite this great success, a more close comparison with neural data revealed that the learned receptive fields do not capture the full frequency distribution as observed in experimental data (van Hateren & van der Schaaf, 1998; Ringach et al., 2002).

While many studies have demonstrated some relationship between neural receptive field properties and aspects of efficient coding, we investigated the quantitative influence of particular aspects of efficient coding on the similarity between the modeled receptive fields and the experimental data (Wiltschut & Hamker, in press). In this recent, previous work, we systematically varied critical model parameters and measured information theoretic properties of efficient coding in these different instances after learning. We then analyzed if these measurements of efficient coding correlate with the similarity between model and biological data – the distribution of spatial frequency tuning (Ringach et al., 2002). We observed in most but not all cases that making the code more efficient enhances the similarity between model and experimental data. However, with respect to the coding quality (e.g. reconstruction error) we observe a saturation, enforcing a highly independent and sparse code does not further improve or even diminish coding quality.

Our developed model (Figure 1) consists of two layers, of which neurons are bidirectionally connected with each other by feedforward ( $W$ ) and feedback ( $A$ ) weights. The activity  $r_i^{On/Off}$  is obtained from images that have been whitened/lowpass filtered and separated into on/off channels (depending on the sign of the pixel value after filtering).

Layer II (model V1) gets activated from Layer I neurons, but is dependent on the activity of other Layer II cells through lateral inhibition developed by anti-Hebbian learning. The Layer II cells feed back to Layer I cells and increase their gain. Due to the learning of the feedback weights, this feedback is predictive. Unlike generative models where the difference between feedforward and feedback is computed, the feedback signal enhances the sensitivity of specific neurons in the previous layer and

thus leads to an attentional tuning.

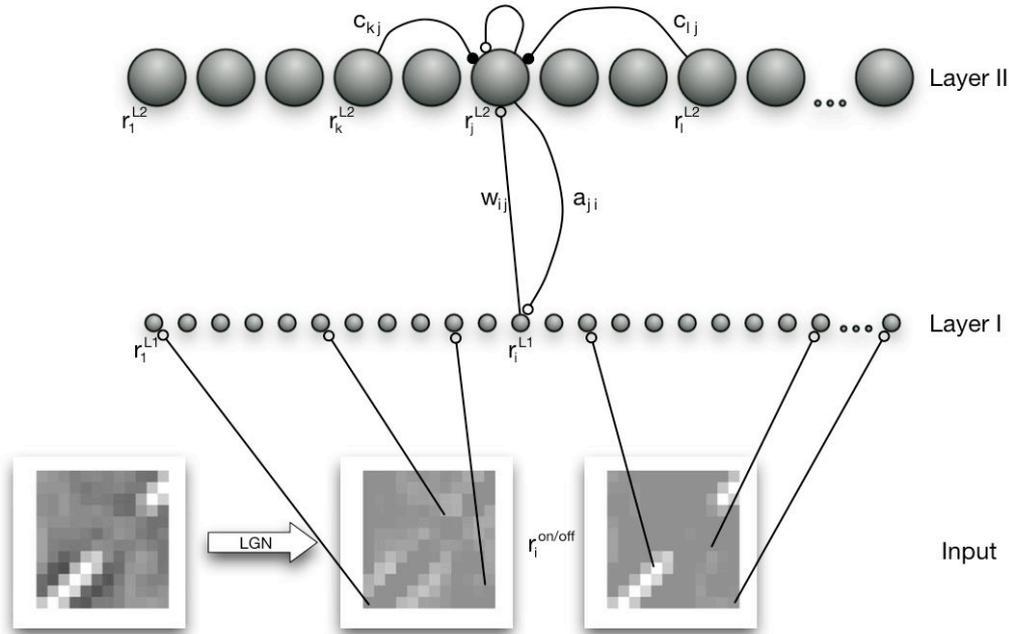


Figure 1: The network consists of two layers. The first layer (Layer I) represents the simulated input modulated by the 'attentional' feedback signal ( $a$ ). The cells of Layer II represent the simulated V1 cells. Each cell of Layer I gains feedback from all cells of Layer II (according to the feedback weight connection) and each Layer II cell obtains its excitation from all Layer I cells (dependent on the feedforward weights ( $w$ ) respectively their receptive field). This excitatory input is linear but this signal is further processed non-linearly. Each cell uses its current activation state to self-enhance its firing rate but is also inhibited by all other cells, dependent on the current lateral inhibition connection weights ( $c$ ).

The learning of the connections between neurons is implemented via a Hebbian principle. Long term potentiation (LTP) requires an above-mean activation of both, pre- and post-synaptic activities, which is well known as the covariance learning rule (Sejnowski, 1977; Willshaw & Dayan, 1990). Long term depression (LTD) occurs by the constraint to limit the overall weight resource and, only for the feedforward connections, if the pre-synaptic activity is below the population mean. For details see (Wiltchut & Hamker, in press).

We have shown that this model with non-linear lateral competition, learned by an anti-Hebbian principle, and Hebbian learning of feedforward and feedback connections, develops receptive fields which are not only similar to V1 macaque data, but which match the distribution of the spatial frequency tuning, particularly if the neural code is efficient (Wiltchut & Hamker, in press). We also compared our results

to Independent Component Analysis (ICA), a standard linear method for learning receptive fields from natural scenes. We used the fast fix-point algorithm (Hyvärinen et al., 2001). Hoyer & Hyvarinen (2000) have shown that this algorithm can produce orientated bandpass filters and disparity tuned neurons similar to those in V1. This particular ICA algorithm shows similar deficits than the results from Lewicki et al. (1999) and more mild deficits compared to those Ringach et al. (2002) has reported for Bell & Sejnowski (1997) ICA and the Olshausen & Field (1996) sparse coding algorithm with respect to the variety of receptive field properties. In all of these studies ICA only captures in part the whole distribution of receptive field properties.

The software we have developed extends the network describe above to learn binocular receptive fields from a set of stereo images. The set of stereo images was taken from a two-camera system mounted in our lab. For the initial testing of the algorithm we used a small set of images from indoor scenes.

### 3 Brief description of results

From each image we randomly selected patches of  $12 \times 12$  pixels. As the on/off channels consist of  $12 \times 12$ , 288 cells were required in the first layer for each eye, i.e., 144 neurons receive input from the on-, the others from the off-cells. We used 288 cells in Layer II to represent the input combinations. The feedforward weights  $w_{ij}$  were initialized randomly with a mean  $\bar{w} = 0.1$ . The feedback and the lateral inhibition weights were initialized to zero. An image patch is presented for 50ms to let the dynamics of the system converge to a stable state. After each trial the feedback and feedforward synapses as well as the lateral inhibition connections are updated according to the final firing rates of the cells.

Figure 2 shows typical examples of tuning curves of binocular cells in the new model using indoor scenes for learning. We observe a large number of cells with disparity tuning. The first cell is tuned to near objects, the other ones are tuned to far objects. However, the second cell shows quite broad tuning characteristics.

### 4 Concluding remarks

Our model has been demonstrated to develop typical V1 receptive fields using the statistics of natural scenes for unsupervised receptive field organization based on Hebbian learning as shown in detail in Wiltschut & Hamker (in press).

The software developed for this deliverable further expands the model to also show disparity tuning similar as it has been observed in area V1. An in depth comparison to V1 data with respect to disparity tuning will be reported in deliverable 3.1b using nat-

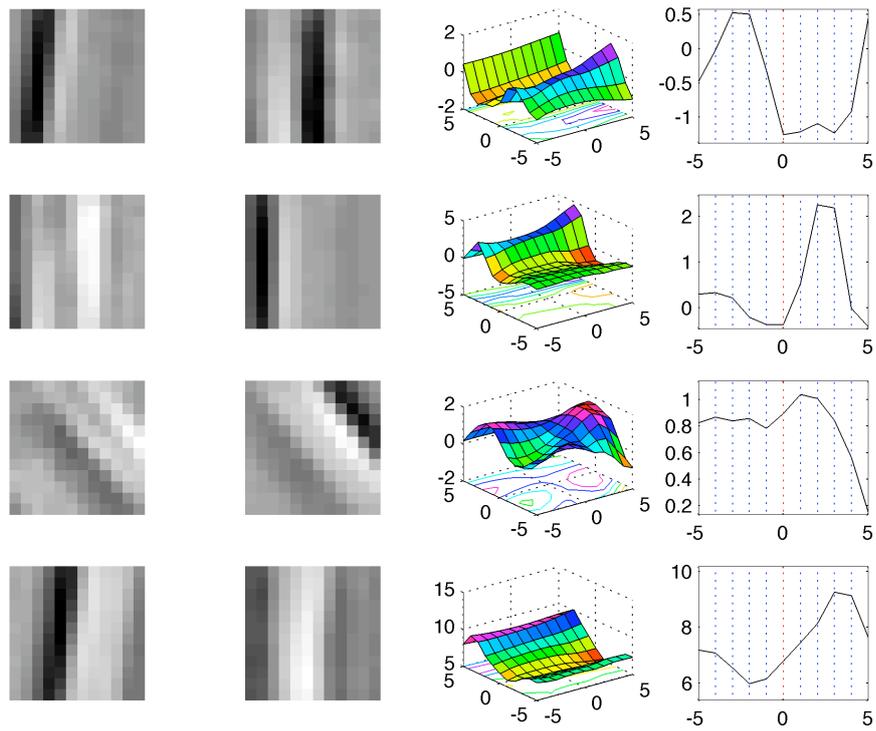


Figure 2: Each row shows the tuning properties of a binocular cell after learning. The two images on the left show the receptive field of the left and right view. The third image shows the horizontal and vertical disparity tuning (in pixels) of the cell and the last image depicts the horizontal disparity tuning properties at zero vertical disparity.

ural outdoor scenes. The feedback connections learned will be particularly important for WP 3.2 where they will be used to attentively bind visual fragments on demand.

## References

- Atick, J. J. & Redlich, A. N. (1990). Towards a theory of early visual processing. *Neural Computation*, 2, 308–320.
- Attneave, F. (1954). Some informational aspects of visual perception. *Psychol Rev*, 61, 183–193.
- Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. In Rosenblith, W. A., Ed., *Sensory Communication*, pages 217–234, Cambridge, MA. MIT Press.
- Bell, A. J. & Sejnowski, T. J. (1997). The “independent components” of natural scenes are edge filters. *Vis. Res.*, 37(23), 3327–3338.
- DeAngelis, G. C.; Ohzawa, I., & Freeman, R. D. (1993). 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.
- Field, D. J. (1994). What is the goal of sensory coding? *Neural Comput*, 6, 559–601.
- Hateren, J. H. V. (1993). Spatiotemporal contrast sensitivity of early vision. *Vision Res*, 33(2), 257–267.
- Hoyer, P. O. & Hyvarinen, A. (2000). Independent component analysis applied to feature extraction from colour and stereo images. *Network*, 11(3), 191–210.
- Hubel, D. H. & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *J Physiol*, 160, 106–154.
- Hyvärinen, A.; Karhunen, J., & Oja, E. (2001). *Independent Component Analysis*. Wiley, New York.
- Laughlin, S. B. (1981). Simple coding procedure enhances a neuron’s information capacity. *Z. Naturforsch.*, 36C, 910–912.
- Lewicki, M. S.; Hughes, H., & Olshausen, B. A. (1999). Probabilistic framework for the adaptation and comparison of image codes. *Comparison of Image Codes, Ó J. Opt. Soc. Am. A*, 16, 1587–1601.
- Marr, D. & Hildreth, E. (1980). Theory of edge detection. *Proc R Soc Lond B Biol Sci*, 207(1167), 187–217.
- Olshausen, B. A. & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381, 607–609.
- Ringach, D. L.; Bredfeldt, C. E.; Shapley, R. M., & Hawken, M. J. (2002). Suppression of neural responses to nonoptimal stimuli correlates with tuning selectivity in macaque V1. *J. Neurophysiol.*, 87(2), 1018–1027. (doi:10.1152/jn.00614.2001.).

- Sejnowski, T. J. (1977). Storing covariance with nonlinearly interacting neurons. *J. Math. Biol.*, 4(4), 303–321. (doi:10.1007/BF00275079).
- Valois, R. L. D.; Yund, E. W., & Hepler, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Res*, 22(5), 531–544.
- van Hateren, J. H. & van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. Biol. Sci.*, 265(1394), 359–366. (doi:10.1098/rspb.1998.0303).
- Willshaw, D. & Dayan, P. (1990). Optimal plasticity from matrix memories: what goes up must come down. *Neural Comp.*, 2, 85–93.
- Wiltshut, J. & Hamker, F. (in press). Efficient coding correlates with spatial frequency tuning in a model of v1 receptive field organization. *Vis Neurosci*.

## 5 Appendix: PVCStereo Documentation

### 5.1 Requirements

PVCStereo is build using the Artificial Neural Network architecture ANNarchy written by Julien Vitay, `jvita_01@uni-muenster.de` in the Lab of Fred H. Hamker at the Westf. Wilhelms-University Münster. The ANNarchy library must be installed prior to the compilation of the PVCStereo sources. For the installation of the ANNarchy library please consult the documentation of ANNarchy. In the following, a brief overview about the classes is provided. A full description of all classes as well as a description for compilation and running will be provided to the project partners upon request.

### 5.2 Classes

Here are the classes, structs, unions and interfaces that have been developed

- class LgnNeuron  
*Class representing a LGN neuron.*
- class LgnLayer  
*Class representing a LGN Layer.*
- class LgnMap

*Class representing a LGN map.*

- class V1NeuronIn  
*Class representing a V1-In neuron.*
- class V1LayerIn  
*Class representing a V1 In Layer.*
- class V1Map  
*Class representing a V1 map.*

### **5.3 File List**

Here is the list of files developed.

- ConnectionManager.cpp
- ConnectionManager.h
- Controls.h
- Main.cpp
- MainGUI.h
- Network.cpp
- Network.h
- PVCStereo.cpp
- PVCStereo.h
- World.cpp
- World.h