

# Bio-inspired disparity estimation system from energy neurons

Flavio Mutti  
Dipartimento di Elettronica e Informazione  
Politecnico di Milano  
Milan, Italy  
mutti@elet.polimi.it

Giuseppina Gini  
Dipartimento di Elettronica e Informazione  
Politecnico di Milano  
Milan, Italy  
gini@elet.polimi.it

**Abstract**—Bio-inspired control systems attempt to reproduce the intelligent behavior by simulating the internal architecture and mechanisms of the biological counterpart. In this paper we propose a bio-inspired algorithm for disparity estimation based on the disparity energy model. In literature several models have been proposed but actually each of these models seems to have unique features and unique lacks due to the intrinsic architecture. Different bio-inspired architectures are reviewed and from the comparison among these algorithms we propose a new architecture that takes into account the good properties of the previous models and tries to overcome the limitations. Simulations are performed on real images, and comparison with published algorithms on Middlebury stereo database are shown. We comment the obtained results from the point of view of humanoid robotics and show how the study of neurophysiology can inspire the design of vision systems.

**Keywords:** disparity map, bio-inspired vision system, neural model, depth perception, primary visual cortex, humanoid robotics

## I. INTRODUCTION

The human primary visual cortex estimates the environment depth starting from a pair of stereo images. The visual signal is captured by the photoreceptors of the retina; after a brightness pre-processing, the signal is sent through the Lateral Geniculate Nucleus (LGN) to the primary visual cortex (V1). From photoreceptors to LGN the visual signal is strictly monocular, so there is no way to estimate the depth of the environment using the stereo image cues (e.g. the retinal disparity). It is known that the depth perception depends primarily on information about retinal disparity and not on other information cues coming from high-level decision areas (like prefrontal cortex) [9]. So, it exists an "automatic process" to estimate disparity that does not imply reasoning. Therefore, if it was possible to simulate the mechanisms that underlie the depth perception, we could design a system to estimate retinal disparity in a bio-inspired way.

From the humanoid robotics point of view, that involves different disciplines from biology to engineering, the ultimate goal is to build a humanoid that can interact with humans "like" a human. One of the main issues is to design a *controller* for humanoid robots is to deal with the vastness of the information available in the surrounding environment; it is not feasible to simply copy a biological system "as is" (and this fact is true in general, even for biological systems much simpler than humans). Rather, the goal is to

discover principles that underly the biological control and try to transfer those to humanoid robotics [1].

Besides, the field of bionics seeks to design robots that mimics biological structures and recent works show that a successful design rely on embodiment [1][2]. It follows that the design of the controller (the central nervous system) is inseparable by the morphology of the robot because both affect the efficiency of the robot [1] [3]. So, if we intend to design a complete controller for a humanoid antropomorphic robot that must interact with humans in hostile environments, and with potentially infinite situations, it could be suitable to try to exploit the intrinsic structure of the brain for information processing.

The idea is to develop a vision system that can be easily integrated in a more complex architecture which must take into account also the structure and the embodiment of the humanoid.

In the following sections we will introduce recent literature on neural approaches to stereo vision based on neuroscience evidence, and fully describe our solution. In the last part we will present our results and compare them to other literature results.

## II. BACKGROUND

Stereo vision generally addresses the research field that studies how to elicit some interesting features from two images; one of the main features of interest is the perception of depth. Then, generally speaking, we are searching for algorithms or methods that allow to perform the perception of depth with adequate reliability.

In this section we introduce the main algorithms and methods published in recent literature.

The first model that explains the intrinsic function of neurons in the primary visual cortex is the *disparity energy model* [7] [8]. This model proposes the existence of two different types of neurons, called respectively simple and complex cells, and explains how they communicate in order to maximally respond in presence of the disparity for what they are tuned. It also shows that the receptive field of binocular simple cells can be approximated as Gabor filters, with proper parametrization (see equation 1). The Gabor filter is typically used in signal processing and it has several good properties (for a detailed description see [6]). However, this model does not explain how it is possible to produce a reliable disparity map.

A first systematic study to explain some properties of the disparity energy model, starting from its mathematical definition, is developed in [10] [11]. They introduce topics as spatial pooling, scale pooling, and orientation pooling; each simulation presented there is done with random dot stereograms synthetic images (RDS) [9]. The usage of synthetic images is reasonable in order to explain properties but in general it is not meaningful to evaluate the performance of the system with them, because the synthetic images have a simplified spatial structure and luminance intensity with respect to natural scenes, and in our experience the result of experiments with RDS are misleading.

In [14] the firing rate of complex neurons is taken as an evidence of the disparity estimation. So, given different complex neurons with different preferred disparities, the estimated disparity is equal to the preferred disparity of the most responsive neuron of the population.

A template-matching approach is proposed in [15]; it is based on a scalar measure of the mismatch between the neural responses and the templates of responses, given a specific disparity.

In [16] a model that successfully integrates spatial pooling, orientation pooling, and scale pooling is proposed; they present a coarse-to-fine mechanism with phase and position-shift integration. Due to the neural architecture, the system is robust to the image acquisition noise but the disadvantage is that the system was designed to work with small disparities and typically in real scenes the disparity range is quite wide.

Another approach to design a bio-inspired vision system is discussed in [17] where the cooperation between phase and position-shift mechanisms is interpreted in a way quite different respect to [16]. Both methods are based on disparity energy model but in [17] the system is intrinsically mono-scale and mono-orientation and it provides a mechanism in order to evaluate a large range of disparities. Besides, the authors suggest to use a normalized feature in order to assess whether the estimation is reliable or not and if an image point belongs to occluded regions. Their results, compared to [16], seem to denote a better ability to estimate the disparity map; however this architecture does not include the orientation and scale pooling, that should improve the disparity estimation. In [18] the previous model is extended with orientation pooling in order to accumulate "evidence" to support a disparity hypothesis. This model is mainly based on a Bayes filter and it uses a Bayes factor to test the hypothesis with the maximum support. Moreover, the proposed model identifies the occluded pixels. In [18] they publish the results of the model tested on Middlebury stereo images [4] [5].

Another approach to disparity map estimation, proposed in [19], is essentially a coarse-to-fine algorithm, with orientation and spatial pooling. To obtain a robust disparity estimation a weighted sum of the complex responses for each orientation is computed. Then, they define a vector disparity as the vector difference from corresponding points in the left and the right images, that permits to evaluate disparities that are not only horizontal. In fact the model can estimate even disparities that have also a vertical component (it is necessary

to compute this component when the principal axes of the stereo cameras are not parallel).

Our model combines the technique proposed in [18] for the computation of large disparities and the capability of the neural architecture of [19] to estimate the vertical component of the disparity. Moreover, in order to improve the robustness, we introduce a weighted coarse-to-fine mechanism in a way similar to [16].

### III. THE NEURAL ALGORITHM

The primary visual cortex (or *area VI*) is the first area that integrates information afferent from both the eyes to produce a three-dimensional representation of the environment based on two-dimensional retinal images; this process is also called *binocular fusion*.

The perception of the environment depth is closely related to the estimation of retinal disparity; retinal images are not strictly equal because of the physical distance between the two eyes. Computing the disparity between the two retinal images allows to estimate the environment depth, relative to the *fixation plan* determined by the eyes convergence.

For our purpose, the problem of depth perception can be reduced to the computation of retinal disparity.

#### A. Image preprocessing

Generally the image sizes are not known a priori, so we need to develop a system able to deal with stereo pairs of different dimension. Obviously, this approach involves to take care about some system parameters, as we will later explain. From now on we consider only aspects that are independent of the size of the images. The acquired images are, in general, color images so it is necessary to convert them to luminance data in order to preserve only the intensity component. After this, the mean luminance is subtracted from the pair of the two images in order to improve the edge enhancement, like the human retina [8].

#### B. Disparity energy neurons

The disparity energy model explains the response properties of the binocular neurons in VI [7]. This model uses two types of neurons, the simple and complex cells, that are tuned to specific disparities [8][10]. The left and the right images coming from the preprocessing stage are then filtered with Gabor filters of different orientation, scale and shape according to the disparity energy model and the coarse-to-fine technique with both the phase and position shift mechanisms [16]. Let  $r_s$  and  $r_q$  be the simple and the complex response of the simple and the complex neurons respectively, and let  $g(x, y, \theta, \phi, \Delta\phi, \omega)$  be the Gabor filter. Then,

$$g(x, y, \theta, \phi, \Delta\phi, \omega) = s(x, y, \theta, \phi, \Delta\phi, \omega)w(x, y, \theta) \quad (1)$$

where  $s(x, y, \theta, \phi, \Delta\phi, \omega)$  is a cosinusoid and  $w(x, y, \theta)$  is a 2D Gaussian-shaped function (known as **envelope**). The cosinusoid is defined as follows,

$$s(x, y, \theta, \phi, \Delta\phi, \omega) = \cos(\omega[x \cos \theta + y \sin \theta] + \phi + \Delta\phi) \quad (2)$$

where  $\omega$  is the preferred spatial frequency,  $\theta$  is the filter orientation,  $\phi$  is the phase parameter that will be used in the complex response mechanism (to define a quadrature pair of simple responses) and  $\Delta\phi$  is the phase difference between a pair of receptive fields (RFs). The envelope is defined as follows,

$$w(x, y, \theta) = k \exp\left(-\frac{[x \cos \theta + y \sin \theta]^2}{2\sigma_x^2} - \frac{[x \sin \theta + y \cos \theta]^2}{2\sigma_y^2}\right) \quad (3)$$

where  $\sigma_x$  and  $\sigma_y$  define the envelope dimensions (and the RF extension) and  $k$ , involved in the filter gain, is defined as

$$k = \frac{1}{2\pi\sigma_x\sigma_y} \quad (4)$$

Therefore the receptive fields (RFs), based on biological evidences [7][8], can be modelled as,

$$g_l(x, y) = g(x, y, \theta, \phi, \frac{\Delta\phi}{2}, \omega) \quad (5)$$

$$g_r(x, y) = g(x - d, y, \theta, \phi, -\frac{\Delta\phi}{2}, \omega) \quad (6)$$

where  $l, r$  subscripts represent the left and the right RF, respectively;  $d$  is the position-shift parameter,  $\Delta\phi$  is the phase-shift parameter. Then the simple cell response is written as,

$$r_s = \left\{ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [g_l(x, y)I_l(x, y) + g_r(x, y)I_r(x, y)] dx dy \right\}^2 \quad (7)$$

where  $I_l$  and  $I_r$  are the input images coming from the preprocessing stage. Finally, the complex response cell is defined as,

$$r_q = r_{s,1} + r_{s,2} \quad (8)$$

where  $r_{s,1}$  and  $r_{s,2}$  are simple responses in quadrature phase, i.e.  $\phi_1 = 0$ ,  $\phi_2 = \frac{\pi}{2}$  and  $\Delta\phi_1 = \Delta\phi_2$ . The preferred disparity of the complex cell response is given by,

$$D_{pref} = \frac{\Delta\phi}{\omega \sin \theta} + d \quad (9)$$

which means that the complex cell will response maximally when the RFs of the complex neuron contain the preferred disparity.

### C. Neural Architecture

In this section we explain the neural architecture of our system (see Figure 1). Motivated by the previous section we integrate some interesting features of other proposed models. For each pixel in the left image we want to estimate the corresponding pixel in the right image to produce the disparity map. The spatial frequency is  $\omega = \frac{\pi}{2}$  and  $\sigma_x = 2$ . The position shift across the population is  $\Delta C = \{0, 1, \dots, 55\}$ . The aspect ratio is 2.

1) *Spatial pooling*: To improve the response of complex cell it is possible to take into account the physiological fact that the RF size of the biological complex cell is larger than that of the complex cell model [10]. Moreover, a given complex cell response is improved by the responses of the complex cells of the near neurons. This observation is included into our model by averaging several pairs of complex cells with overlapping RFs. Spatial pooling can be mathematically defined as,

$$r_c(x_0, y_0) = \frac{1}{(a+1)^2} \sum_{i=x_0-\frac{a}{2}}^{x_0+\frac{a}{2}} \sum_{j=y_0-\frac{a}{2}}^{y_0+\frac{a}{2}} r_q(i, j) w(i, j) \quad (10)$$

where  $r_q$  is the complex cell response at the  $(i, j)$  stereo images location (see equation 8),  $w(i, j)$  is a spatial weighting function and  $r_c$  is the spatial pooling response of the complex cell with RFs centered at  $(x_0, y_0)$  over stereo images. In our system the chosen weighting function is a symmetric two-dimensional gaussian with  $\sigma_{pooling} = 2\sigma_x$ .

2) *Normalized response*: In [18] they propose to evaluate the population responses at different locations in order to estimate the position-shift component and to refine the estimation with the chosen population via phase-shift mechanism. To evaluate the position-shift component they suggest to use a normalized feature  $R$  defined as,

$$R = \frac{P - M}{M} \quad (11)$$

where  $P$  and  $M$  are respectively the peak and the mean of the population response curve. It can be demonstrated that the feature  $R$  takes values between 0 and 1, since  $M \geq (P - M)$ . In order to choose the most probable position shift, they choose the population response curve (each of them is localized at different horizontal locations) that maximizes the feature  $R_{\Delta c}$ , where  $\Delta c$  denotes the current position disparity. Due to the good properties of this normalized feature we use it in order to estimate the disparity.

3) *Orientation pooling*: According to [18], we implement a similar orientation pooling mechanism.

$$\widehat{R} = \sum_{\theta} w_{\theta} R_{\Delta c, \theta} \quad (12)$$

where  $\widehat{R}$  formally depends from the position shift  $\Delta C$  and the scale,  $R_{\Delta c, \theta}$  is the normalized response at equation 11 and the weights  $w_{\theta}$ , that are estimated through an exhaustive search in the space problem using a set of Middlebury stereo images.

According to previous results [16] we use 5 orientations ranging from  $-60^\circ$  to  $60^\circ$  in  $30^\circ$  steps.

4) *Scale pooling*: Differently from [18] we propose to introduce a scale pooling phase in order to make robust the estimated disparity. However, despite [16] and [19] we introduce the scale pooling as weighted average across scales. Formally,

$$\bar{R} = \sum_s w_s \widehat{R} \quad (13)$$

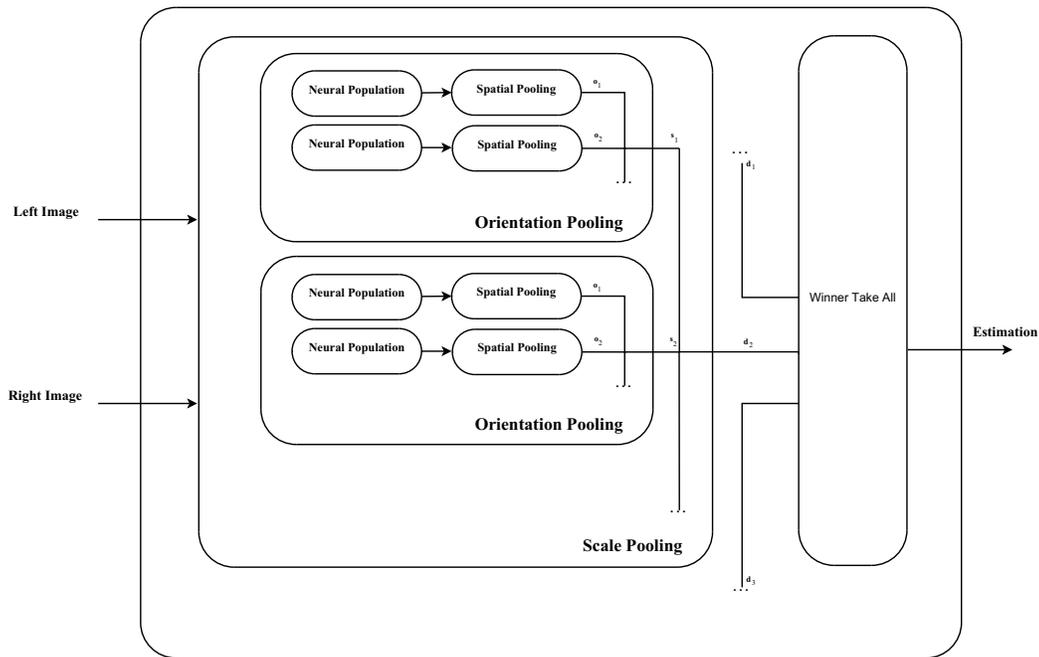


Fig. 1: Neural architecture

where  $w_s$  are the weights that are estimated through an exhaustive search. Due some empirical results, we choose to use only two scales because the overall performance seems to increase for two scales only.

Now, taking the maximum  $\bar{R}$  at each position shift location, we estimate the most probable disparity.

#### D. Disparity direction

For each pixel it is possible to determine the direction of the estimated disparity. We take the preferred direction of the filter (i.e. the normal direction respect to the principal axis of the filter) that is the direction of the the bi-dimensional preferred disparity associated to the corresponding complex neuron.

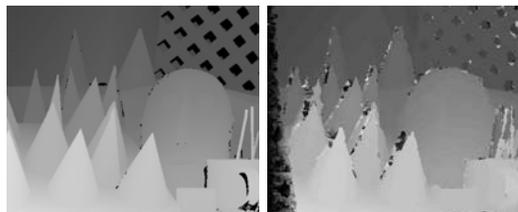
In [19] the authors propose a weighted sum of the complex cell responses for each orientation (i.e. center of gravity). Here, we propose a weighted average of the estimated disparity for each orientation with optimized weights (estimated with an exhaustive search through the orientations). Formally,

$$V(x, y) = \sum_i w_{\theta_i} d_{\theta_i} \quad (14)$$

where  $w_{\theta_i}$  is the estimated weight that depends from the orientation  $\theta$  (again, it is estimated through an exhaustive search) and  $d_{\theta_i}$  is the vector with module equal to the estimated disparity at the given orientation  $\theta$ . The resultant vector  $V(x, y)$  will have the direction of the estimated disparity (in our case we expect to have always horizontal vectors because our test images have only horizontal disparities). Our simulations show the effectiveness of the proposed formula, see Table I; for each pixel in the images we compute the

Stereo images	Mean Square Error [ $rad^2$ ]
Venus	0.043
Cones	0.036
Teddy	0.094
Tsukuba	0.180

TABLE I: The angle deviation from the optimality



(a) Cones ground truth disparity map (b) Cones estimated disparity map

Fig. 2: Cones estimation

estimated disparity direction and after that we extract the mean square error.

#### IV. COMPARISON AND RESULTS

In this section we present the results obtained through simulation. The bio-inspired model was first coded in Matlab to prove its correctness and afterward, in order to minimize the computational time, some key functions (e.g. 2D convolution) were implemented in CUDA.

Qualitatively we can perform a dense stereo estimation map of 383x434 pixels (the size of some Middlebury images)

BioPsyASW [80]	64.2	3.62	65	5.52	69	14.6	65	3.15	71	4.20	72	20.4	74	11.5	71	18.2	68	23.2	65	4.93	47	13.0	60	11.7	43	11.2
BP+MLH [40]	65.2	4.17	72	6.34	77	14.6	66	1.96	66	3.31	67	16.8	67	10.2	63	18.9	70	24.0	69	4.93	48	15.5	68	12.3	50	11.1
H-Cut [76]	65.8	2.85	86	4.86	65	14.4	64	1.73	63	3.14	65	20.2	72	10.7	65	19.5	71	25.8	74	5.46	69	15.6	69	15.7	66	11.7
SAD-IGMCT [52]	68.0	5.81	83	7.14	81	22.6	84	2.61	69	3.33	69	25.3	79	9.79	59	15.5	59	25.7	73	5.08	61	11.5	46	15.0	84	12.5
DPVI [67]	72.8	4.76	75	5.83	71	16.6	72	4.89	77	5.66	77	22.9	77	11.0	68	16.2	63	23.4	67	9.64	74	15.6	70	23.5	83	13.3
DP [1b]	73.4	4.12	71	5.04	66	12.0	57	10.1	66	11.0	86	21.0	75	14.0	74	21.6	74	20.6	59	10.5	79	19.1	76	21.1	76	14.2
Bipartite [78]	74.6	2.54	81	4.41	60	13.6	62	6.62	79	7.46	78	18.6	71	16.9	81	24.1	80	30.2	79	15.1	87	21.8	85	23.0	82	15.4
PhaseBased [31]	77.1	4.26	73	6.53	78	15.4	70	6.71	80	8.16	80	26.4	81	14.5	75	23.1	75	25.5	72	10.8	81	20.5	83	21.2	77	15.3
RegionalSup [38]	77.8	3.99	69	6.05	73	14.2	63	8.14	82	9.68	83	36.8	87	18.3	85	26.7	85	32.1	80	9.16	73	19.3	79	19.9	74	17.0
IMCT [62]	78.0	4.54	74	5.90	72	19.8	80	3.16	72	3.83	71	23.2	78	18.0	84	23.1	76	35.3	84	12.7	82	18.5	77	27.9	86	16.3
SSD+MF [1a]	78.2	5.23	81	7.07	79	24.1	85	3.74	74	5.16	75	11.9	57	16.5	80	24.8	81	32.9	82	10.6	80	19.8	80	26.3	84	15.7
<b>YOUR METHOD</b>	<b>78.5</b>	<b>7.74</b>	<b>87</b>	<b>9.74</b>	<b>88</b>	<b>24.4</b>	<b>86</b>	<b>6.19</b>	<b>78</b>	<b>7.60</b>	<b>79</b>	<b>34.5</b>	<b>84</b>	<b>13.2</b>	<b>73</b>	<b>21.5</b>	<b>73</b>	<b>33.4</b>	<b>83</b>	<b>6.67</b>	<b>88</b>	<b>16.8</b>	<b>72</b>	<b>19.4</b>	<b>71</b>	<b>16.8</b>
SO [1c]	80.2	5.08	78	7.22	82	12.2	58	9.44	85	10.9	85	21.9	76	19.9	86	28.2	86	26.3	75	13.0	84	22.8	86	22.3	79	16.6
MI-nonpara [85]	81.7	5.59	82	7.54	83	18.8	77	7.50	81	8.99	81	35.0	85	17.4	82	25.7	84	36.9	85	10.2	78	19.9	81	22.6	81	18.0
STICA [16]	82.4	7.70	86	9.63	87	27.8	87	8.19	83	9.58	82	40.3	88	15.8	77	23.2	77	37.7	86	9.80	75	17.8	74	28.7	87	19.7
PhaseDiff [23]	82.4	4.89	77	7.11	80	16.3	71	8.34	84	9.76	84	26.0	80	20.0	87	28.0	87	29.0	78	19.8	88	28.5	86	27.5	85	18.8
Rank+ASW [84]	82.7	6.51	85	8.43	85	19.7	79	10.5	87	12.0	87	32.7	83	15.7	76	24.1	79	32.8	81	14.1	85	23.1	87	21.7	78	18.4
LCDM+AdaptWgt [75]	82.8	5.98	84	7.84	84	22.2	83	14.5	88	15.4	88	35.9	86	20.8	88	27.3	86	38.3	87	8.90	72	17.2	73	20.0	75	19.5
Infection [10]	84.0	7.95	88	9.54	86	28.9	88	4.41	76	5.53	76	31.7	82	17.7	83	25.1	83	44.4	88	14.3	86	21.3	84	38.0	88	20.7

Fig. 6: Comparison among the proposed neural architecture for disparity estimation and some state-of-art algorithms [4]; the table is extracted from the online evaluation page of Middlebury Database [5].

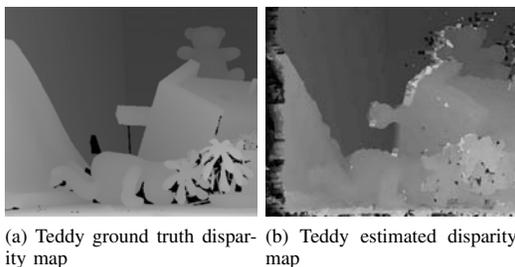


Fig. 3: Teddy estimation

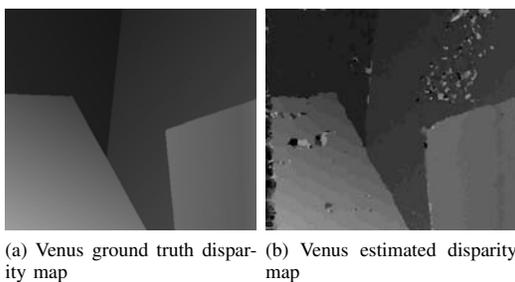


Fig. 4: Venus estimation

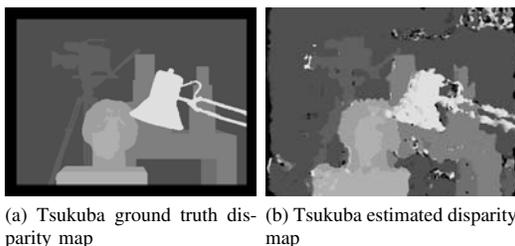


Fig. 5: Tsukuba estimation

in about 12 seconds. The performed simulations are relative to the disparity estimation and the estimated disparity direction. The estimated disparity maps are then submitted to the Middlebury evaluation system and the results are presented (see Figure 6).

The evaluated stereo images are 'Cones', 'Teddy', 'Venus' and 'Tsukuba' (see Figures 2,3,4,5).

It is worth noting that the proposed algorithm is biologically plausible and this should be taken into account when comparing our approach to the other algorithms. We have obtained an improvement of the performance with respect to [18].

The simulations also show that the architecture, with the weighted sum of the direction of the oriented Gabor filter, is able to correctly identify the horizontal directions of the disparity (remember that the stereo images from Middlebury database have only horizontal disparities). Simulations should be performed in order to validate the model even for non-horizontal only disparities.

With respect to the performance reported in [19] we have obtained comparable results, in terms of bad pixel errors.

## V. CONCLUSION AND FUTURE WORKS

### A. Conclusions

In this paper we have presented a bio-mimetic system that computes a disparity map starting from a pair of stereo images. Previous works show the possibility to develop a bio-inspired system and here we proposed a different bio-inspired mechanism in order to improve the performance reported in [18] [19].

In fact, experimental evidences show that the system is more reliable for small disparities than for large disparities [16]. However, as previously shown, natural images have a wide range of possible disparities even in the same scene. One way to overcome this issue is to use coarse Gabor filter large enough to cover large disparities, but empirical evidence (results not published) seems to indicate that the

estimation is not reliable. Besides, the computational cost is too expensive and further research is needed to adopt the simulator for a real-time implementation. Another approach, implemented in our system, is to use a smaller coarse Gabor filter in order to cover small but reliable disparities [16] with an initial position-shift [18] mechanism at the coarse scale.

Experimental evidences and the results presented in [18] seem to confirm the reliability of this type of approach. We proposed a strategy in order to integrate the pooling mechanisms, proposed in [16] and [19], with the position-shift selection at coarse scale based on a bio-mimetic feature analysis proposed in [18].

The obtained results point out that the stereo simulator of the primary visual cortex is a possible candidate for a physical implementation in hardware. Other works show this possibility but in general these implementations do not include the pooling mechanisms of orientations and scales [20] [21], while it has been pointed out how their contribution is fundamental in natural scenes.

The output of the system is a disparity map with associated disparity directions (in the current implementation only horizontal directions); with these decoded features it is possible to correctly estimate the depth of the environment.

### B. Future Works

Future work includes the study of the statistics of the scenes in order to improve the disparity estimation and to reduce the number of neurons in the population. Moreover we will study how to include the color information in the system, that currently works on gray scalar stereo pairs.

### REFERENCES

- [1] R. Pfeifer, et al., *Self-Organization, Embodiment, and Biologically Inspired Robotics*, Science, vol. 318, pp. 1088-1093, 2007
- [2] J. F. V. Vincent, O. A. Bogatyreva, N. R. Bogatyrev, A. Bowyer, A. K. Pahl, *Biomimetics: its practice and theory*, Journal of Royal Society Interface, 3, 41, 2006
- [3] M. Lungarella, O. Sporns, *Mapping information flow in sensorimotor networks*, PLoS Comp. Biology, 2, e144, 2006
- [4] D. Scharstein and R. Szeliski, *A taxonomy and evaluation of dense two-frame stereo correspondence algorithms*. International Journal of Computer Vision, 47(1/2/3):7-42, April-June 2002. Microsoft Research Technical Report MSR-TR-2001-81, November 2001. [vision.middlebury.edu/stereo/](http://vision.middlebury.edu/stereo/)
- [5] Tai S. Lee, *Image Representation Using 2D Gabor Wavelets*, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 18, no. 10, 1996
- [6] I. Ohzawa, G. C. De Angelis and R. D. Freeman, *Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors*, Science, 249, 1037-1041, 1990.
- [7] I. Ohzawa, *Mechanisms of stereoscopic vision: the disparity energy model*, Curr. Opin. in Neurobio., 1998.
- [8] B. Julesz, *Foundations of Cyclopean Perception*, University of Chicago Press, 1971.
- [9] N. Qian, *Computing stereo disparity and motion with known binocular cell properties*, Neural Computation, vol. 6, pp. 390-404, 1994.
- [10] N. Qian, *Binocular Disparity and the Perception of Depth*, Neuron, vol. 18, pp. 359-368, 1997.
- [11] N. Qian and Y. Zhu, *Physiological computation of binocular disparity*, Vision Research, vol. 37, pp. 1811-1827, 1997.
- [12] Y. Zhu and N. Qian, *Binocular receptive fields, disparity tuning, and characteristic disparity*, Neural Computation, vol. 8, pp. 1611-1641, 1996.
- [13] D.J. Fleet, H. Wagner and T.J. Sejnowski, *Neural encoding of binocular disparity: Energy models, position shifts and phase shifts*, Vis. Res., vol. 36, pp. 1839-1857, 1996
- [14] J.J. Tsai and J. D. Victor, *Reading a population code: A multi-scale neural model for representing binocular disparity*, Vis. Res., vol. 43, pp. 445-466, 2003
- [15] Y. Chen and N. Qian, *Coarse-to-fine Disparity Energy Model with Both Phase-Shift and Position-Shift Receptive Field Mechanisms*, Neural Computation, vol. 16, pp. 1545-1577, 2004.
- [16] Erik K. C. Tsang and Bertram E. Shi, *Estimating disparity with confidence from energy neurons*, Advances in Neural Information Processing Systems, vol. 20, 2007
- [17] Erik K. C. Tsang and Bertram E. Shi, *Disparity Estimation by pooling Evidence From Energy Neurons*, IEEE Trans. On Neural Network, vol. 20, no. 11, 2009
- [18] Manuela Chessa, Silvio P. Sabatini and Fabio Solari, *A Fast Joint Bioinspired Algorithm for Optic Flow and Two-Dimensional Disparity Estimation*, Computer Vision System, pp. 184-193, 2009
- [19] K. Shimonomura, T. Kushima and T. Yagi, *Binocular robot vision emulating disparity computation in the primary visual cortex*, Neural Networks, vol. 21, pp. 331-340, 2008.
- [20] Erik K. C. Tsang, Stanley Y. M. Lam, Yicong Meng and Bertram E. Shi, *Neuromorphic implementation of active gaze and vergence control*, Proceedings of the IEEE International Symposium on Circuits and Systems, pp. 1076-1079, 2008