Bio-inspired Binocular Disparity with Position-Shift Receptive Field

Fernanda C. E C. Faria, Jorge Batista, Helder Araújo

Bio-inspired Binocular Disparity with Position-Shift Receptive Field

Fernanda da C. e C. Faria, Jorge Batista and Helder Araújo

Institute of Systems and Robotics,
Department of Electrical Engineering and Computers, University of Coimbra – Polo II,
3030-290 Coimbra, Portugal
{fernanda, batista, helder}@isr.uc.pt

Abstract. Depth perception starts with the binocular interaction receptive fields of simple cells modeled by two Gabor functions followed by a half-squaring function. Simple cells do not have reliable disparity computation. Complex cells combine two simple cells in quadrature. They are better adapted to encode disparity information. The image disparity can be determined by fixing the receptive field in one eye and varying it in the other. Pooling information of spatial frequency and orientation is very important to improve the quality of results of real world stereograms. In this work, a bio-inspired method to calculate binocular disparity based on the energy model for depth perception of real world images is described and implemented. The performance of the proposed method is also evaluated in what concerns the algorithm computational cost, root-mean-square error and percentage of wrongly matched pixels.

Keywords: Depth perception; Binocular disparity; Bio-inspired stereo vision; Energy model.

1 Introduction

In the literature several methods to compute the binocular disparity are described such as statistical approaches, probabilistic approaches and artificial neural networks [13]. The focus of this work is on the application of bio-inspired models.

The brain combines information from both eyes and constructs a visual world that is perceived as single and three dimensional. Primary visual cortex (V1) is the first site at which single neurons can be activated by stimuli in both eyes. Primary visual cortex neurons encode information specifically about the relationship between the images in the two eyes [1]-[4].

Biologically plausible models for stereo vision inspired in the visual cortex employ receptive field (RF) profiles of binocular cells for disparity computation. A region of a scene that is visible by left and right eyes is named binocular view field. All points in the binocular view field have different positions in the left and right eyes [2], [4], [5]. Binocular disparities are the small positional differences between the images from the left and right eyes. The slightly different viewpoint is due the horizontal

separation of the eyes in the head. Depth perception is deduced through binocular disparity [2]-[4].

Stereoscopic vision at the neuronal level based on events within the primary visual cortex can be studied through two subdivisions of cortical cells, named simple cells and complex cells. Simple cells are composed of orientation selective RFs that have spatially separate regions. These regions are mutually antagonistic, i.e., they respond to either ON (excitatory) or OFF (inhibitory) to a stimulus [3], [5]-[9]. The receptive fields of binocular simple cells are selective to binocular disparity and they are also affected by several stimulus parameters, such as [2], [4], [5]: position, contrast polarity, spatial frequency and orientation. The other type of V1 cells, complex cells, are composed of orientation selective RFs and respond to a stimulus anywhere within the RF, i.e., there is not spatially separate regions. Complex cells are not sensitive to changes in stimulus position and contrast polarity. The fact that they do not suffer from the influence of irrelevant parameters makes them well adapted to encode disparity information [3]-[6].

The binocular energy model is a model for disparity-selective complex cells based on a combination of simple-cell subunits. Binocular interaction RFs of a simple cell is described as the product of left and right eye RFs. This interaction can be demonstrated by a linear binocular filter followed by a half-power function [2], [4], [10], [11].

This work describes a bio-inspired method to calculate binocular disparity based on energy model for depth perception of real world images [10]-[11].

2 Contribution to Technological Innovation

Visual perception is essential for many applications, especially in Robotics. For example, in applications related to human-robot interaction, the machine needs to cope with and share the human environment in a friendly way. For applications such as assistants in surgery, service robots, caretakers for the elderly and disabled people, a technological solution is necessary requiring advanced perception capabilities as well as advanced skills that allow a robot to perform the desired tasks.

The properties of a scene analyzed by a robot head can involve tracking movements, focusing by vergence and 3D perception of the scene structure. The relative orientation of the eyes, the target location and depth are important for surveillance activities [15] [16].

The algorithm presented in this paper is a preliminary work related to the development of bio-inspired visual perception modules for robotic heads. The model created is based on two algorithms [10], [11]. The principal idea is to develop a faster algorithm to compute disparity maps employing the concepts of energy model for real world stereograms.

3 Stereoscopic Model

Binocular disparity forms the basis of stereoscopic depth perception. The binocular energy model is based on a combination of simple cells. Each complex cell is the linear combination of the outputs of four simple type subunits. A typical binocular simple cell is identical to two Gabor functions (equation 1), one for each of its receptive fields in the left and the right retinas [10].

$$g(x, y, \phi) = \exp\left(-\frac{x'^2 - y'^2}{2\sigma^2}\right)\cos(\omega x' - \phi) \quad (\mathbf{1})$$

where ω is the cell preferred spatial frequency ($2\pi f$), θ is its preferred orientation, ϕ is its phase, $x'=x\cos(\theta)+y\sin(\theta)$ and $y'=y\cos(\theta)-x\sin(\theta)$. The standard deviation σ is computed as follows:

$$\sigma = \frac{\sqrt{\ln 2}}{2\pi f} \times \frac{2^{1.5}+1}{2^{1.5}-1} \quad (\mathbf{2})$$

where the half-maximum bandwidth is 1.5 octaves [10].

The left and right RFs are shifted in position (equations 3 and 4 – the receptive field is fixed in one eye and varying it in the other) [11].

$$g_l(x, y) = g(x, y, \phi) \quad (\mathbf{3})$$

$$g_r(x, y) = g(x + d, y, \phi) \quad (\mathbf{4})$$

where d is a positional difference between the left and right eyes Gabor functions.

The response of a simple cell can be determined by first filtering, for each eye, the retinal image by the corresponding receptive field profile (equation 5), and then adding the two contributions from the two eyes (equation 6) [12].

$$L = \int_{-\infty}^{+\infty} I_l(x, y)\, g_l(x, y)\, dx\, dy \quad (\mathbf{5})$$

$$S = L + R \quad (\mathbf{6})$$

The response of a complex cell can be modeled by summing the squared outputs of a quadrature pair of simple cells, i.e., their spatial phases differ by 90 degrees while all the other parameters of the cells are identical (equation 7). The combination of simple cells in quadrature eliminates the Fourier phase dependence. Therefore, complex cells encode reliable disparity information [7], [9], [12].

$$C = S_1^2 + S_2^2 \quad (\mathbf{7})$$

Briefly, the process of depth perception begins with the binocular interaction RFs of simple cells. They can be modeled by two Gabor functions followed by a half-squaring function. Simple cells depend on the Fourier phase of the stimulus. As a result they do not perform a reliable disparity computation. Complex cells combine two simple cells in quadrature and eliminate the Fourier phase dependence. They are better adapted to encode disparity information than simple cells.

3.1 Computing Disparity

To compute the image disparity, the most responsive complex cell at each spatial location is selected at zero phase disparity for the first spatial frequency (*f*). We use 6 spatial frequencies (0.5, 1.0, 2.0, 4.0, 8.0 and 16.0 cycles per degree) and 6 RF orientations (30, 60, 90, 120, 150 and 180 degrees). Next the disparity responses of the complex cells population with the fixed *f* are computed, for the full range of the positional differences (*d*) for each θ. Complex cells at a local maximum or a local minimum are selected and if it is not the first spatial frequency *f*, then the selected complex cell is the cell with the closest filtered *d* value computed. If spatial frequency $f = 0.5$, then the selected complex cell is the one with the biggest *d* value. Another situation is when a local maximum or a local minimum does not exist then the *d* value is -99. The value -99 means that the *d* value will be ignored in the pooling final disparity map. The result images and the filtered images are saved. A 5x5 median filter is used. The size of the receptive field decreases when *f* increases. The range of values for *d* is split into several steps. The steps have a small and fixed size. Fig. 1 shows a block diagram of the algorithm.

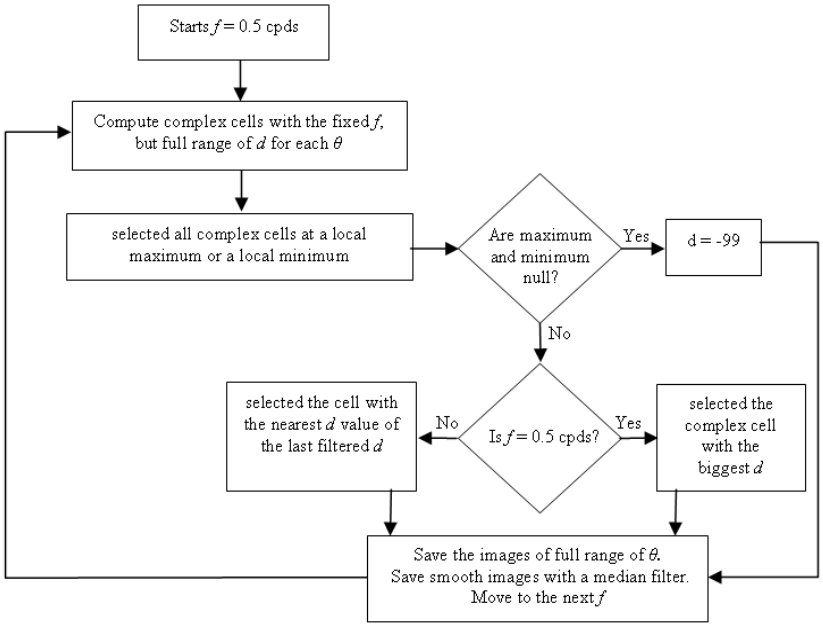

Fig. 1. Block diagram of the algorithm.

The outputs of the algorithm are 36 images (6 spatial frequencies and 6 orientations) that will be pooled for the computation of the final disparity map. The images employed in the pooling are the images saved without filtering. The values of the filtered images are used only during the process of selection of d (Fig. 1).

3.2 Pooling

A disparity map created by pooling the cells with multiple spatial frequencies and orientations produces results that are more reliable and accurate than those that would be obtained without the pooling step. The pooling is the same method described in [10] where each spatial position of the 36 disparity maps with different spatial frequencies and orientations are combined to compute the average. The resulting average is compared with each spatial position of all maps and the position whose value is farthest from the mean is removed. The process is repeated until the total of spatial positions is reduced by half.

4 Evaluation Methodology and Results

The algorithm evaluation was done employing two quantitative measures of fit quality: root-mean-square error (R) and percentage of wrongly matched pixels (B) [13]. R is computed using equation 8,

$$R = \sqrt{\frac{1}{N}\sum_{(x,y)} | d_C(x,y) - d_T(x,y) |^2} \quad (8)$$

where d_C is the computed disparity map, d_T is the ground truth map and N is the total number of pixels.

Equation 9 is used to calculated B with a disparity error $\delta_d = 1$.

$$B = \frac{1}{N}\sum_{(x,y)} (| d_C(x,y) - d_T(x,y) | > \delta_d), \quad (9)$$

Fig. 2 and Fig. 3 show the resulting disparity maps for the images from the Middlebury stereo repository [13] (http://cat.middlebury.edu/stereo). Fig. 4 also presents an output disparity map for the method described in section 3.

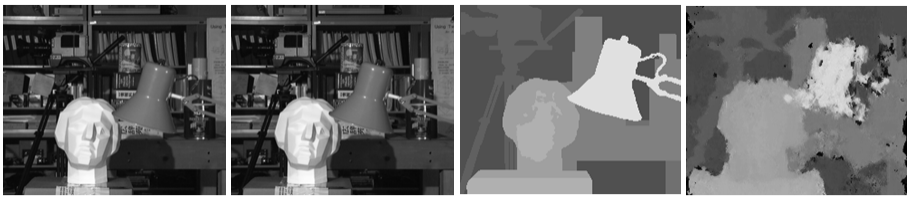

Fig. 2. (a) Left image; (b) right image; (c) the ground truth disparity map; (d) the estimated disparity map.

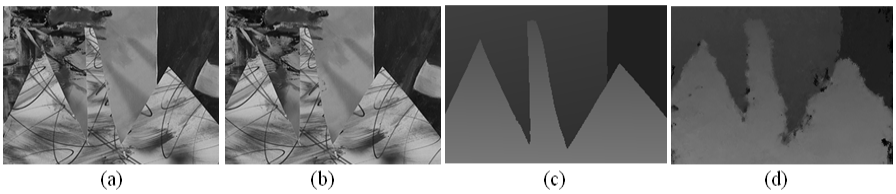

Fig. 3. (a) Left image; (b) right image; (c) the ground truth disparity map; (d) the estimated disparity map.

Fig. 4. (a) Left image; (b) right image; (c) the estimated disparity map.

Table 1 shows the results of root-mean-square error (*R*) and the percentage of wrongly matched pixels (*B*) to Figures 2 (d) and 3 (d). Errors of table 1 are only evaluated in non-occluded regions [13] (http://cat.middlebury.edu/stereo).

Table 1. Evaluation results.

	R	B
Fig.2 (d)	2.0	20
Fig.3 (d)	1.5	11

5 Discussion of the Results

Several tests were performed with the algorithm described, using a Pentium 4 with 3.00GHz and 2.00GB of RAM memory. The operating system used was Suse Linux 10.

The algorithm proposed by Chen and Qian [11] (using a coarse-to-fine approach) was implemented in C++ and tested with the goal of analyzing its performance with real world images and to check whether it would be possible to run it in real-time. All functions that are used to perform convolution were implemented in the frequency domain to allow a decrease in the processing time. However, in spite of several optimizations, it was not possible to obtain real-time performance.

The coarse-to-fine algorithm spent approximately twenty seven minutes to process an image with 256x256 pixels (Fig. 4. (a) and (b)). The coarse-to-fine algorithm can be improved using multiple processors and a library with multithreading support. The algorithm proposed by Read and Cumming [10], was also implemented in C++. However its processing time was higher than the processing time of the coarse-to-fine algorithm. The test with the images presented in Fig. 4. (a) and (b) required five days and nineteen hours of running time.

The proposed algorithm in this work required 7 minutes and 33 seconds to obtain the disparity map of Fig. 4 (c). For Fig. 2 (d) the algorithm spent 8 min and 45 sec. and for Fig. 3. (d) the disparity map was computed in 6 min. and 20 sec. The method described method is based partially on the algorithms presented in [10] and [11].

The root-mean-square error and the percentage of bad matching pixels presented in table 1 are similar to the results shown in [10] but do not have yet the precision of stereo algorithms in 13 (http://cat.middlebury.edu/stereo).

6 Conclusion and Future Work

The estimation of disparity maps using the energy model is slow. Real images are inherently complex due to the non-uniformity of the disparity values. This problem can be handled by pooling images with different orientations and spatial frequencies. On the other hand, pooling requires additional processing time. One alternative to develop a real time system (using models of populations of cortical neurons) is to employ DSPs and FPGA chips [14].

As future work, we intend to improve the precision of the algorithm and its suitability for tracking applications. Depth accuracy is crucial for binocular vergence control.

References

1. Cumming, B.G., DeAngelis, G.C.: The Physiology of Stereopsis. Annual Review of Neuroscience, Vol. 24, pp. 203--238 (2001)
2. Anzai , A., Ohzawa, I., Freeman, R.D.: Neural Mechanisms for Processing Binocular Information. I. Simple Cells. Journal of Neurophysiology, 82, pp. 891--908 (1999)

3. Parker, A.J.: Binocular Depth Perception and the Cerebral Cortex. Nature Reviews, Neuroscience, Vol. 8, pp. 379--391 (2007)
4. Qian, N.: Computing Stereo Disparity and Motion with Known Binocular Cell Properties. Neural Computation, 6, pp. 390--404 (1994)
5. Ohzawa, I., DeAngelis, G.C., Freeman, R.D.: Stereoscopic Depth Discrimination in the Visual Cortex: Neurons Ideally Suited as Disparity Detectors. Science, Vol. 249, pp. 1037--1041 (1990)
6. Hubel, D.H., Wiesel, T.N.: Receptive Fields, Binocular Interaction, and Functional Architecture in the Cat's Visual Cortex. J. Physiol, 160, 106--154 (1962)
7. Qian, N.: Binocular Disparity and the Perception of Depth. Neuron, Vol. 18, No. 3, pp. 359--368 (1997)
8. Ohzawa, I., DeAngelis, G., Freeman, R.: Encoding of Binocular Disparity by Simple Cells in the Cat's Visual Cortex. J. Neurophysiol, 75, pp. 1779--1805 (1996)
9. Read, J.: Early Computational Processing in Binocular Vision and Depth Perception. Progress in Biophysics and Molecular Biology, 87, pp. 77--108 (2005)
10. Read, J.C.A., Cumming, B.G.: Sensors for Impossible Stimuli May Solve the Stereo Correspondence Problem. Nature Neuroscience, Vol. 10, No. 10, pp. 1322--1328 (2007)
11. Chen, Y., Qian, N.: A Coarse-to-Fine Disparity Energy Model with Both Phase-Shift and Position-Shift Receptive Field Mechanisms. Neural Computation, 16, pp. 1545--1577 (2004)
12. Qian, N., Zhu, Y.: Physiological Computation of Binocular Disparity. Vision Research, Vol. 37, No. 13, pp. 1811--1827 (1997)
13. Scharstein, D., Szeliski, R.: A Taxonomy and Evaluation of Dense Two-frame Stereo Correspondence Algorithms. Int. J. Comput. Vis., 47, pp. 7--42 (2002)
14. Tsang, E.K.C., Lam, S.Y.M., Meng, Y., Shi, B.E.: Neuromorphic Implementation of Active Gaze and Vergence Control. ISCAS - IEEE Intl. Symp. on Circuits and Systems, pp. 1076--1079 (2008)
15. Batista, J., Peixoto, P., Araújo, H.: A Focusing-by-Vergence System Controlled by Retinal Motion Disparity. ICRA – Proc. of the IEEE International Conference on Robotics and Automation, San Francisco, CA, pp. 3209--3214 (2000)
16. Batista, J., Peixoto, P., Araújo, H.: Binocular Tracking and Accommodation controlled by Retinal Motion Flow. ICPR – Proc. of the Int. Conference on Pattern Recognition, Barcelona, Spain, Vol. 1, pp. 171--174 (2000)