

# Log-Polar Stereo for Anthropomorphic Robots

Enrico Grosso and Massimo Tistarelli

DIST - University of Genoa, Computer Vision Laboratory  
Via Opera Pia, 13 - 16145 Genoa, Italy  
{grenri,tista}@dist.unige.it

**Abstract.** Stereoscopic vision has a fundamental role both for animals and humans. Nonetheless, in the computer vision literature there is limited reference to biological models related to stereoscopic vision and, in particular, to the functional properties and the organization of binocular information within the visual cortex.

In this paper a simple stereo technique, based on a space variant mapping of the image data and a multi-layered cortical stereoscopic representation, mimicking the neural organization of the early stages of the human visual system, is proposed. Radial disparity computed from a stereo pair is used to map the relative depth with respect to the fixation point. A set of experiments demonstrating the applicability of the devised techniques is also presented.

## 1 Introduction

Stereopsis is a fundamental visual cue for depth computation and for the organization of spatial perception. Stereo vision has been proposed for the recovery of the 3D structure of a scene [1] but stereo techniques have been also used for object and face recognition [2,3].

The realization of anthropomorphic robots has been inspired for a long time by physiological and neuro-physiological studies of the human sensory-motor system. In particular the studies on the human visual system show how the structure of retina implies a space-variant representation of the visual stimuli. This is due to the size and distribution of the receptive fields in the retina [4]: the resolution is very high in the center of the visual field (fovea) and decreases from the fovea toward the periphery. In the literature, the log-polar mapping has been proposed to model the non-uniform distribution of the receptive fields in the retina [12,13]. By using this transformation the image information is represented at different spatial-frequencies, yet maintaining a fine detail within the neighborhood of the fixation point.

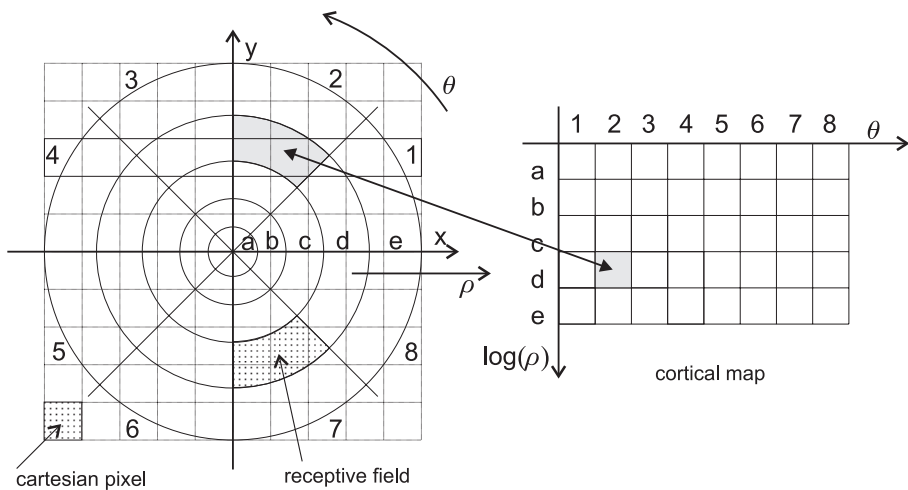
Despite the great importance of stereopsis in many visual tasks [6,7] and the good properties of the log-polar mapping [8] very few researchers exploited the advantages of coupling stereo vision and log-polar sampling [9,10,11].

This paper presents a stereo matching technique specially devised to benefit of the properties of the non-uniform, log-polar image representations. The organization of the paper is as follows. Section 2 describes the mathematical

formulation of the log-polar sampling while the cortical model is described in section 3. The stereo matching technique is described in section 4. Finally, section 5 and 6 describe some experimental results and outline future developments.

## 2 Log-Polar Sampling

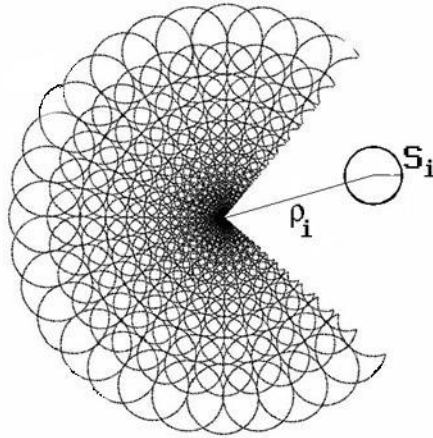
Many studies about the human visual system show that the retina performs a spatial sampling that can be modeled by a discrete distribution of elements (receptive fields) whose size increase linearly with the eccentricity. The transformation of the retinal image onto its cortical projection can be described by a mapping of the cartesian coordinates on the retinal plane  $(x, y)$  into polar coordinates  $(\rho, \theta)$  and then in the cortical plane  $[\log(\rho), \theta]$ . The diagram in figure 1 sketches this mapping for one receptive field covering several pixels on the cartesian plane. The formalization of the retino-cortical transformation (figure



**Fig. 1.** Retino-cortical transformation

2) used in this paper likely differs from the models proposed by [12] and [13]. The advantage of the proposed transformation is the possibility to control in a simple way the overlap among neighboring receptive fields. The parameters required for log-polar sampling are the number of eccentricities ( $N_r$ ), the number of receptive fields per eccentricity ( $N_a$ ) and the radial and angular overlap of neighboring receptive fields ( $O_r$  and  $O_a$ ). The overlap which occurs along each circle is controlled by the parameter  $K_0$ :

$$K_0 = \frac{S_i}{\rho_i}$$



**Fig. 2.** Retinal model

where  $S_i$  is the radius of the considered receptive field and  $\rho_i$  is the eccentricity. For adjacent fields the  $K_0$  value can be obtained considering that

$$2\pi\rho_i \cong 2S_iN_a.$$

Hence, we can observe that there is an overlap along each circle only when

$$K_0 > \frac{\pi}{N_a}.$$

Concerning the overlap between receptive fields at different eccentricities a second parameter  $K_1$  is defined

$$K_1 = \frac{S_i}{S_{i-1}} = \frac{\rho_i}{\rho_{i-1}}.$$

The relationship between  $K_0$  and  $K_1$  for adjacent receptive fields can be expressed as:

$$K_1 = \frac{\rho_i}{\rho_{i-1}} = \frac{1 + K_0}{1 - K_0}.$$

There is an overlap along each radial coordinate when

$$K_1 < \frac{1 + K_0}{1 - K_0}.$$

In practice, given the parameters  $N_r, N_a, O_r, O_a$ , the  $K$  values are computed as follows:

$$K_0 = \pi \frac{O_a}{N_a}, \quad K_1 = \frac{O_r + K_0}{O_r - K_0}.$$

Concerning the smallest eccentricity  $\rho_0$  used for the mapping, it obviously depends on the size of the receptive fields in the fovea  $S_0$ :

$$\rho_0 = \frac{S_0}{K_0}$$

In practice it is reasonable to assume  $\rho_0 \in [0.5 - 5]$ , in such a way to preserve as much as possible the original resolution of the cartesian image.

### 3 Multilayer Cortical Model

The analysis of the primary visual cortex shows a layered organization, where each layer differs in response complexity; for example receptive fields can have very different sizes and respond strictly to a properly oriented visual stimulus. Crossing the cortex in a direction parallel to the layer surfaces, the organization of the information shows alternation from left eye to right eye and back, with thin stripes 0.5 millimeter wide (ocular dominance columns).

In order to simulate this structure and the computational processing underlying depth perception we introduce a cortical model obtained by fusing two input cortical images (see figure 3).

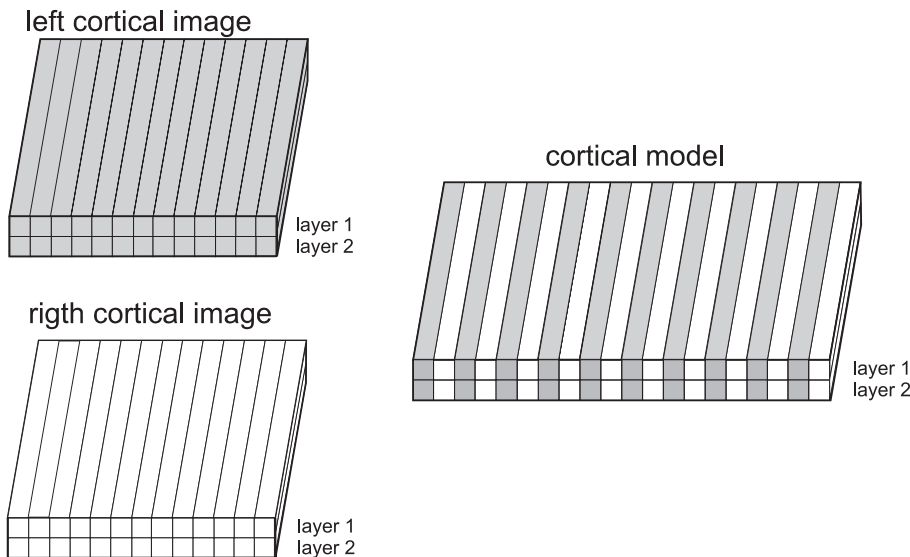


Fig. 3. Cortical model

In this model the left and right cortical images are combined along the radial direction. The first layer simply contains the grey level information while the second layer is the result of the application of a Laplacian of Gaussian and a zero crossing filter over the receptive fields of the retinal model.

The Laplacian of Gaussian mask is approximated with a Difference of Gaussians (DoG) [12] that best matches the circular symmetry of the corresponding receptive field:

$$F(\rho, \sigma_i, \sigma_e) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(\frac{-\rho^2}{2\sigma_i^2}\right) - \frac{1}{\sqrt{2\pi}\sigma_e} \exp\left(\frac{-\rho^2}{2\sigma_e^2}\right)$$

where  $\sigma_i$  and  $\sigma_e$  are proportional, respectively, to the diameters of the surround and center regions and  $\sigma_i/\sigma_e = 2$ . In order to implement a discrete filter we calculate a priori a mask for any possible size of receptive field. For zero crossing extraction a simple 3x3 mask is applied to the log-polar representation. The slope of the zero crossing is then coded as a gray level in the range 0-254. In figures 4 to 6, the application of the multilayer cortical model to a stereo image pair taken from a real scene, is shown.

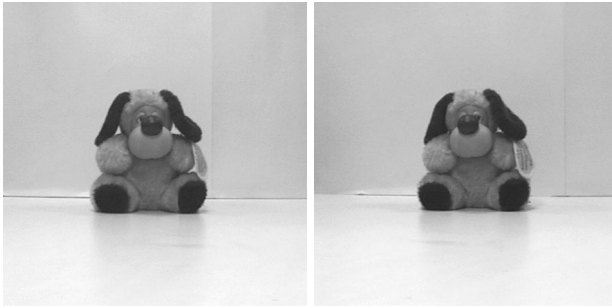
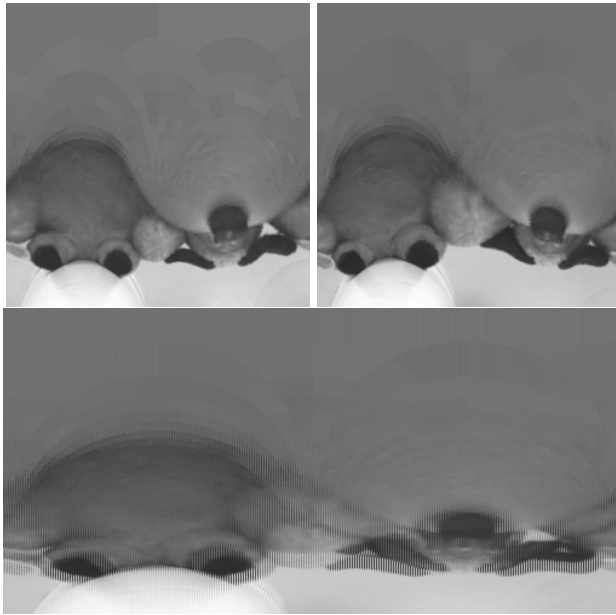


Fig. 4. Cartesian stereo images

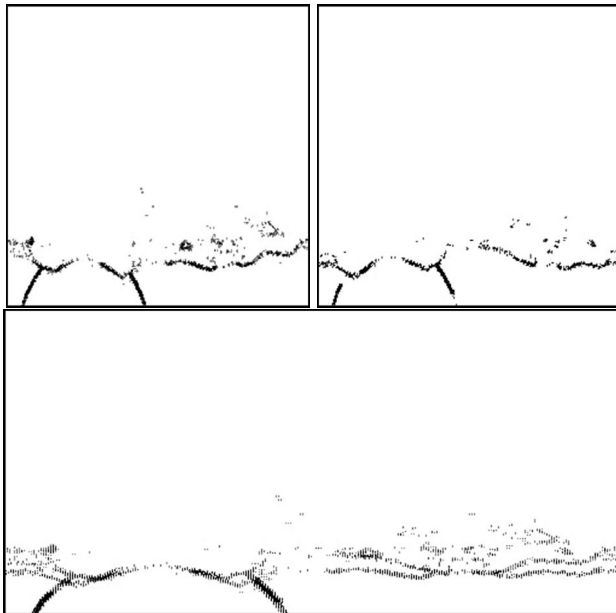
## 4 Cortical Stereo

The image projection of objects onto a stereo camera pair is subject to changes in position, form and, according to light, even color. The problem of establishing the correspondence between points in stereo images has been deeply investigated in the past and a number of techniques has been proposed for cartesian images. Considering space-variant images some additional issues have to be taken into account:

1. the epipolar constraint, which is usually exploited for stereo matching, on the cortical plane becomes a non-linear function which projects curves instead of straight lines on the image plane;

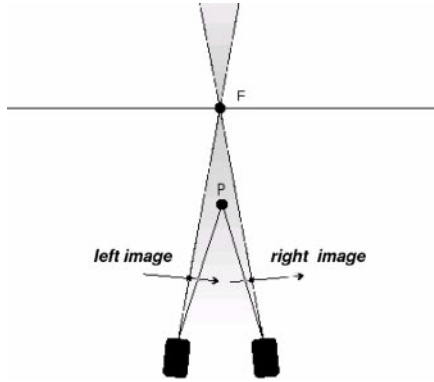


**Fig. 5.** (top) Log-polar transformation of left and right grey level images; (bottom) The resulting **layer 1** of the cortical model where left and right columns are combined.



**Fig. 6.** (top) Log-polar transformation of left and right contour images; (bottom) The resulting **layer 2** of the cortical model where left and right columns are combined.

- in few special cases the projections on the two images of the same point in space lie on the opposite sides (hence very far one from the other) of the log-polar plane. This may happen whenever the point in space is located within the inter-ocular cone of the visual field between the cameras and the fixation point, as shown in figure 7. On the other hand, in a real system the fixation mechanism will control the vergence of the stereo cameras in order to fixate the closest point within the inter-ocular visual cone.



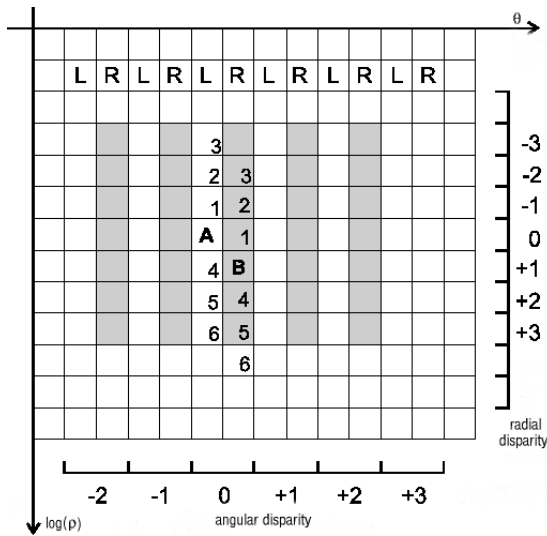
**Fig. 7.** A point belonging to the inter-ocular visual cone produces projections on the log-polar plane which can be very far one from the other. As a consequence this point cannot be matched.

The stereo matching process relies on the multi-layer information of the cortical representation. First of all candidate points are selected by looking at the zero crossing image (layer 2). Then, candidates in adjacent columns are compared by using the gray level information (layer 1). Figure 8 depicts this process in the cortical plane: two candidate contour points  $A(\theta_A, \log(\rho_A))$  and  $B(\theta_B, \log(\rho_B))$  are compared by computing the following expression:

$$C(\theta_A, \log(\rho_A), \theta_B, \log(\rho_B)) = \sum_i [I_c(\theta_A, \log(\rho_A) + i) - I_c(\theta_B, \log(\rho_B) + i)] \cdot w_i$$

where  $I_c()$  is the cortical layer 1 and  $w_i \in [0, 1]$  is a normalizing weight factor. The index  $i$  usually cover a small range around the candidate point, few pixels at most. If the computed value of  $C()$  exceeds a given threshold the two candidates are matched and they are removed from the candidate list. The disparity computed along the radial direction is expressed by:

$$d_r = \log(\rho_A) - \log(\rho_B)$$



**Fig. 8.** Two candidate corresponding points, derived from layer 2, are compared by using gray level information in a small neighborhood (1x7 pixels wide).

This value can be interpreted with reference to figure 9 where the sign of the disparity is related to the position of the object with respect to the fixation point. For instance, an object located behind the fixation point and in the right side has negative radial disparity .

### 5 Experimental Results

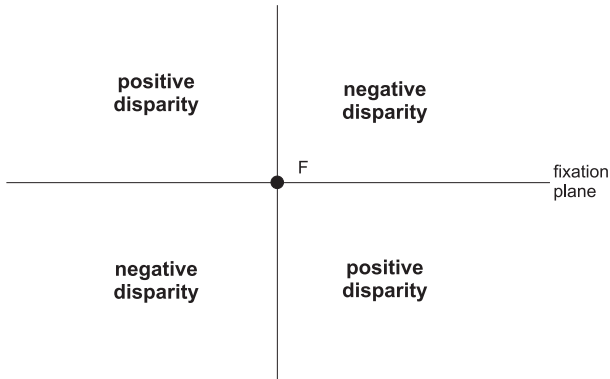
In order to assess the validity of the proposed approach, several experiments have been performed both on synthetic and real images. Two off-the-shelf black/white cameras have been used, with a stereo baseline of about 200 mm.

For the first experiment two synthetic stereo images were generated with the fixation point set at 1500 mm from the virtual cameras. The scene contains a vase, on the left side, lying between the cameras and the fixation point in space and a cube, on the right side, located behind the fixation point. A diagram of the scene and the two cartesian stereo images are shown in figure 10.

The log-polar mapping was applied to the original stereo images generating the two layers of the cortical model (figure 11). The edge and gray-level space-variant representations of the two images are fused together to obtain the final cortical representation (figure 12).

In figure 13, the result of the stereo matching, based on the multi-layer cortical representation, is shown. In the presented maps the gray level codes the disparity between -127 (black) and 128 (white), with 0 represented as light gray. As it can be noticed, most of the pixels of the two objects have a negative disparity. This result can be readily and correctly interpreted with the table of figure





**Fig. 9.** Disparity areas with respect to the fixation point (top view)

9. As it can be noticed there are few errors on the upper right corner of the square and small portions of the vase. By comparing the computed disparity represented on the cortical plane and the ocular dominance representation in figure 12, it appears that most of the errors are due to ambiguities in the matching between corresponding contours. For example, considering the wrong disparity in the upper right edge of the square, as the edges of the two images are almost coincident there is an ambiguity along the direction of the edge.

The same experiment described for the synthetic scene has been performed on images acquired from a real scene. For the first experiment two stereo images were acquired with the fixation point set at 1500 mm from the cameras. The scene contains two puppets, one on the left side, lying between the cameras and the fixation point in space and another, on the right side, located behind the fixation point. A diagram of the scene and the two stereo images are shown in figure 14. In figure 15 the cortical representation is shown.

In figure 16, the result of the stereo matching is shown. Also in this experiments few errors are due to to ambiguities in the matching between corresponding contours.

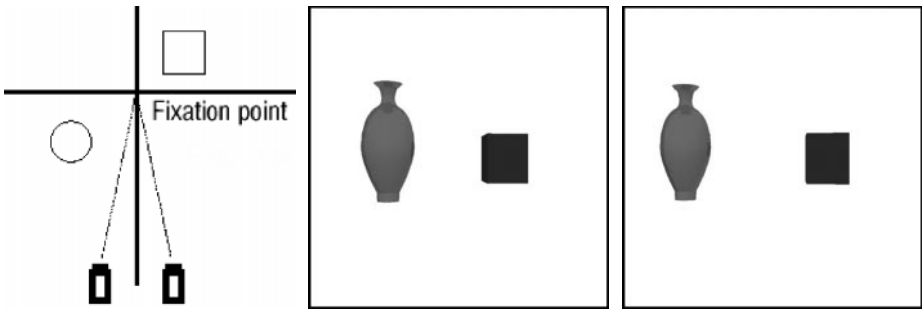
In the last experiment two stereo images were acquired inside the Computer Vision laboratory with the fixation point set at 1500 mm from the cameras. The scene contains several clearly distinguished objects located at different positions with respect to the fixation point which is 2500 mm far from the stereo cameras. A diagram of the scene and the two stereo images are shown in figure 17. Figures 18 and 19 show the cortical model and the final result of the stereo matching algorithm. Even though the scene is far more complex than those presented in the previous experiments, still the most prominent regions in the images (the man seated on the right and the pole on the left) clearly show disparity values which are consistent with their position in space.

## 6 Conclusion

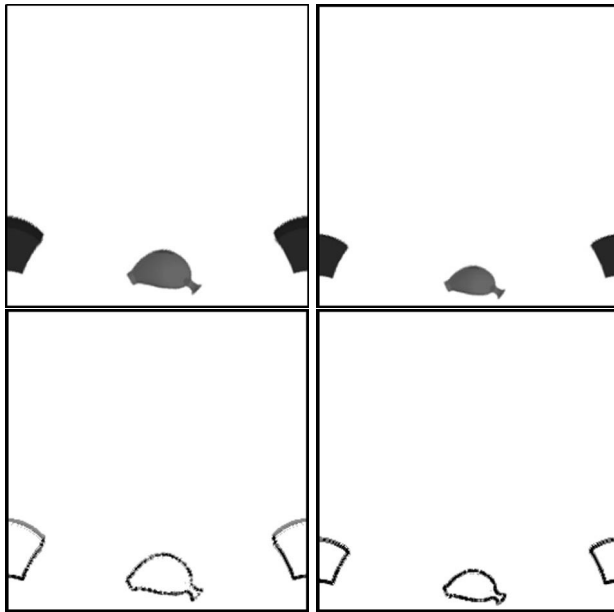
In this paper a stereo matching technique, based on space-variant information, is proposed. Radial disparity is used to map the relative distance from the fixation point. This information, so far very rough, turns out to be quite robust even in complex environments. It is therefore an interesting cue for an active vision system, provided with fixation capabilities. Future work will be devoted to the refinement of the proposed technique, by using multiple layers (computed at different spatial resolutions) and by assessing the resolution in depth attainable in a neighborhood of the fixation point.

## References

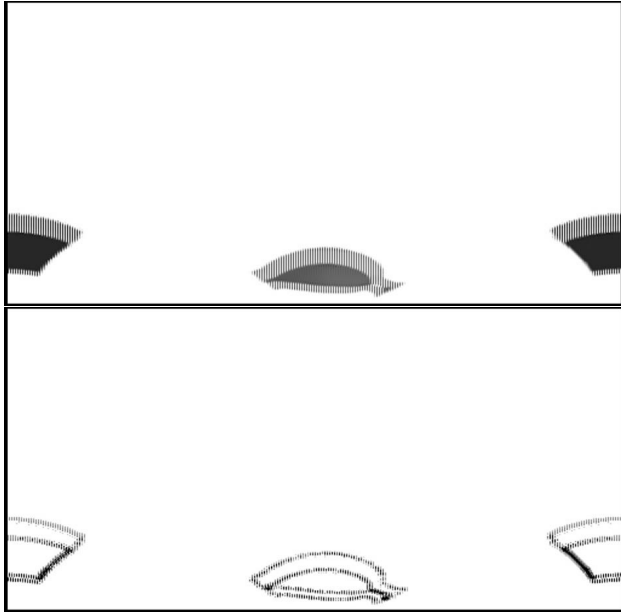
1. S. Grossberg and N.P. McLoughlin, "Cortical dynamics of three-dimensional surface perception: binocular and half-occluded scenic images," *Neural Network*, Vol. 10, Num. 9, pp. 1583-1605, 1997.
2. F.L. Lim and G.A.W. West and S. Venkatesh, "Use of log-polar space for foveation and feature recognition," *IEE Proceedings-Vision, image and Signal Processing*, Vol. 144, Num. 6, pp. 323-331, 1997.
3. Ming Xu, T. Akatsuka, "Detecting head pose from stereo image sequence for active face recognition," *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, Japan 1998.
4. D. H. Hubel, "Eye, Brain, and Vision," *Scientific American Library*, 1988.
5. C.F.R. Weiman and G. Chaikin, "Logarithmic Spiral Grids for Image Processing and Display," *Comp. Graphic and Image Processing*, Vol. 11, pp. 197-226, 1979.
6. E. Grosso and M. Tistarelli, "Active/Dynamic Stereo Vision," *IEEE Trans. on PAMI*, Vol. 17, Num. 9, pp. 868-889, 1995.
7. M. Xie and C.M. Lee and Z.Q. Li and S.D. Ma, "Depth assessment by using qualitative stereo-vision," *1997 IEEE International conference on intelligent processing System*, Beijing, China, 1997.
8. G. Sandini and V. Tagliasco, "An Anthropomorphic Retina-like Structure for Scene Analysis," *CGIP*, Vol. 14, Num. 3, pp. 365-372, 1980.
9. C. Capurro, F. Panerai and G. Sandini, "Vergence and tracking fusing log-polar images," *Proceedings of the 13th International Conference on Pattern Recognition*, Wien, Austria 1996.
10. C. F. R. Weiman, "Binocular stereo via log-polar retinas," *Proceedings of the SPIE - the International Society for Optical Engineering*, Vol 2488, pp. 309-320, 1995.
11. N. Oshiro and A. Nishikawa and N. Maru and F. Miyazaki, "Binocular tracking based on retino-cortical mapping," *Transactions of the Institute of Systems, Control and information Engineers*, Vol. 10, Num. 6, pp. 287-296, 1997.
12. C. Braccini and G. Gambardella and G. Sandini and V. Tagliasco, "A Model of the Early Stages of the Human Visual System: Functional and Topological Transformation Performed in the Peripheral Visual Field," *Biol. Cybern.*, Vol. 44, pp. 47-58, 1982.
13. M. Tistarelli and G. Sandini, "On the Advantages of Polar and Log-Polar Mapping for Direct Estimations of Time-to-Impact from Optical Flow," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 15, Num. 4, 1993.



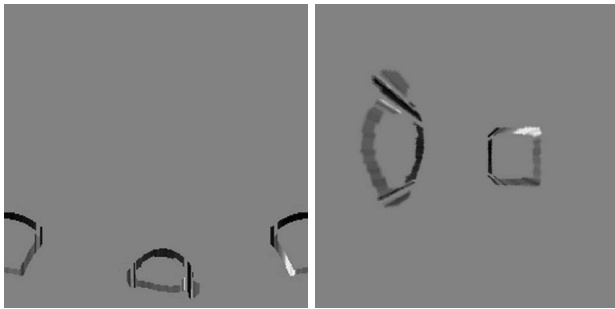
**Fig. 10.** (left) Top view of the setup adopted for the synthetic scene; (right) Synthetic stereo-images used in the first experiment.



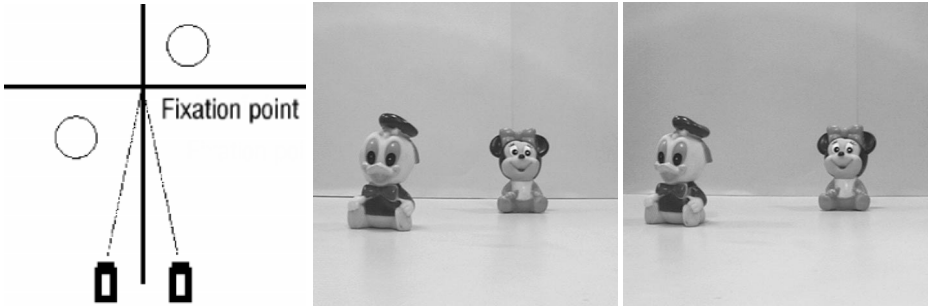
**Fig. 11.** (top) Space-variant stereo images and (bottom) computed zero crossings after the application of the difference of Gaussians receptive field to the original stereo images.



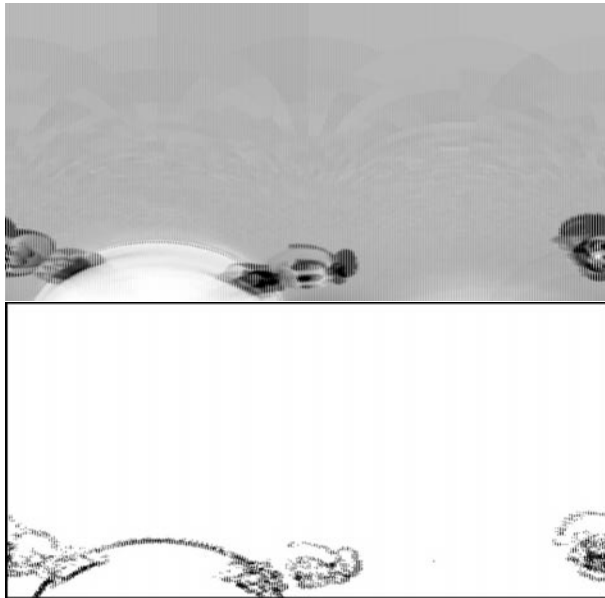
**Fig. 12.** Cortical representation for layer 1 (top) and layer 2 (bottom).



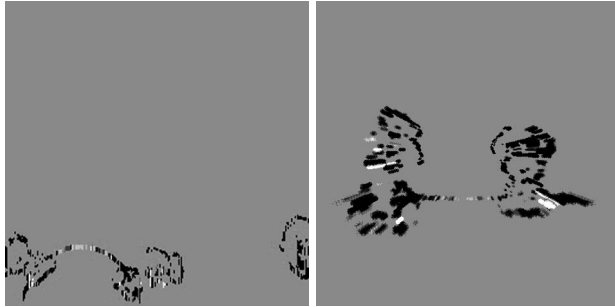
**Fig. 13.** Computed disparity represented on the (left) cortical and (right) cartesian plane.



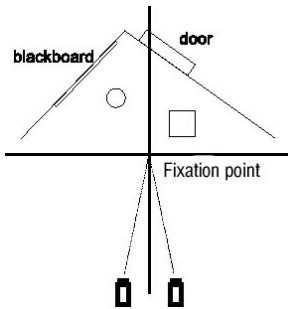
**Fig. 14.** (left) Setup for the stereo image acquisition of the “puppets” scene; (right) Stereo-images acquired from the “puppets” scene.



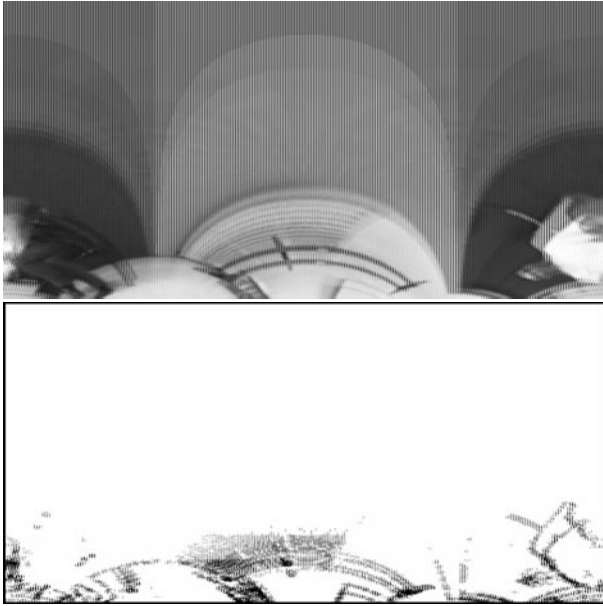
**Fig. 15.** Cortical representation for layer 1 (top) and layer 2 (bottom).



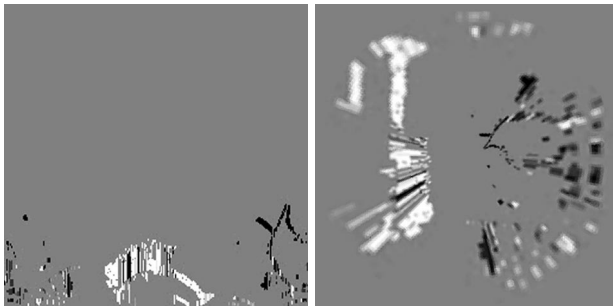
**Fig. 16.** Computed disparity represented on the (left) cortical and (right) cartesian plane.



**Fig. 17.** (left) Setup for the stereo image acquisition of the “laboratory” scene; (right) Stereo-images acquired from the “laboratory” scene.



**Fig. 18.** Cortical representation for layer 1 (top) and layer 2 (bottom).



**Fig. 19.** Computed disparity represented on the (left) cortical and (right) cartesian plane.