Design strategies for direct multi-scale and multi-orientation visual processing in the log-polar domain

Fabio Solari, Manuela Chessa, Silvio P. Sabatini

Department of Biophysical and Electronic Engineering, University of Genoa, Via all'Opera Pia 11a, 16145 Genova, ITALY

Abstract

Despite the well known advantages that a space-variant representation of the visual signal offers, the required adaptation of the algorithms developed in the Cartesian domain before applying them in the log-polar space has limited a wide use of such representation in visual processing applications. Here, we present a set of original rules for designing a discrete log-polar mapping in order to directly apply, without modification, the standard algorithms based on spatial multi-scale and multi-orientation filtering in the log-polar domain, and effectively take advantage of the space-variance and of the data reduction. Such rules are based on a quantitative analysis of the relationships between the spatial filtering and the space-variant representation. We assess the devised rules by using a distributed approach based on a bank of Gabor filters to compute reliable disparity maps, by providing quantitative measures of the computational load and of the accuracy of the computed visual features.

Key words: Log-polar mapping, Gabor filtering, Design criteria, Active vision, Disparity computation

1. Introduction

Inspired by the retina of mammals, characterized by a decreasing of the photo-receptors from the center of the visual field (fovea) towards the periphery (Schwartz, 1977), the log-polar imaging is now a well established paradigm for simplifying a wide number of computational problems in pattern recognition and active vision (see (Berton et al., 2006; Traver and Bernardino, 2010; Yeasin and Sharma, 2005) for reviews). The log-polar mapping simultaneously provides a wide field-of-view, high spatial resolution on the region of interest, and a significant data reduction. All these features are well suitable for active vision applications (Aloimonos et al., 1988; Schwartz et al., 1995), since the visual systems continuously interact with the environment, by purposefully moving the eyes, to bring the interesting objects into the foveas (Bernardino and Santos-Victor, 1998). In such applications, the necessary real-time visual data processing is facilitated by the compression obtained by the mapping. At the same time, the log-polar mapping guarantees useful properties for pattern recognition problems (Wilson and Hodgson, 1992), such as rotation and scaling invariance.

In the literature, many approaches to directly solve image processing and image understanding tasks for space-variant representation of the visual signal have been described (Fischl et al., 1997; Nattel and Yeshurun, 2002; Smeraldi and Bigun, 2002; Traver and Pla, 2003; Wallace and McLaren, 2003). Although, in theory, the conformal mapping should permit a direct application of the visual operators developed for Cartesian images to log-polar ones, these authors discuss the necessity of properly adapting the algorithms before applying them on the space-variant images. Nevertheless, the extraction of visual features based on multi-orientation and multi-scale spatial filtering (Bigun, 2006; Granlund and Knutsson, 1995) has not been explicitly addressed yet.

In this paper, the relationships between the different parameters of a discrete log-polar mapping and of a bank of multi-scale and multi-orientation band-pass filters are analyzed, with the aim of demonstrating that a proper choice of such parameters allow us to directly use the algorithmic solutions developed for the Cartesian domain on log-polar images, without any modification. It is worth noting that the inherent space-variance of log-polar mapping is exploited to properly cope with the multi-scale issue. The validity of the devised design strategies are proved with reference to the computation of binocular disparity through a distributed phase-based algorithm (Chessa et al., 2009a), previously developed for the Cartesian domain.

2. Log-polar blind-spot model

In the literature, several log-polar mapping models are described (Bolduc and Levine, 1998; Florack, 2007; Jurie, 1999). In this paper, the central blind-spot model is chosen (Traver and Pla, 2008). The log-polar transformation $\mathbf{T} : (x, y) \mapsto (\xi, \theta)$, from the Cartesian domain to the cortical domain, can be backwards expressed in the following way:

$$\begin{cases} x = \rho_0 a^{\xi} \cos\theta \\ y = \rho_0 a^{\xi} \sin\theta \end{cases} \tag{1}$$

where a represents the base of the non-linearity of the mapping, ρ_0 is the radius of the blind spot and $(\rho, \eta) = (\sqrt{x^2 + y^2}, \arctan(y/x))$ are the usual polar coordinates.

To deal with digital images discrete coordinates have to be considered. Given a Cartesian image of $(m \times n)$ pixels, it is transformed into a cortical image of R rings and S sectors, where the discrete log-polar coordinates are denoted by (u, v). Thus, the growth rate of the size of the receptive fields between two consecutive rings (see Figure 1) can be expressed as $a = \exp(\ln(\rho_{max}/\rho_0)/R)$, where $\rho_{max} = \frac{1}{2}\min(m, n)$.

Figure 1 shows the log-polar receptive fields superimposed to the Cartesian domain and the cortical domain. The red circle, with radius $S/2\pi$, represents the locus where the size of log-polar pixels is equal to the size of Cartesian pixels. In particular, in the area inside the red circle a single Cartesian pixel contributes to many log-polar pixels (oversampling), whereas outside this region many Cartesian pixels contribute to a single log-polar pixel, thus avoiding the aliasing due to the undersampling (Jerry, 1977). This is pointed out in the receptive field bordered in violet in Figure 1.

Other important parameters of the log-polar mapping must be defined, in order to highlight properties of the transformation, such as the aspect ratio of

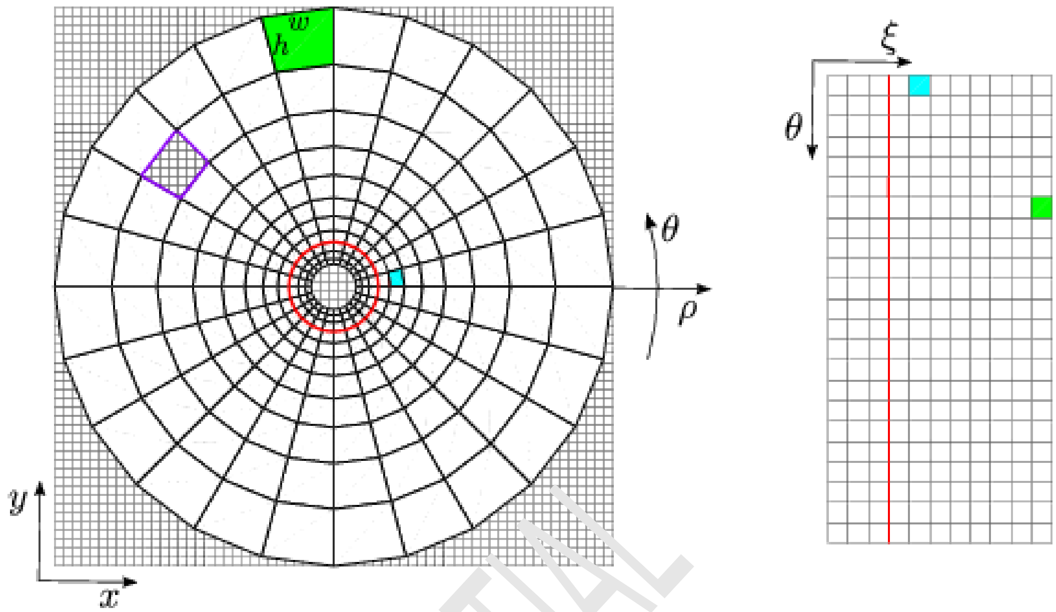


Figure 1: Cartesian domain with the superposition of the log-polar receptive fields (left) and cortical domain (right). The green and the cyan areas represent two receptive fields at different angular and radial positions (thus with different size w and h) that are mapped in the two corresponding cortical pixels. The red circle delimits the oversampling and undersampling areas.

the log-polar pixel γ, i.e. the ratio between its width $w = \frac{2\pi}{S}\rho_0 a^{u-1}$ and its height $h = \rho_0 a^{u-1}(a-1)$. In the following, the importance of the parameter γ for the visual processing will be analyzed.

3. Design rules of the log-polar mapping and of the filters for the extraction of visual features

In general, image feature extraction, based on spatial filtering, has two main drawbacks: the computational load of the filtering stage and the necessity of exploiting a multi-scale approach. The log-polar mapping intrinsically mitigates these issues, since the input image is compressed and a space-variant processing is obtained. Thus, the direct extraction of the features in the cortical domain has a lower computational load and intrinsically performs a multi-scale processing, as a function of the cortical location. To "optimally" design the log-polar mapping for visual processing tasks, it is important to study the relationships between the usual processing in the Cartesian domain and the direct extraction of the features in the cortical domain, by characterizing the filters with respect to the different parameters of the log-polar mapping. In particular, we consider Gabor filters (Daugman, 1985; Gabor, 1946), since they minimize the joint uncertainty in both the spatial and the frequency domain. The filters are normalized by their energy and can be expressed as:

$$g(x, y; \phi, \alpha) = \frac{1}{\sqrt{\pi}\sigma} \exp\left(-\frac{x_\alpha^2 + y_\alpha^2}{2\sigma^2}\right) \exp(j\omega_0 x_\alpha + \phi), \tag{2}$$

where σ determines the spatial support of the filter, ω_0 is the spatial peak tuning frequency, ϕ is the phase of the sinusoidal modulation and (x_α, y_α) are the rotated spatial coordinates in the Cartesian domain. Analogously, the Gabor filter can be directly defined in the cortical domain $g(\xi, \theta)$, or we can consider a filter mapped into the cortical domain $g(x(\xi, \theta), y(\xi, \theta))$. It is worth noting that, due to the non-linearity of the log-polar mapping, the mapped filters are distorted (Mallot et al., 1990; Wallace and McLaren, 2003). Thus, a filtering operation directly in the cortical domain could introduce undesired distortions in the filter outputs. To analyze this issue, we consider the response E of the filter $g(\xi, \theta)$ to the signal $s(\xi, \theta)$, that can be expressed by the inner product $E = \langle g(\xi, \theta), s(\xi, \theta)\rangle$. Specifically, to characterize the filtering operations we consider the response of a filter to a mapped and to a matched filter. The response for a mapped filter is:

$$E_{mapped} = \langle g(\xi, \theta), g(x(\xi, \theta), y(\xi, \theta))\rangle, \tag{3}$$

whereas the response for the matched filter is:

$$E_{matched} = \langle g(\xi, \theta), g(\xi, \theta)\rangle. \tag{4}$$

A filtering in the cortical domain results in a space-variant filtering operation in the Cartesian domain, where both the scale and the orientation of the filters vary. To guarantee a proper multi-orientation and multi-scale processing, we have to verify in which conditions the distortion of the mapped filters are minimal.

3.1. Response of a single filter as a function of the cortical location

In order to exploit the advantages provided by a space-variant processing, it is necessary that the filtering operations perform a uniform feature extraction, without introducing undesired anisotropies in the parametric space, thus allowing a direct application of the spatial filtering in the cortical domain, without specific modifications.

The specific visual feature extraction we are addressing constrains the choice of the parameters ρ_0, ρ_{max} and R. Once fixed these parameters, an analysis[1] of the influence of the parameters of the log-polar mapping on the response E_{mapped} of the Gabor filters is shown in Figure 2, to take into account the joint effects of the spatial support and orientation α of the Gabor filters and of the position (ξ_0, θ_0) in the cortical plane for two different aspect ratios γ of the log-polar pixel. The responses of the differently oriented filters (colored profiles in the polar plots) for an aspect ratio $\gamma = 1$ (first row) are compared to the responses obtained with $\gamma > 1$ (second row). The different colors in the polar plots represent different spatial supports of the filters. It is worth noting the anisotropy of the responses when the log-polar pixel is not squared: the responses are highly influenced both by the orientation α of the filter, and by the position (ξ_0, θ_0) in the cortical plane. For an aspect ratio $\gamma = 1$ the spatial support of the filters slightly affects the responses by lowering them without introducing any anisotropy to the responses.

3.2. Response of a single filter as a function of the parameters of the mapping

A further analysis to systematically investigate how the energy ratio between the response E_{mapped} of a mapped filter and the response $E_{matched}$ of the matched filter, is affected by the relationships between the parameters of the

[1] It is worth noting that the analysis of the parameters, presented in this paper, has been verified for different Cartesian image size ($m \times n$) and for different cortical image size ($R \times S$). Moreover, the real and imaginary parts of the Gabor filters have been considered both separately and jointly.

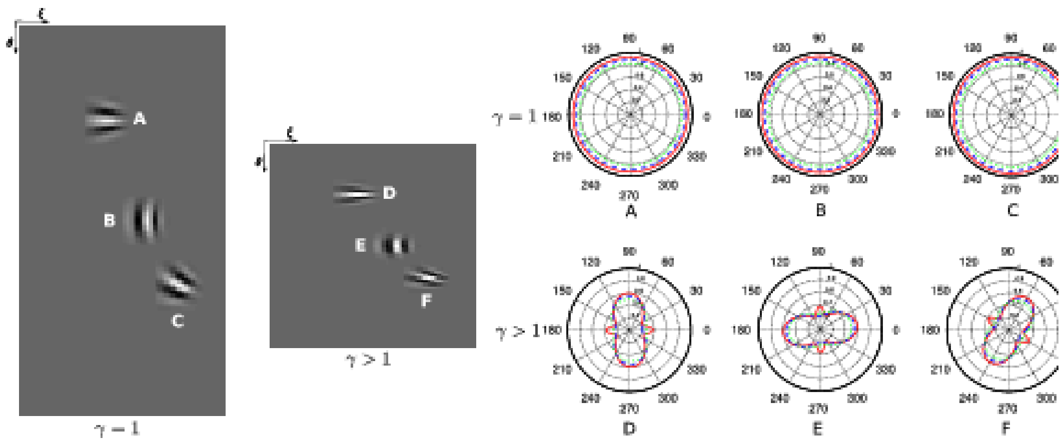


Figure 2: (Left) Mapped filters, with orientation $\alpha = 0$ in the Cartesian domain and constant spatial support in the cortical plane for two values of the aspect ratio γ. (Right) The polar plots show the responses (E_{mapped}) of the Gabor filters as a function of different orientations α. Three different positions in the cortical plane have been considered: A-C for $\gamma = 1$ D-F for $\gamma > 1$. For each of the three positions in the cortical plane and for the two aspect ratios three different Cartesian spatial supports have been considered: 11×11, 21×21 and 31×31 pixels, whose responses are plotted in red solid, dashed blue and dotted green, respectively. The black thick line corresponds to the energy of the corresponding matched filter.

log-polar mapping and of the Gabor filter is shown in Figure 3. Each subfigure shows the variation of the energy ratio $E_{mapped}/E_{matched}$ with respect to pairs of parameters of the mapping and of the filters (left side) and the profile of the mapped filters for four different combinations of such parameters (right side). If the aspect ratio of the log-polar pixel is approximately 1, the energy ratio $E_{mapped}/E_{matched}$ remains high, independently of the eccentricity ξ_0 in the cortical plane and of the orientation α of the Gabor filter (see Figure 3a-b). Conversely, values of γ different from 1 yield to lower responses of the filters with respect to the eccentricity ξ_0 and to an anisotropy with respect to the orientation α of the filter. Moreover, Figure 3c shows that the maximum response is obtained when the spatial support of the filter is small (e.g. 11×11 pixels). It is worth noting that under these conditions the deformations of the mapped filters are relatively small (see inset A of Figure 3c). Once fixed $\gamma = 1$, the influence of the spatial support of the filter can be evidenced from Figure 3d-e. The response of the filters decreases with an increase of the spatial support, independently of the eccentricity in the cortical plane and of the orientation α of the filter. This can be also evidenced from the deformed profiles of the Gabor filters (see Figure 3d-e, profiles marked by D and C). For a given value of the spatial support (e.g. 11×11 pixels) the responses of the filters neither depend on the eccentricity in the cortical plane nor on the orientation of the filter (see Figure 3f).

In this Section, we have devised the constraints for the parameters of the log-polar mapping and of the spatial filters, in order to obtain a signal processing in the cortical domain equivalent to the one in the Cartesian domain. It is worth noting that in (Traver and Pla, 2008) the authors state that a log-polar pixel with aspect ratio equals to 1 is necessary to correctly compute the gradient orientation. The analysis conducted here shows that this rule can be generalized in order to efficiently use the Gabor filters as *local jets* to measure important elements of the visual signal (Adelson and Bergen, 1991; Fleet and Jepson, 1990; Fleet et al., 1991; Haglund, 1992; Koenderink and van Doorn, 1987).

4. Feature extraction through a bank of filters

In this Section, we address the problem of extracting visual features from the responses of a direct filtering in the cortical domain. In particular, we consider the computation, for each orientation, of the local phase in the image signal. Performing the phase measurement directly in the cortical domain requires the verification of the quadrature conditions for the real and the imaginary part of the Gabor-like mapped filters (if one adopts a direct measure of the local phase), or a uniform preservation of the phase of the filters (if one adopts a distributed representation of the local phase). Here, the latter approach is considered. Though, it is worth noting that the following analysis is not limited to the distributed approach, since the equivalence between direct phase measurement and energy distributed models has been demonstrated (Qian and Mikaelian, 2000).

In particular, to perform such analysis, we analyze the different responses E_i

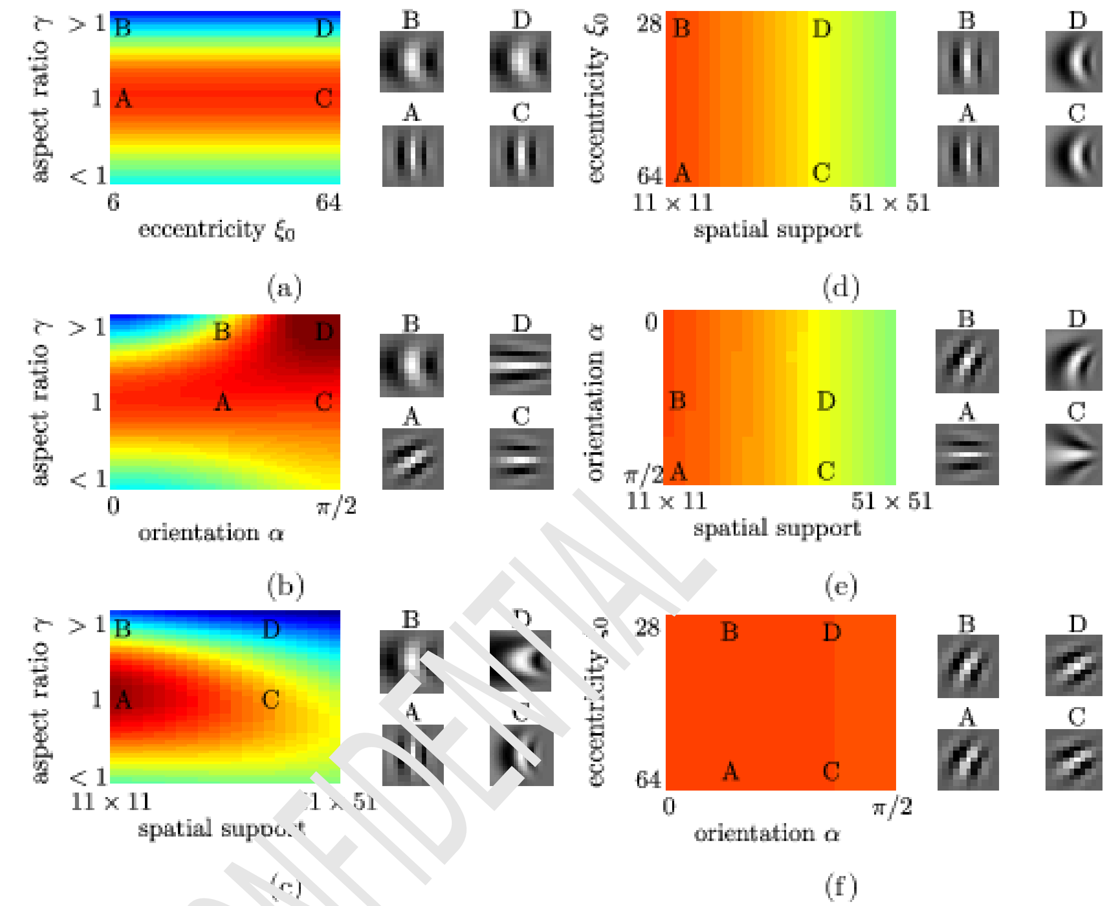
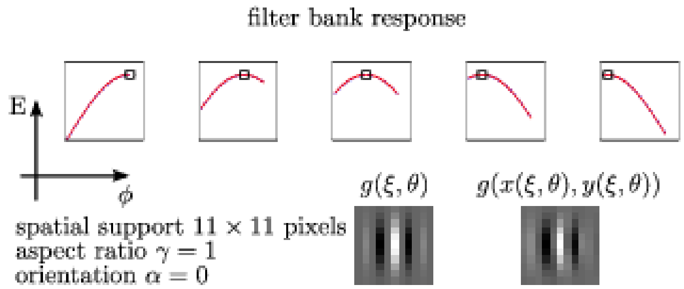


Figure 3: Variation of the energy ratio $E_{mapped}/E_{matched}$ with respect to the parameters of the log-polar mapping and of the Gabor filters (left side of each subfigure) and profiles of the mapped filters for particular choices of such parameters, marked by capital letters A-D (right side of each subfigure). Hot colors mean high energy ratios, whereas cold colors mean low energy ratios. (a) Aspect ratio of the log-polar pixel γ with respect to the eccentricity ξ_0 in the cortical plane. The maximum energy ratio is obtained for a squared log-polar pixel independently of the eccentricity in the cortical plane. (b) Aspect ratio of the log-polar pixel γ with respect to the orientation α of the Gabor filter. The energy is constant independently of the orientation of the filter when the aspect ratio is $\gamma = 1$. (c) Aspect ratio of the log-polar pixel γ with respect to the spatial support of the Gabor filters. The maximum energy ratio is obtained for a squared log-polar pixel ($\gamma = 1$) and a small spatial support (11×11 pixels). It is worth noting that under these conditions the filters show no deformation (see A), otherwise high deformations are present (see B-C-D). (d) Eccentricity ξ_0 in the cortical plane with respect to the spatial support of the filter. The maximum energy is obtained for a small spatial support. (e) Orientation α with respect to the spatial support of the Gabor filter. The maximum energy ratio is obtained for a small spatial support. (f) Eccentricity ξ_0 in the cortical plane with respect to the orientation α of the filter. The energy is constant independently of the values and the filters present only small deformation (see A-B-C-D).

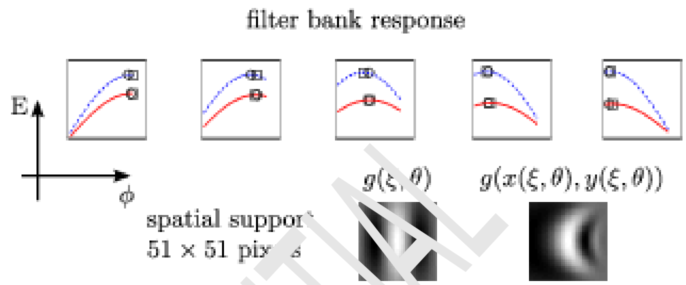
of a bank of Gabor filters, each characterized by a different value of the phase ϕ_i, for a given phase ϕ^{IN} of a filter considered as the input signal. It is worth noting that, a reliable detection of the input phase can be obtained when the peak of the responses E_i occurs for the value ϕ^{IN} of the input signal and when the shape of the response curve is bell-shaped and symmetric with respect to its peak.

Figure 4 describes how the filter bank response is affected by the spatial support of the filter, the aspect ratio γ of the log-polar pixel, and the orientation α of the filter, respectively. Since the stability of the phase-based approach has been demonstrated (Fleet and Jepson, 1993), the response of the filter bank for "non-optimal" choices of the parameters is again analyzed. For each subfigure the comparison between the responses of a bank of matched Gabor filters (dashed blue profiles) and those of a bank of mapped filters (solid red profiles), for five different values of the phase ϕ^{IN}, is shown. The peak of the responses E_i and the value of the phase input are marked by a circle and a square, respectively. Furthermore, the filter $g(\xi, \theta)$ and the corresponding mapped filter $g(x(\xi, \theta), y(\xi, \theta))$ are shown.

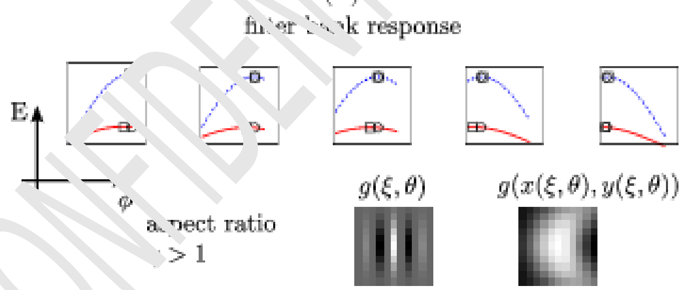
Figure 4a shows the filter bank response for different phase values ϕ^{IN} in the input signal for a bank of filters with the reference set of parameters: spatial support equal to 11×11 pixels, aspect ratio $\gamma = 1$ and filter orientation $\alpha = 0$. Figure 4b shows the filter bank response with a spatial support equal to 51×51 pixels. For a small spatial support the response of a bank of Gabor filter and the one of a bank of mapped filter is similar, with the peak response coincident with the input phase value. The aspect ratio γ of the log-polar pixel also affects the filter bank response (not only the response E of a single filter). Figure 4c shows how the peak of the response does not coincide with the input phase ϕ^{IN}. Moreover the profile of the mapped filter shows a significant deformation. Finally, if the spatial support of the filter is small and the aspect ratio of the log-polar pixel is equal to 1, the orientation of the filter does not affect the filter bank response (see Figure 4d).



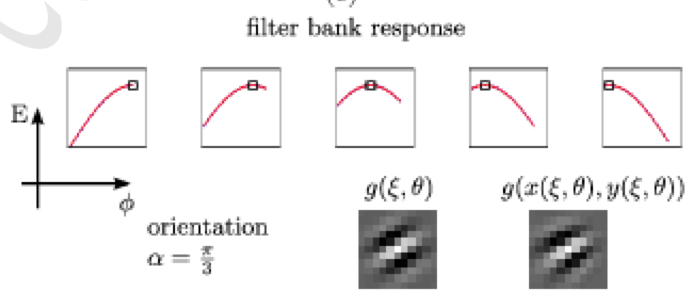
(a)



(b)



(c)



(d)

Figure 4: Comparison between the responses of a bank of matched filters (dashed blue profiles) and those of a bank of mapped filters (solid red profiles), for five different values of the phase ϕ^{IN}. The peak of the responses E_i and the value of the phase input are marked by a circle and a square, respectively. The filter $g(\xi,\theta)$ and the corresponding mapped filter $g(x(\xi,\theta),y(\xi,\theta))$ are shown. (a) The response of the Gabor filters for the reference set of parameters: spatial support 11×11 pixels, aspect ratio $\gamma = 1$ and orientation $\alpha = 0$. (b) The response of the filter bank when the spatial support of the Gabor filters is 51×51 pixels. (c) The response of the filter bank when the aspect ratio γ of the log polar pixel is larger than 1. (d) The response of the filter bank when the Gabor filter orientation α is $\pi/3$.

5. Remarks about the log-polar mapping for the visual processing

5.1. Fovea design

In the literature, different techniques to handle the information in the blind spot have been proposed. A Cartesian representation in the foveal region can be used (Bolduc and Levine, 1998), although it is necessary to take into account the discontinuity between the central area and the peripheral area. An alternative representation it is to consider a polar fovea (Berton et al., 2006). However, the polar fovea model has the major drawback of generating an aspect ratio γ that changes with the distance from the origin[2], thus it is not useful for the visual processing, as we have demonstrated in Section 3. It is also worth noting that the size of the blind spot is usually small if compared with the spatial support of the filter, thus the problem associated with the discontinuity issue diminishes.

5.2. Multi-scale analysis

To optimally detect different features at different levels of resolution in the input image, a multi-scale approach is necessary. Considering that Gabor filters act in the frequency domain as band-pass filters, centered at their own spatial peak frequency, whereas information in natural images is spread on a wide range of frequencies, it is necessary to use a technique that allow to capture information from the whole range. In general, a multiresolution analysis can be efficiently implemented through a coarse-to-fine strategy that allows us to recover feature values larger than the spatial support of the filter. Thus, the number of spatial scales depends on the specific processing task addressed. The space variance of the log-polar mapping, i.e. the linearly increase of the filter size with respect to the eccentricity, can be exploited to efficiently implement a multi-scale analysis. A pyramidal approach (Burt and Adelson, 1983) can be considered as a "vertical" multi-scale, i.e. the variation of the filter size

[2] In the polar region of the mapping the aspect ratio is $\gamma = (k_p 2\pi/S)u$, where k_p is a constant that takes into account the continuity condition between the polar and the log-polar mapping, and u represents the eccentricity.

at a single location, whereas the log-polar spatial sampling acts as an "horizontal" multi-scale, i.e. the variation of the filter size across different location (Schwartz, 1985; Bonmassar and Schwartz, 1997). The "vertical" multi-scale is also addressed in the literature as "cortical pyramids" (Colombo et al., 1996).

To exploit the "horizontal" multi-scale properties, an additional rule to design the log-polar mapping is introduced, in order to take into account the desired filter size at the maximum eccentricity, and, consequently, the maximum log-polar pixel size W_{max}. Consequently, it is necessary to devise how the parameters of the log-polar mapping can be expressed as a function of W_{max}. The novel rule that relates the total number of rings R with W_{max} can be expressed by:

$$R = -\frac{\ln(\rho_{max}/\rho_0)}{\ln((\rho_{max} - W_{max})/\rho_{max})}. \tag{5}$$

Hence, the log-polar mapping is defined by three parameters: ρ_0, ρ_{max} and W_{max}, by assuming the aspect ratio of the log-polar pixels $\gamma = 1$.

5.3. Vector feature mapping

In computer vision, important visual features, such as the optic flow and the disparity for a stereo active vision system with convergent axes, are described by vector fields. Since the visual features ($\mathbf{d}$) are computed in the cortical domain, the transformation of a vector field from the (ξ, θ) domain to the (x, y) domain can be expressed in terms of general coordinates transformation (Chan Man Fong et al., 1997):

$$\begin{bmatrix} d_x \\ d_y \end{bmatrix} = \begin{bmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial \xi} & \frac{\partial y}{\partial \theta} \end{bmatrix} \begin{bmatrix} d_\xi \\ d_\theta \end{bmatrix}. \tag{6}$$

Combining Eqs. 1 and 6, we obtain:

$$\begin{bmatrix} d_x \\ d_y \end{bmatrix} = \rho_0 a^\xi \ln(a) \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} d_\xi \\ d_\theta \end{bmatrix}. \tag{7}$$

The scalar coefficient of Eq.7 represents the scale factor of the log-polar vector and the matrix describes the rotation due to the mapping.

6. Experimental validation

In this Section the analyzed design rules are assessed by using a multi-scale and multi-orientation approach for the extraction of visual features. In particular, a distributed phase-based algorithm for vector disparity evaluation has been considered (see Appendix A for details).

6.1. Disparity computation in log-polar images

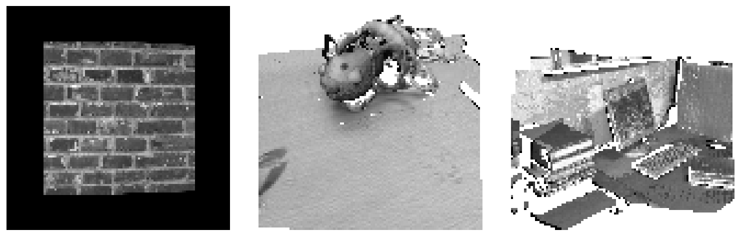
The issue of disparity estimation for log-polar foveated systems has been addressed by several authors in the literature (Bernardino and Santos-Victor, 1996, 2002; Grosso and Tistarelli, 2000; Manzotti et al., 2001; Schindler, 2006). In (Grosso and Tistarelli, 2000) sparse disparity maps are obtained by using a matching of Laplacian features. Bernardino and Santos-Victor (1996) and Manzotti et al. (2001) present techniques for extracting a disparity measure for vergence control, thus neglecting the contribution of the vertical disparity. In (Bernardino and Santos-Victor, 2002) the authors follow a Bayesian approach to estimate both horizontal and vertical disparities. However, the lack of quantitative results prevents an explicit comparison with our approach.

The algorithm used for the experimental validation presented in this paper is suitable to be directly applied on cortical images, since 2D vector disparity is computed without an explicit search of the correspondences, between the left and the right images, along the epipolar lines. In this way, it is not necessary to take into account that the straight lines in the Cartesian domain become curves in the log-polar space (Schindler, 2006).

6.2. Results

The 2D vector disparity is computed for stereo image pairs acquired by an active vision system: the two cameras of the system can actively fixate points in the 3D workspace through vergence and version movements. In order to quantitatively benchmark the proposed approach, stereo sets with available ground truth disparities are necessary. To this aim, the tool described in (Chessa et al., 2009b) is used. Such a tool, exploiting the ground truth available from a

3D model of the observed scene, virtual or real, and the related projected stereo images, provides a way to validate the behavior of an active vision system in a controlled and realistic scenario. Figure 5 shows the left image of the stereo pairs used in the following analysis.



(a) (b) (c)

Figure 5: Left images of the considered stereo pairs. (a) A synthetic frontoparallel plane. (b)-(c) Real-world scenarios acquired by a 3D laser scanner.

The visual task, we are considering, constraints the choice of the blind spot radius and the number of scales. Since in active vision systems information in the fovea is the most important, the ρ_0 is kept small, i.e. in the range between 3 and 1[illegible] pixels. Moreover, since the presence of large disparities in the periphery affects the number of scales, both "horizontal" multi-scale, see Eq. 5, and "vertical" multi scale are used. The values of W_{max} range between 4 and 8, and the number of "vertical" scales is chosen between 1 and 3.

According to these choices, we first transform the stereo image pair into the cortical domain, then the 2D vector disparity is computed in the cortical domain (ξ, θ) by using a bank of Gabor filters with spatial support 11×11 pixels, peak frequency $\omega_0 = 1/4$, bandwidth 0.833 octave and 8 different orientations α. Finally, the vector disparity is transformed into the Cartesian domain (x, y) by using Eq. 7, where we perform the quantitative benchmark with respect to the ground truth data.

Figure 6 shows the resulting estimate of the horizontal and vertical disparities for a frontoparallel plane, with the camera axes vergent in the center of the plane. Figure 6 (first row) shows the disparities computed in the cortical domain, by using two "vertical" scales. It is worth noting that two "vertical" spatial scales are not sufficient to recover the correct disparity range if the bank of filters is applied into the Cartesian domain, directly (see Figure 6 (second row)).

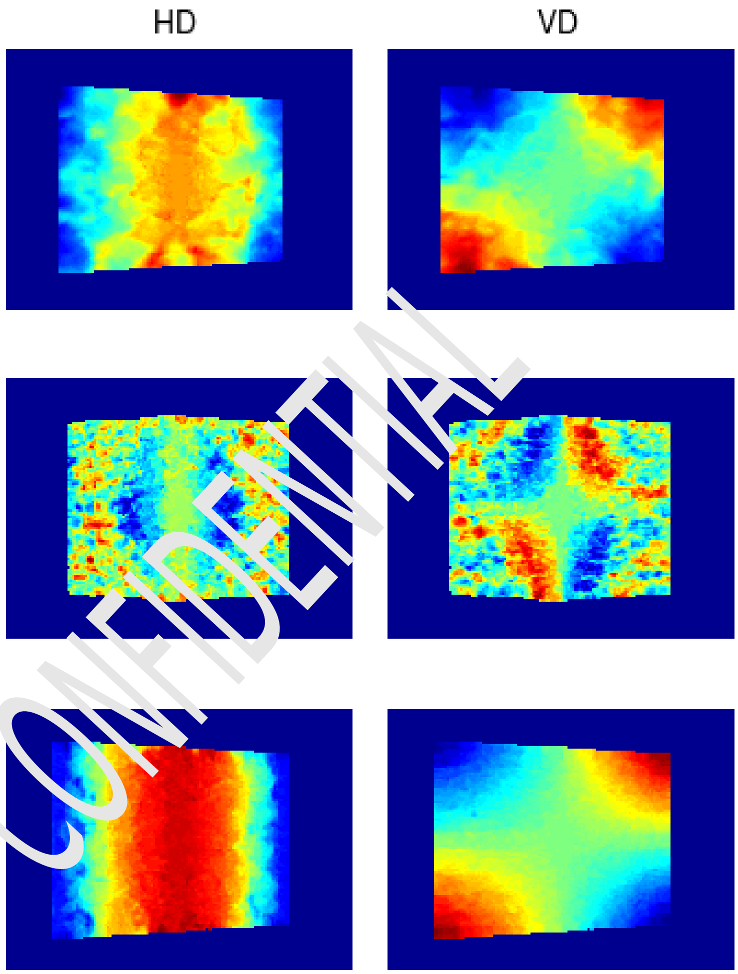


Figure 6: Horizontal (HD) and vertical (VD) disparities for a frontoparallel plane, when the optical axes are vergent in the center of the plane. (First row) Computation in the cortical domain with two "vertical" scales. (Second row) Computation in the Cartesian domain with two "vertical" scales and with five "vertical" scales (third row).

In order to quantitatively analyze the reliability of the results, the computed disparity maps are compared to the available ground truth maps. Figures 7, 8 and 9 show the computed disparity maps obtained from stereo pairs representing a plane and two more complex real-world scenes, acquired by a laser scanner, respectively. Furthermore, the reliability of the disparity values with respect to

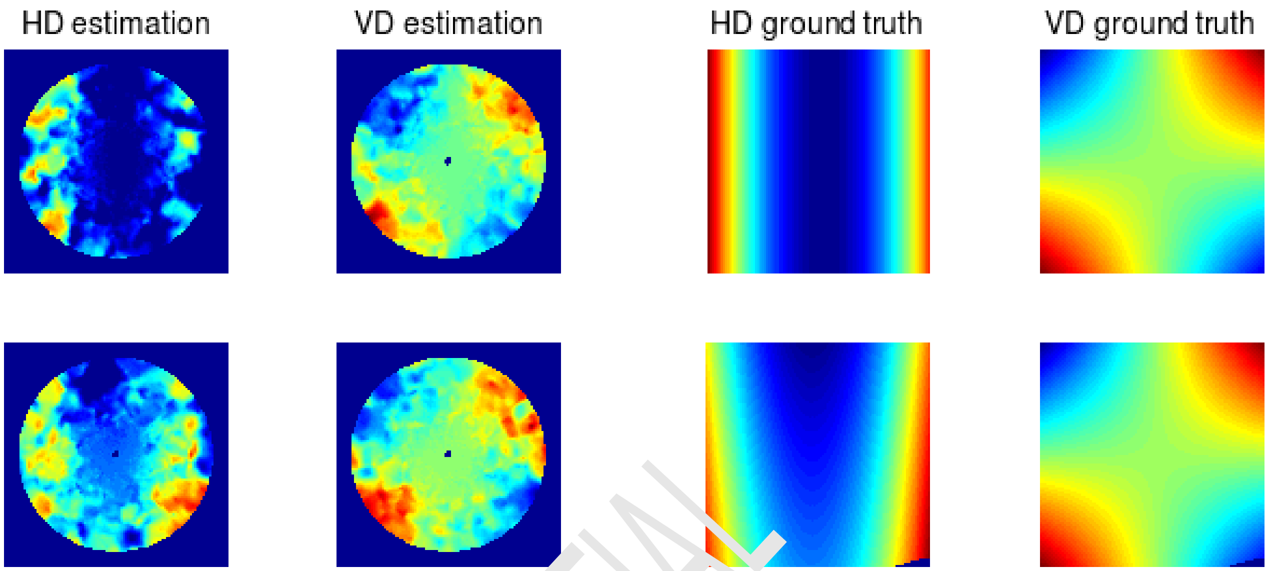


Figure 7: Comparison between the estimated disparity maps and the ground truth for a frontoparallel plane (first row) and for a slanted plane (second row). In both situations the axes of the two cameras are vergent in the center of the planes. The average errors in the computation of the horizontal (HD) and vertical (VD) disparities are: 1.72 and 0.69 pixels for the first row, and 1.71 and 0.79 pixels for the second row. The ground truth disparity range is between −16 and 16 pixels.

the parameters of the mapping is analyzed. Tables 1 and 2 show how the size of the cortical image, defined by ρ_0 and W_{max}, and the number of the considered "vertical" scales affect both the execution time, and the global average error on the computation of the disparities, with respect to the ground truth. In addition to the global average error, computed by considering all the pixels of the image, the average error around the fovea (a region with a radius half of the image size) and in the periphery is computed separately. This approach is necessary, since the central part of the image is mostly important for active vision tasks and the error in the peripheral area is affected by the increased size of the log-polar pixel. The analysis show that the average error in the region around the fovea is small, i.e. less than 1 pixel in every condition.

The execution time is expressed as a fraction of the algorithm execution time in the Cartesian domain with the optimum set of parameters, in this way the

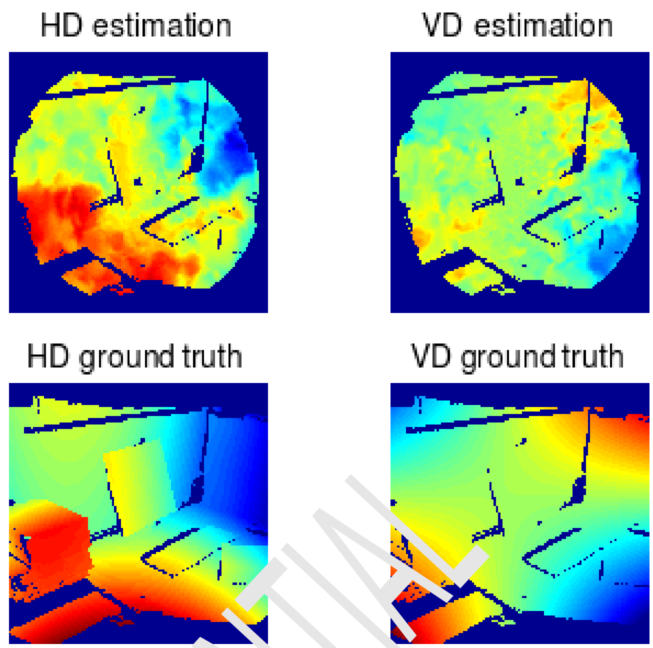


Figure 8: Comparison between the estimated disparity maps (first row) and the ground truth (second row) for a stereo pair obtained from a real scenario acquired by a laser scanner. The average errors in the computation of the horizontal (HD) and vertical (VD) disparities are 1.50 and 0.57 pixels, respectively. The ground truth disparity range is between -13 and 21 pixels.

obtained results are not bound to a specific implementation. It is worth noting that the time necessary for the forward and backward log-polar transformation is a small percentage of the total execution time.

7. Conclusion

In this paper, we have addressed the problem of the multi-orientation and multi-scale filtering in the log-polar domain. The extraction of features based on spatial filtering has a great importance for many applications of image processing and computer vision. Nevertheless, this topic has not been fully investigated in the literature. To this aim, a systematic analysis of the relationships between the parameters of the discrete log-polar mapping and of a bank of Gabor filters has been carried out. The major outcome of this analysis is the definition of a set of general design rules, that allow us to use algorithms, which were

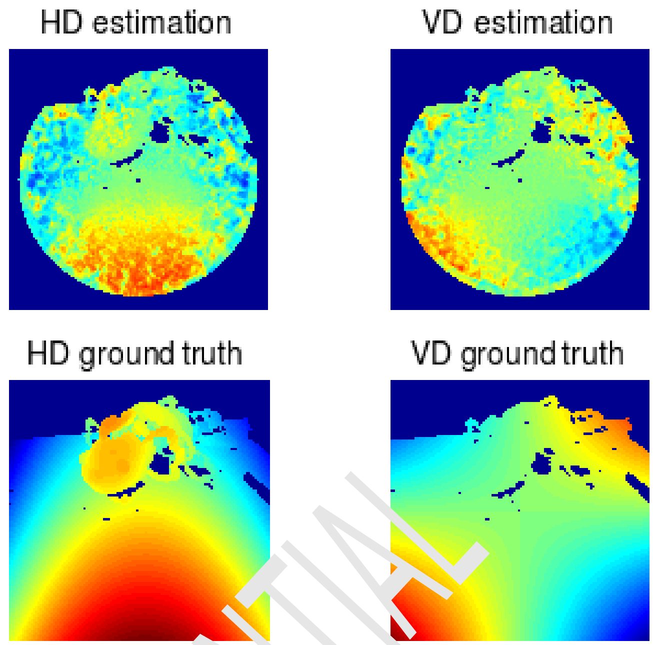


Figure 9: Comparison between the estimated disparity maps (first row) and the ground truth (second row) for a stereo pair obtained from a real scenario acquired by a laser scanner. The average errors in the computation of the horizontal (HD) and vertical (VD) disparities are 1.29 and 0.54 pixels, respectively. The ground truth disparity range is between -6 and 10 pixels.

originally designed in the Cartesian domain, directly in the log-polar space, without requiring specific modifications. Moreover, we have deduced a novel rule to efficiently implement a multi-scale analysis, by exploiting the space-variance of the log-polar mapping.

The validity of such analysis has been proved by applying a distributed phase-based approach for the computation of binocular disparity based on a bank of Gabor filters on log-polar stereo pairs. The obtained results show that it is possible to recover reliable values of the horizontal and vertical disparities by directly applying the algorithm in the cortical domain, thus achieving a consistent reduction in the execution time.

The possibility of efficiently exploiting a space-variant representation is of great importance in the development of active systems capable of interacting with the environment, since a precise processing of the visual signal is possible in the foveal area, where the feature errors are small enough to allow a fine exploration of the object of interest. At the same time, the coarse computation of the feature in the peripheral area provides enough information to detect new saliencies and to bring the focus of attention there.

To allow a future quantitative comparison with the results presented in this paper, the stereo pairs and the ground truth data considered for the analysis are made publicly available at `www.pspc.dibe.unige.it/Research/vr.html`.

Acknowledgments

This work has been partially supported by EU FP7-ICT 217077 Project "EYESHOTS" and by EU FP7-ICT 215866 Project "SEARISE".

Appendix A. Distributed disparity computation

In this Appendix an approach to extract disparity from a sequence of stereo image pairs, using a distributed bio-inspired architecture that resorts to a population of tuned units, is described (Chessa et al., 2009a).

The population that has been used to compute the features is based on a bank of oriented Gabor filters (cf. Eq. 2), each having the same peak spatial frequency ω_0. Following the *phase-shift model* (Fleet et al., 1996), to obtain the tuning to a specific disparity, a pair of filters, $g^L(\mathbf{x})$ and $g^R(\mathbf{x})$, is applied in the same position $\mathbf{x}_0 = (x_0, y_0)$ of the left and the right images, respectively. The filter pair share the same properties, but are characterized by a phase difference $\Delta\phi = \phi^L - \phi^R$. For each spatial orientation α, a set of K binocular phase differences are chosen to obtain the tuning to different disparities. The sensitivity to binocular disparity is obtained from a quadrature pair of binocular energy units (Ohzawa et al., 1990; Fleet et al., 1996), described by :

$$E(\mathbf{x_0}; \alpha, d) = |E^L(\mathbf{x_0}; \alpha) + E^R(\mathbf{x_0}; \alpha)|^2,$$

where $E^L(\mathbf{x_0}; \alpha) = \langle g^L(\mathbf{x} - \mathbf{x}_0), I^L(\mathbf{x})\rangle$, $E^R(\mathbf{x_0}; \alpha) = \langle g^R(\mathbf{x} - \mathbf{x}_0), I^R(\mathbf{x})\rangle$ and $I^R(\mathbf{x}) = I^L(\mathbf{x} + \mathbf{d})$. The binocular energy $E(\mathbf{x_0}; \alpha, d)$ has its maximum when the product of the projection d, along the orientation α of the filter, of the stimulus disparity $\mathbf{d}$ and the spatial peak frequency ω_0 equals the binocular phase difference $\Delta\phi$: $d = \Delta\phi/\omega_o$. To obtain precise feature computation a

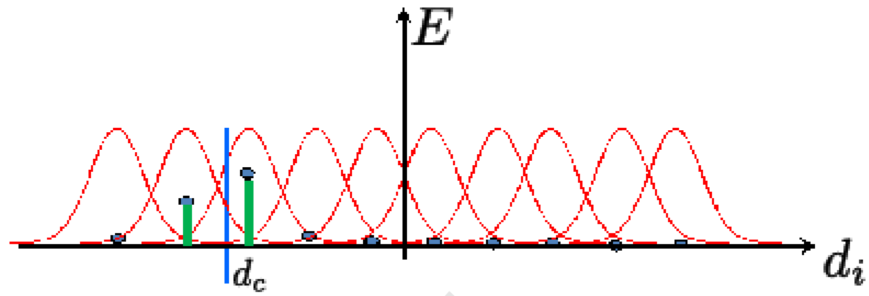


Figure A.10: Population decoding by a weighted sum of the population response. The tuning curves of each unit are represented in red, the responses $E(\mathbf{x_0};\alpha,d_i)$ are shown as the green bars and the decoded *component disparity* d_c is represented by the blue bar.

weighted sum (i.e. a population vector) of the responses for each orientation α is performed. The *component disparity* d_c is obtained by:

$$d_c(\mathbf{x_0};\alpha) = \frac{\sum_{i=1}^{K} d_i E(\mathbf{x_0};\alpha,d_i)}{\sum_{i=1}^{K} E(\mathbf{x_0};\alpha,d_i)},$$

where d_i are the K tuning disparities and $E(\mathbf{x_0};\alpha,d_i)$ are the binocular energies obtained for each spatial unit (see Figure A.10). Then, the vector disparity $\mathbf{d}$ is obtained trough a combination of the different component disparities. Finally, by exploiting the pyramidal approach a coarse-to-fine refinement is obtained. The feature values computed at a coarser level of the pyramid are used to warp the outputs of the filtering stage at the finer level, then the residual values of the feature are computed.

References

E. Schwartz, Spatial Mapping in the Primate Sensory Projection: Analytic Structure and Relevance to Perception, Biological Cybernetics 25 (1977) 181–194.

F. Berton, G. Sandini, G. Metta, Encyclopedia of Sensors, chap. Anthropomorphic visual sensors, American Scientific Publishers, 1–16, 2006.

V. Traver, A. Bernardino, A review of log-polar imaging for visual perception in robotics, Robotics and Autonomous Systems 58 (4) (2010) 378 – 398.

M. Yeasin, R. Sharma, Machine Learning and Robot Perception, chap. Foveated Vision Sensor and Image Processing A Review, Springer-Verlag, 57 – 98, 2005.

J. Aloimonos, I. Weiss, A. Bandyopadhyay, Active vision, International Journal of Computer Vision 1 (4) (1988) 333–356.

E. Schwartz, D. Greve, G. Bonmassar, Space-Variant Active Vision: Definition, Overview and Examples, Neural Networks 8 (7-8) (1995) 1297–1308.

A. Bernardino, J. Santos-Victor, Visual behaviours for binocular tracking, Robotics and Autonomous Systems 25 (3-4) (1998) 137–146.

J. Wilson, R. Hodgson, A pattern recognition system based on models of aspects of the human visual system, in: International Conference on Image Processing and its Applications, 258 –26[illegible], 1992.

B. Fischl, M. Cohen, E. Schwartz, The local structure of space-variant images, Neural Networks 10 (5) (1997) 815 – 831.

E. Nattel, Y. Yeshurun, Direct feature extraction in a foveated environment, Pattern Recognition Letters 23 (13) (2002) 1537–1548.

F. Smeraldi, J. Bigun, Retinal Vision applied to Facial Features Detection and Face Authentication, Pattern Recognition Letters 23 (2002) 463–475.

V. Traver, F. Pla, Dealing with 2D translation estimation in log-polar imagery, Image Vision Comput. 21 (2) (2003) 145–160.

A. Wallace, D. McLaren, Gradient detection in discrete log-polar images, Pattern Recognition Letters 24 (14) (2003) 2463–2470.

J. Bigun, Vision with Directions: A Systematic Introduction to Image Processing and Computer Vision, Springer-Verlag, Berlin Heidelberg, 2006.

G. Granlund, H. Knutsson, Signal Processing for Computer Vision, Kluwer Academic Publishers, Dordrecht, 1995.

M. Chessa, S. Sabatini, F. Solari, A Fast Joint Bioinspired Algorithm for Optic Flow and Two-Dimensional Disparity Estimation, in: ICVS, 184–193, 2009a.

M. Bolduc, M. D. Levine, A Review of Biologically Motivated Space-Variant Data Reduction Models for Robotic Vision, Computer Vision and Image Understanding 69 (2) (1998) 170–184.

L. Florack, Modeling Foveal Vision, in: Scale Space and Variational Methods in Computer Vision 2007, 919–928, 2007.

F. Jurie, A new log-polar mapping for space variant imaging. - Application to face detection and tracking, Pattern Recognition 32 (1999) 865–875.

V. Traver, F. Pla, Log-polar mapping template design: From task-level requirements to geometry parameters, Image Vision Comput. 26 (10) (2008) 1354–1370.

A. Jerry, The Shannon sampling theorem - Its various extensions and applications: A tutorial review, Proceedings of the IEEE 65 (11) (1977) 1565–1596.

J. Daugman, Uncertainty Relation for Resolution in Space, Spatial Frequency, and Orientation Optimized by Two-Dimensional Visual Cortical Filters, J. Opt. Soc. Amer. A A/2 (1985) 1160–1169.

D. Gabor, Theory of Communication, J. Inst. Elec. Eng. 93 (1946) 429–459.

H. A. Mallot, W. Seelen, F. Giannakopoulos, Neural mapping and space-variant image processing, Neural Networks 3 (3) (1990) 245–263.

E. Adelson, J. Bergen, The Plenoptic and the Elements of Early Vision, in: M. Landy, J. Movshon (Eds.), Computational Models of Visual Processing, MIT Press, 3–20, 1991.

D. Fleet, A. Jepson, Computation of component Image Velocity from local phase information, International Journal of Computer Vision 1 (1990) 77–104.

D. Fleet, A. Jepson, M. Jenkin, Phase-Based Disparity Measurement, CVGIP: Image Understanding 53 (1991) 198–210.

L. Haglund, Adaptive Multidimensional Filtering, Ph.D. thesis, Linkoing University, Sweden, 1992.

J. Koenderink, A. van Doorn, Representation of local geometry in the visual system, Biol. Cybern. 55 (1987) 367–375.

N. Qian, S. Mikaelian, Relationship Between Phase and Energy Methods for Disparity Computation, Neural Comput. 12 (2) (2000) 279–292.

D. J. Fleet, A. D. Jepson, Stability of Phase Information, IEEE Trans. Pattern Anal. Mach. Intell. 15 (12) (1993) 1253–1268.

P. Burt, E. Adelson, The Laplacian Pyramid as a Compact Image Code, IEEE Trans. Commun. COM-31 (1983) 532–540.

E. Schwartz, Image Processing Simulations of the Functional Architecture of Primate Striate Cortex, Investigative Ophthalmic and Vision Research (Supplement) 26 (3) (1985) 164.

G. Bonmassar, E. Schwartz, Space-Variant Fourier Analysis: The Exponential Chirp Transform, IEEE Trans. Pattern Anal. Mach. Intell. 19 (10) (1997) 1080–1089.

C. Colombo, M. Rucci, P. Dario, Image Technology: Advances in Image Processing, Multimedia and Machine Vision, chap. Integrating selective attention and spacevariant sensing in machine vision, L.C. Sanz (Ed.), Springer, 109–127, 1996.

C. Chan Man Fong, D. Kee, P. Kaloni, Advanced Mathematics For Applied And Pure Sciences, Crc Press, 1997.

A. Bernardino, J. Santos-Victor, Vergence control for robotic heads using log-polar images, in: Intelligent Robots and Systems, 1264–1271, 1996.

A. Bernardino, J. Santos-Victor, A Binocular Stereo Algorithm for Log-Polar Foveated Systems, in: Biologically Motivated Computer Vision, 127–136, 2002.

E. Grosso, M. Tistarelli, Log-Polar Stereo for Anthropomorphic Robots, in: Computer Vision - ECCV 2000, 299–313, 2000.

R. Manzotti, A. Gasteratos, G. Metta, G. Sandini, Disparity Estimation on Log-Polar Images and Vergence Control, Computer Vision and Image Understanding 83 (2) (2001) 97–117.

K. Schindler, Geometry and construction of straight lines in log-polar images, Computer Vision and Image Understanding 103 (3) (2006) 196–207.

M. Chessa, F. Solari, S. Sabatini, A Virtual Reality Simulator for Active Stereo Vision Systems, in: VISSAPP, 444–449, 2009b.

D. Fleet, H. Wagner, D. Heeger, Neural Encoding of Binocular Disparity: Energy Models, Position Shifts and Phase Shifts, Vision Research 36 (12) (1996) 1839–1857.

I. Ohzawa, G. DeAngelis, R. Freeman, Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors, Science 249 (1990) 1037–1041.

CARTESIAN DOMAIN								
image size ($m \times n$)	number of scales	AEH	AEV	AEH fovea	AEV fovea	AEH periphery	AEV periphery	execution time
331×331	5	0.82	0.27	0.10	0.10	0.83	0.28	100%
331×331	2	3.50	2.43	0.30	0.21	3.59	2.68	89%
331×331	1	3.73	2.74	0.76	0.54	3.77	3.09	67%

CORTICAL DOMAIN								
($R \times S$) ρ_0, W_{max}	number of scales	AEH	AEV	AEH fovea	AEV fovea	AEH periphery	AEV periphery	execution time
100×159 $3, 5$	2	1.37	0.45	[illegible]	0.29	1.53	0.61	36%
100×184 $5, 5$	2	1.21	[illegible]	0.28	0.22	1.47	0.57	36%
100×159 $3, 5$	1	1.92	0.64	0.43	0.35	2.09	0.83	29%
100×184 $5, 5$	1	[illegible]	0.71	0.46	0.29	2.07	1.01	20%
64×117 $5, 7$	1	2.13	0.84	0.50	0.34	2.23	1.04	9%

Table 1: Performance comparison between the computation of the disparity in the Cartesian and in the log-polar domain for different sizes of the cortical image. The values refer to the test image shown in Figure 5a. The global average error for the horizontal disparity (AEH) and for the vertical disparity (AEV), together with the local error around the fovea (AEH fovea and AEV fovea) and in the periphery (AEH periphery and AEV periphery) are shown. The execution time is expressed as a percentage of the execution time in the Cartesian domain with five spatial scale.

CARTESIAN DOMAIN								
image size ($m \times n$)	number of scales	AEH	AEV	AEH fovea	AEV fovea	AEH periphery	AEV periphery	execution time
534×524	3	1.02	0.84	0.24	0.10	1.20	0.55	100%
534×524	2	1.45	0.98	0.28	0.11	1.91	0.83	92%
534×524	1	1.98	1.04	0.65	0.20	2.59	1.24	70%

CORTICAL DOMAIN								
($R \times S$) ρ_0, W_{max}	number of scales	AEH	AEV	AEH fovea	AEV fovea	AEH periphery	AEV periphery	execution time
232×373 $5, 4$	2	1.29	0.54	0.41	0.19	1.51	0.67	42%
232×373 $5, 4$	1	1.61	0.53	0.58	0.20	1.93	0.67	31%
154×247 $5, 6$	2	1.70	0.86	0.52	0.29	2.15	1.10	21%
154×247 $5, 6$	1	1.84	0.79	0.63	0.29	2.33	1.03	15%
115×183 $5, 8$	2	2.01	1.00	0.74	0.44	2.61	1.25	12%
115×183 $5, 8$	1	2.08	1.10	0.74	0.46	2.69	1.41	12%

Table 2: Performance comparison between the computation of the disparity in the Cartesian and in the log-polar domain for different sizes of the cortical image. The values refer to the test image shown in Figure 5b. The global average error for the horizontal disparity (AEH) and for the vertical disparity (AEV), together with the local error around the fovea (AEH fovea and AEV fovea) and in the periphery (AEH periphery and AEV periphery) are shown. The execution time is expressed as a percentage of the execution time in the Cartesian domain with three spatial scale.