



Design strategies for direct multi-scale and multi-orientation feature extraction 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

ARTICLE INFO

Article history:

Received 18 March 2010

Available online 21 September 2011

Communicated by P. Bhattacharya

Keywords:

Space-variant processing

Foveated images

Spatial filtering

Design criteria

Disparity computation

Visual processing

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. In this paper, we present a set of original rules for designing a discrete log-polar mapping that allows a direct application in the log-polar domain of the algorithms, based on spatial multi-scale and multi-orientation filtering, originally developed for the Cartesian domain. The advantage of the approach is to gain, without modifications, an effective space-variance and data reduction. Such design strategies 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 band-pass filters to compute reliable disparity maps, by providing quantitative measures of the computational load and of the accuracy of the computed visual features.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Inspired by the retina of mammals (Schwartz, 1977), the log-polar imaging is now a well established paradigm for simplifying a wide number of computational problems in pattern recognition and computer vision (see Berton et al., 2006; Traver and Bernardino, 2010 for reviews). The log-polar mapping simultaneously provides a wide field-of-view, a high spatial resolution on the region of interest, and a significant data reduction, besides rotation and scaling invariance properties. All these features are well suitable for robotics and active vision tasks (e.g. Monaco et al., 2009), and for pattern recognition applications (e.g. Amiri and Rabiee, 2010).

In the literature, many approaches to directly solve image processing and image understanding tasks in the space-variant representation of the visual signal have been described. Although, in theory, the conformal mapping should permit a direct application of the visual operators developed for the Cartesian images to log-polar ones, these approaches have to cope with the necessity of properly adapting the algorithms before applying them on the space-variant images. In particular, Fischl et al. (1997) analyze the properties of the log-polar transform from an analytic point of view, by addressing an application of the Laplacian operator, nevertheless neglecting the problems related to the discretization of the log-polar coordinates. Nattel and Yeshurun (2002) present some common global and local operators specifically designed to

be directly applied in the log-polar domain, and Smeraldi and Bigun (2002) use modified Gabor filters designed in the log-polar frequency plane for facial features detection and face authentication. Low-level feature extraction in log-polar images through an approach based on neural network is addressed in (Gomes and Fisher, 2003). Traver and Pla (2003) tackle the problem of 2D translation estimation in log-polar images, to take into account how a simple linear translation is transformed by applying the log-polar mapping. Three strategies for gradient detection, based on difference of Gaussian (DOG) operators in the space-variant geometry are presented and compared by Wallace and McLaren (2003). The proposed solution is a DOG operator, adapted to the distance from the image center. Recently, Pamplona and Bernardino (2009) claim that one of the major difficulties of working with log-polar images is to apply the usual Cartesian operators in the log-polar domain. Thus, they present a method for transferring some of the common Cartesian operators into the corresponding cortical operations. The proposed approach, based on the transformations of the operators between the two domains (see Mallot et al., 1990), requires specific modifications of the algorithms in order to take into account the log-polar transformation, which are quite expensive from the computational point of view. Finally, Zhang and Tay (2011) describe a space-variant approach for vergence control, based on disparity computed through normalized cross-correlation, adapted in the log-polar domain (see also Bernardino and Santos-Victor, 1996).

Notwithstanding the specific solutions present in the literature and described above, a general approach to extract visual features directly into the cortical domain is still missing. Moreover, the is-

* Corresponding author.

E-mail addresses: fabio.solari@unige.it (F. Solari), manuela.chessa@unige.it (M. Chessa).

sues related to the discrete log-polar transformations are often not explicitly addressed.

The main motivation of this paper is to present a general approach to the computation of visual features, directly in the cortical domain: in particular, 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 (Bigun, 2006; Granlund and Knutsson, 1995) are analyzed. The aim is to demonstrate that by a proper choice of such parameters we can directly use the algorithmic solutions developed for the Cartesian domain on log-polar images, without any modification. Moreover, the inherent space-variance of log-polar mapping is exploited to properly cope with the multi-scale issue. From this analysis, as a major contribution of this paper, it is thus possible to devise a set of design strategies, whose validity is proved with reference to the computation of complex visual features. As a test bed, we consider the computation of binocular disparity through an algorithm, previously developed for the Cartesian domain (Chessa et al., 2009a).

The paper is structured as follows: in Section 2 the symbols and notations used in the text are explained; the log-polar blind-spot model is described in Section 3; the relationships between the parameters of the log-polar mapping and of the local spatial filtering are analyzed in Section 4; in Section 5 the devised design strategies are explained; the experimental validation is presented in Section 6; and the conclusions are reported in Section 7. Finally, some details about the generality of the proposed approach are presented in the Appendix.

2. Symbols and notations

We use the following notations and symbols for denoting the different operations and terms in the subsequent text.

(x, y)	Continuous Cartesian coordinates
(ξ, θ)	Continuous log-polar coordinates (eccentricity, and angle)
(ρ, η)	Continuous polar coordinates
ρ_0	Radius of the innermost ring (size of the blind spot)
ρ_{max}	Radius of the outermost ring (size of the field of view)

a	Growth rate of the size of the log-polar pixel (receptive field) in the radial direction
$m \times n$	Size of Cartesian image (columns by rows)
$R \times S$	Size of log-polar image (rings by sectors)
(u, v)	Discrete log-polar coordinates
w, h, γ	Width, height, and aspect ratio of a log-polar pixel (receptive field)
$g(x, y)$	A generic spatial filter defined in the Cartesian domain
σ	Standard deviation of the Gaussian envelope (related to the spatial support)
ω_0, ϕ, α	Spatial peak frequency, phase, and orientation of a 2D Gabor filter
(x_ξ, y_ξ)	Continuous rotated Cartesian coordinates
$g(\xi, \theta)$	A filter defined in the cortical domain
$g(x(\xi, \theta), y(\xi, \theta))$	A filter mapped into the cortical domain
E	Response of a filter to a signal (inner product)
E_{mapped}	Response of a filter to a mapped filter
$E_{matched}$	Response of a filter to a matched filter
(ξ_0, θ_0)	Radial and angular position in the cortical domain (eccentricity, and angle)
E_i, ϕ_i	Response and phase of the i th filter of a bank
ϕ^{IN}	Phase of the input signal
W_{max}	Maximum size of a log-polar pixel
(d_ξ, d_θ)	Vector visual feature in the cortical domain
(d_x, d_y)	Vector visual feature in the Cartesian domain

3. Log-polar blind-spot model

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

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

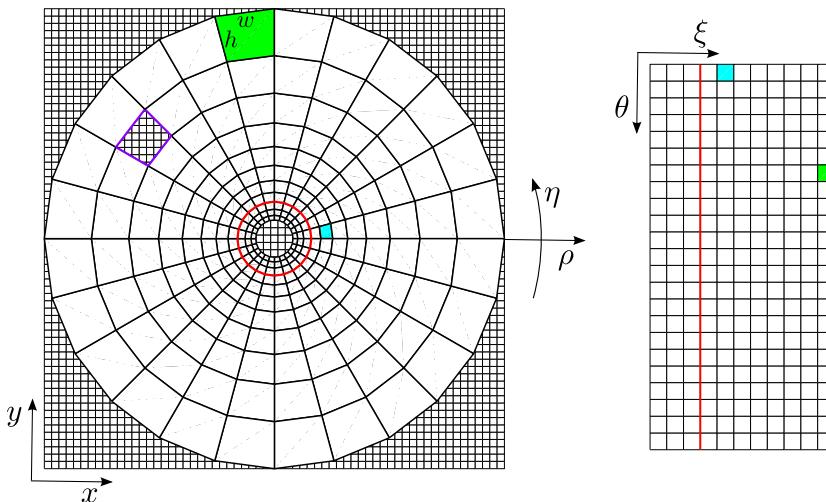


Fig. 1. Cartesian domain with the superposition of the log-polar receptive fields (left) and cortical domain (right). The green and the cyan areas represent two receptive fields at different angular and radial positions (thus with different size w and h) that are mapped in the two corresponding cortical pixels. The red circle delimits the oversampling and undersampling areas. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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.

It is worth noting that, 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 $(u, v) = (\lfloor \xi \rfloor, \lfloor \theta S/(2\pi) \rfloor)$ with $\lfloor \cdot \rfloor$ representing the integer part. Thus, the growth rate of the size of the receptive fields (log-polar pixel) between two consecutive rings (see Fig. 1) can be expressed as:

$$a = \exp \left(\ln \left(\frac{\rho_{max}}{\rho_0} \right) / R \right), \quad (2)$$

where $\rho_{max} = \frac{1}{2} \min(m, n)$.

Fig. 1 shows the log-polar receptive fields in 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 and June, 1977). This is pointed out in the receptive field bordered in violet. In order to highlight the properties of the discrete transformation, other

important parameters of the log-polar mapping must be defined, such as the aspect ratio of the log-polar pixel:

$$\gamma = \frac{2\pi}{S(a-1)} = \frac{w}{h}, \quad (3)$$

where w is the width, defined as $w = (2\pi/S)\rho_0 a^{u-1}$, and h is the height, defined as $h = \rho_0 a^{u-1}(a-1)$, of the log-polar pixel. In the following, the importance of the parameter γ for the visual processing will be analyzed.

4. Analysis of the parametric relationships between log-polar mapping and local spatial filtering

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 burden 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 $g(x, y)$, defined in the Cartesian domain, with respect to the

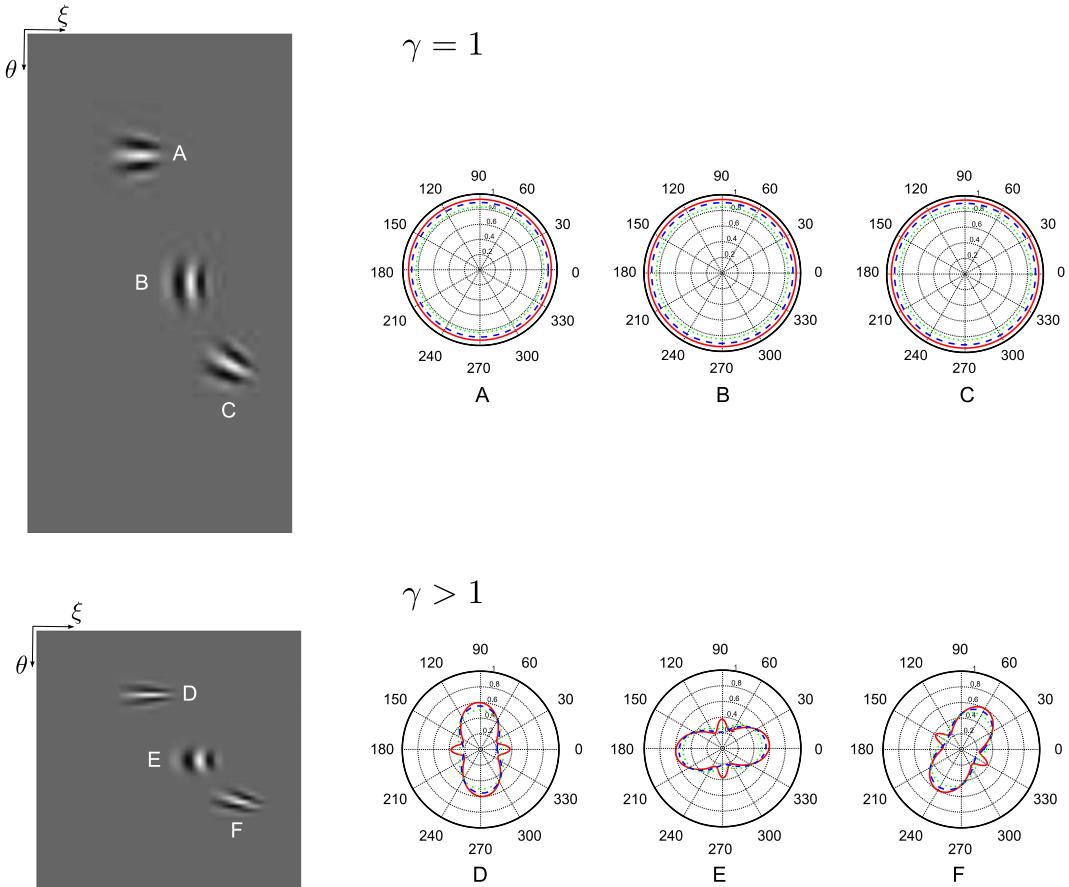


Fig. 2. (Left) Three mapped filters in the cortical domain, defined with same orientation $\alpha = 0$, different spatial support, and different position in the Cartesian domain. Two different aspect ratios γ are considered: $\gamma = 1$ and $\gamma > 1$ (similar results are obtained for $\gamma < 1$). (Right) The polar plots show the responses (E_{mapped}), normalized between 0 and 1, of the Gabor filters as a function of different orientations α in the Cartesian domain. The reference response of the corresponding matched filter is represented by the outermost black thick circle. Three different positions in the cortical plane have been considered: A–C for $\gamma = 1$, and D–F for $\gamma > 1$. For each of the three positions and for the two aspect ratios, three different cortical spatial support have been considered: 11 × 11, 21 × 21 and 31 × 31 pixels, whose responses are plotted in red solid, dashed blue and dotted green, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

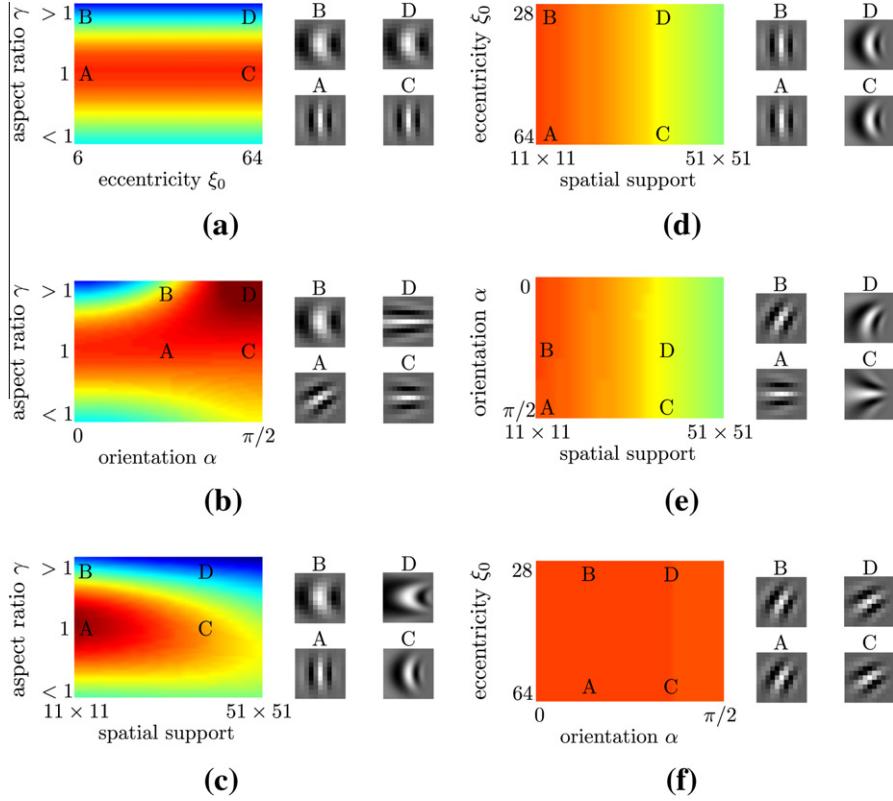


Fig. 3. Variations of the energy ratio $E_{\text{mapped}}/E_{\text{matched}}$ with respect to the parameters of the log-polar mapping and of the Gabor filters (left side of each subfigure). The profiles of the mapped filters for particular choices of such parameters are marked by capital letters A–D (right side of each subfigure). Hot colors represent high energy ratios (i.e. $E_{\text{mapped}}/E_{\text{matched}}$ is close to 1, thus the distortions of the mapped filter are minimal), whereas cold colors mean low energy ratios. In each subfigure, two parameters have been varied and the others are kept constant. The default values are: $\alpha = 0$, $\gamma = 1$, spatial support 11×11 pixels, $\xi_0 = 40$ (the value is related to the dimension of the used cortical image). (a) The maximum energy ratio is obtained for a squared log-polar pixel independently of the eccentricity in the cortical plane. (b) The energy is constant independently of the orientation of the filter when the aspect ratio is $\gamma = 1$. (c) The maximum energy ratio is obtained for a squared log-polar pixel ($\gamma = 1$) and a small spatial support (11×11 pixels). Under these conditions the filters show no deformation (see A), otherwise high deformations are present (see B–D). (d) The maximum energy is obtained for a small spatial support. (e) The maximum energy ratio is obtained for a small spatial support. (f) The energy is constant independently of the values of the parameters and the filters present only small deformation (see A–D).

different parameters of the log-polar mapping. Analogously, the spatial 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 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 responses of a filter to its mapped and to its matched filter. The response to the mapped filter is:

$$E_{\text{mapped}} = \langle g(\xi, \theta), g(x(\xi, \theta), y(\xi, \theta)) \rangle, \quad (4)$$

whereas the response to the matched filter is:

$$E_{\text{matched}} = \langle g(\xi, \theta), g(\xi, \theta) \rangle. \quad (5)$$

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.

In particular, the following analysis will be performed by using Gabor filters (Gabor, 1946), commonly used for visual feature extraction, 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 Daugman (1985):

$$g(x, y; \sigma, \alpha, \phi) = \frac{1}{\sqrt{\pi}\sigma} \exp\left(-\frac{x_z^2 + y_z^2}{2\sigma^2}\right) \exp(j\omega_0 x_z + \phi), \quad (6)$$

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, α is the orientation of the filter, and (x_z, y_z) are the rotated spatial coordinates in the Cartesian domain.

4.1. 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 visual feature extraction we are addressing constrains the choice of the parameters ρ_0 , ρ_{\max} and R . The radius of the innermost ring ρ_0 determines the size of the blind spot, and the radius of the outermost ring ρ_{\max} determines the size of the field of view. The size of the log-polar image depends on the number of rings R . These parameters affect the space-variance of the mapping (see Eq. (2)). Once fixed ρ_0 , ρ_{\max} and R , we analyze¹ how the response

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.

E_{mapped} of the Gabor filters is affected by the discrete log-polar mapping, and by the parameters of the filter (see Fig. 2). The filter is characterized by its spatial support and by its orientation (*see Eq. (6)*). Moreover, due to the space-variance of the mapping, it is important to consider different positions (ξ_0, θ_0) in the cortical plane. Finally, the discrete mapping is affected by the angular and radial sampling, that is the aspect ratio of the log-polar pixel (see Eq. (3)). The responses of the differently oriented filters (colored profiles in the polar plots) for an aspect ratio $\gamma = 1$ (first row of Fig. 2) are compared to the responses obtained with $\gamma > 1$ (second row of Fig. 2). 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.

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

A further analysis is presented 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 log-polar mapping and those of the Gabor filter. The results are shown in Fig. 3. Each subfigure shows the variation of the energy ratio $E_{\text{mapped}}/E_{\text{matched}}$ with respect to pairs of parameters of the mapping and of the filters (left side), and the profiles 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 $E_{\text{mapped}}/E_{\text{matched}}$ remains high, independently of the eccentricity ξ_0 in the cortical plane and of the orientation α of the Gabor filter (see Fig. 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, Fig. 3c shows that the maximum response is obtained when the spatial support of the filter is small (e.g. 11×11 pixels) and γ is close to 1. It is worth noting that under these conditions the deformations of the mapped filters are relatively small (see inset A of Fig. 3c). Once fixed $\gamma = 1$, the influence of the spatial support of the filter can be evidenced from Fig. 3d–e. The response of the filter 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 Fig. 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 Fig. 3f).

The methodology of the analysis presented in this Section and the resulting dependencies among the different parameters are not bounded to a specific filter (i.e. Gabor Filter), but they can be considered general and applied to a wide class of filters (see Appendix A for further results).

4.3. Response of a bank of filters as a function of the parameters of the mapping

In this Section, we address the problem of extracting visual features through a direct filtering in the cortical domain. In particular, we consider the computation, for each orientation, of the local phase in the image signal, by adopting a distributed representation of the local phase. 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).

To perform such analysis, we analyze the different responses E_i 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

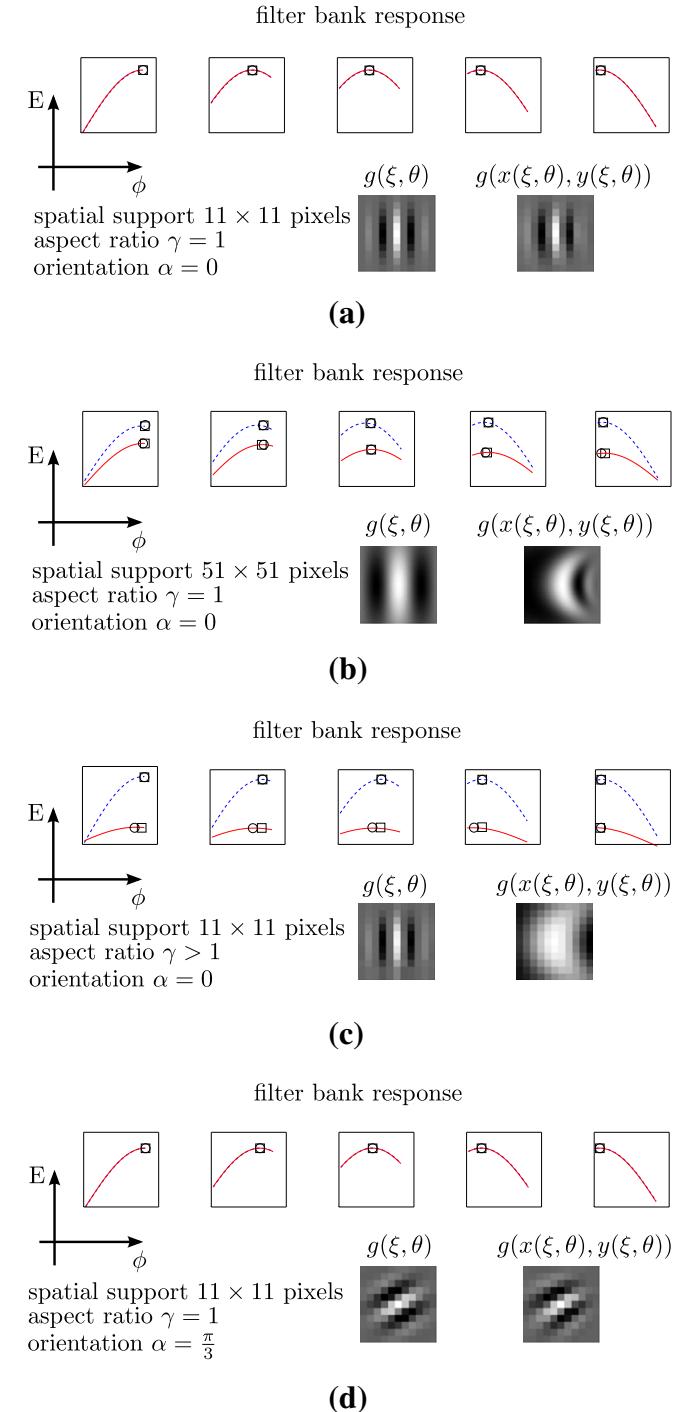


Fig. 4. Comparison between the responses of a bank of matched filters (dashed blue profiles) and those of a bank of mapped filters (solid red profiles) for different combinations of the parameters of the mapping and of the filter (a–d). Five different values of the phase ϕ^{IN} are considered for each filter bank response. 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 represented as gray level images. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

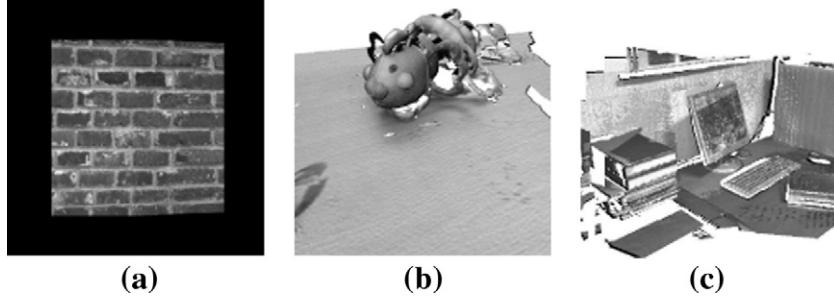


Fig. 5. Left images of the considered stereo pairs. (a) A synthetic frontoparallel plane. (b) and (c) Real-world scenarios acquired by a 3D laser scanner.

response curve is bell-shaped and symmetric with respect to its peak.

Fig. 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 analyzed. Fig. 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 $\alpha = 0$. Fig. 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. A bank of filters with larger spatial supports generates a flatter response curve, thus compromising the reliability of the detection of the peak. The aspect ratio γ of the log-polar pixel also affects the filter bank response (not only the response E of a single filter). Fig. 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 Fig. 4d).

5. Design rules of the log-polar mapping for visual processing

5.1. Proper selection of the parameters of the mapping and of the filters

In the previous Section, we have devised the constraints for the parameters of the log-polar mapping and of the spatial filters, to make the signal processing in the cortical domain equivalent to the one in the Cartesian domain. In order to extract visual features in a reliable way, the response of a filter to a given input signal should depend on the feature we are interested in, only. Any other variation, due to distortions or anisotropy introduced by the log-polar transformation, should be avoided. Thus the following design rules have to be considered:

- The aspect-ratio of the log-polar pixel has to be around 1. 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 in this paper shows that this rule can be generalized in order to efficiently use the local spatial operators to measure important elements of the visual signal (Adelson and Bergen, 1991; Granlund and Knutsson, 1995).
- The spatial support of the filter has to be small, since the growth of the size of the filters introduces a uniform decrease of the responses, due to the higher distortion of the mapped filters,

that reduces the reliability of the feature extraction. We can observe that the minimum spatial support of a filter is bounded by the sampling constraints.

5.2. 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 is to consider a polar fovea (Bertoni 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 undesired distortions and anisotropies are introduced, as we have previously demonstrated. It is also worth noting that the size of the blind spot is usually small, if compared with the spatial support of the filter, and thus the problem associated with the discontinuity issue diminishes.

5.3. 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 the common spatial filters, used for feature extraction, are band-pass filters, centered in their own spatial peak frequency, whereas information in natural images is spread over a wide range of frequencies, it is necessary to use a technique that allows us 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. The number of spatial scales depends on the specific processing task addressed. The space variance of the log-polar mapping, i.e. the linear 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 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 locations (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. The spatial support of the filter is a function of the log-polar pixel size $w \times h$, thus the choice of the desired size of the filter at the maximum

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.

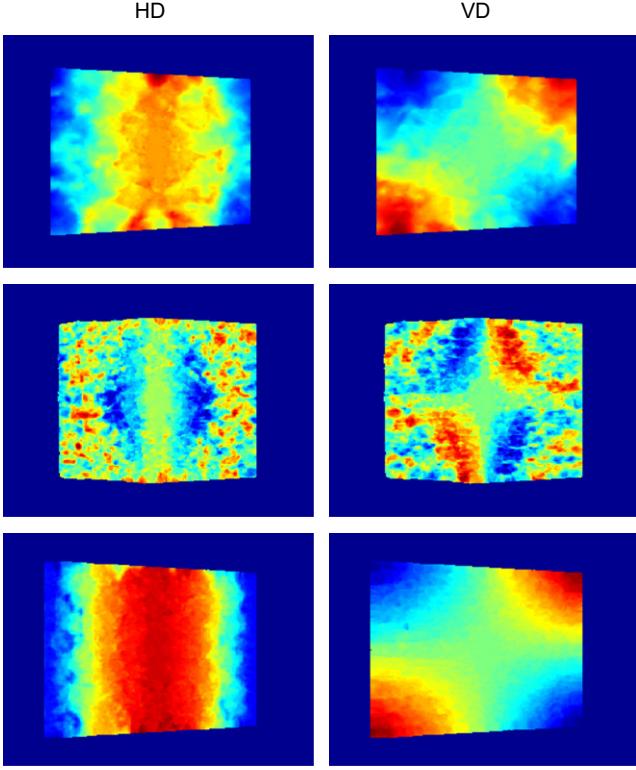


Fig. 6. Horizontal (HD) and vertical (VD) disparities for a frontoparallel plane, when the optical axes are vergent in the center of the plane. (First row) Computation in the cortical domain with two “vertical” scales. (Second row) Computation in the Cartesian domain with two “vertical” scales and with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

eccentricity ρ_{max} constraints 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} . By assuming the aspect ratio of the log-polar pixels $\gamma = 1$, that is $w = h = \rho_0 a^{u-1}(a - 1)$ and by considering Eq. (2), the novel rule that relates the total number of rings R to W_{max} can be expressed by:

$$R = -\frac{\ln(\rho_{max}/\rho_0)}{\ln((\rho_{max} - W_{max})/\rho_{max})}. \quad (7)$$

Hence, the log-polar mapping is defined by three parameters: ρ_0 , ρ_{max} and W_{max} .

5.4. 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):

$$\begin{bmatrix} d_x \\ d_y \end{bmatrix} = \begin{bmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial \xi} & \frac{\partial y}{\partial \theta} \end{bmatrix} \begin{bmatrix} d_\xi \\ d_\theta \end{bmatrix}. \quad (8)$$

Combining Eqs. (1) and (8), we obtain:

$$\begin{bmatrix} d_x \\ d_y \end{bmatrix} = \rho_0 a^\xi \ln(a) \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} d_\xi \\ d_\theta \end{bmatrix} \quad (9)$$

The scalar coefficient of Eq. (9) 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 described 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 (Chessa et al., 2009a).

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. In (Weiman, 1994) a preliminary study on the use of local Gabor filters for the computation of binocular disparity in log-polar images is presented, but no results are shown. Ahrns and Neumann (1998) use Gabor filters to compute optical flow, then they integrate it with a Kalman filter, in order to obtain depth estimation with a structure from motion approach. Sparse disparity maps are obtained by using a matching of Laplacian features in (Grosso and Tistarelli, 2000). Bernardino and Santos-Victor (1996) and

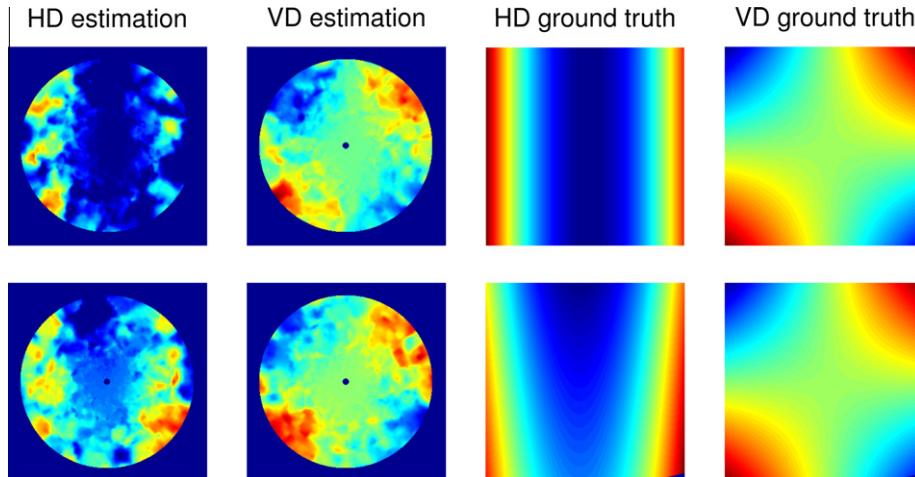


Fig. 7. Comparison between the estimated disparity maps and the ground truth for a frontoparallel plane (first row) and for a slanted plane (second row). In both situations the axes of the two cameras are vergent in the center of the planes. The average errors in the computation of the horizontal (HD) 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.

Manzotti et al. (2001) present techniques for extracting a disparity measure for vergence control, thus neglecting the contribution of the vertical disparity, whereas 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.

It is worth noting that 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. Fig. 5 shows the left images of the stereo pairs used in the following analysis. 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. 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 11 pixels. Moreover, since the presence of large disparities in the periphery affects the number of scales, both “horizontal” multi-scale (see Eq. (7)) and

“vertical” multi-scale are used. Given the maximum value of disparity we aim to compute, 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, and to the design rules we have formulated in Section 5, we first transform the stereo image pair into the cortical domain, then the 2D vector disparity is computed in the cortical domain (ξ, θ) . The following set of parameters are used for all the experiments: 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. (9), where we perform the quantitative benchmark with respect to the ground truth data.

Fig. 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. Fig. 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 (see Fig. 6 (second row)), where instead five “vertical” scales are necessary (third row).

In order to quantitatively analyze the reliability of the results, the computed disparity maps are compared to the available ground truth maps. Figs. 7–9 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. We can observe that the resulting error in the disparity estimation is small with respect to the range of the ground truth disparity values.

Furthermore, the reliability of the disparity values with respect to the different parameters of the mapping is analyzed. Tables 1 and 2 show how the size of the cortical image, defined by $\gamma = 1$ and by ρ_0, ρ_{max} and W_{max} (see Eq. (7)), and the number of the con-

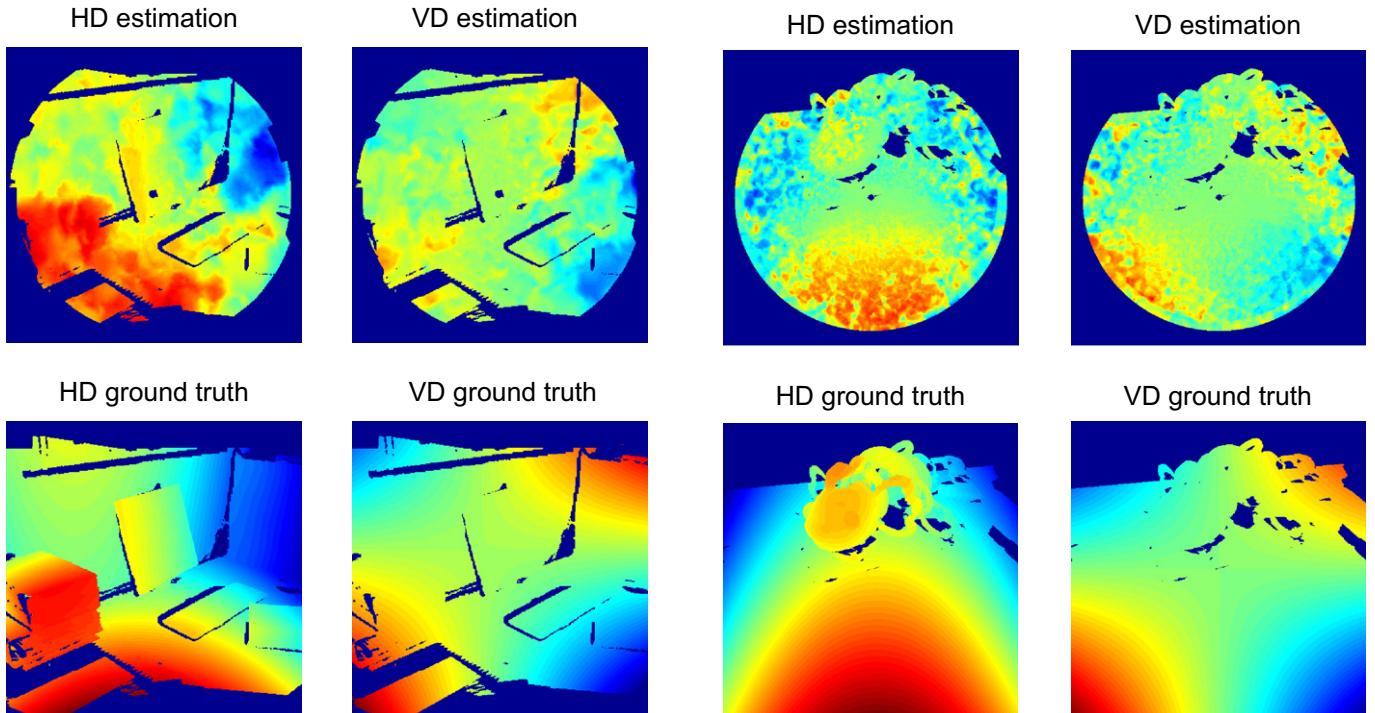


Fig. 8. Comparison between the estimated disparity maps (first row) and the ground truth (second row) for a stereo pair obtained from a real scenario acquired by a laser scanner. The average errors in the computation of the horizontal (HD) and vertical (VD) disparities are 1.50 and 0.57 pixels, respectively.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

Fig. 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.

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 Fig. 5a. AEH and AEV are the global average errors for the horizontal and for the vertical disparity, respectively. AEH fovea and AEV fovea are the local errors around the fovea. AEH periphery and AEV periphery are the errors in the periphery. The execution time is expressed as a percentage of the execution time in the Cartesian domain with five spatial scale.

	Number of scales	AEH	AEV	AEH fovea	AEV fovea	AEH periphery	AEV periphery	Execution time (%)
<i>Cartesian domain</i>								
<i>(m × n)</i>								
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
<i>Cortical domain</i>								
<i>R × S, ρ₀, W_{max}</i>								
100 × 159, 3, 5	2	1.37	0.45	0.29	0.29	1.53	0.61	36
100 × 184, 5, 5	2	1.31	0.39	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	1.95	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 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 Fig. 5b. AEH and AEV are the global average errors for the horizontal and for the vertical disparity, respectively. AEH fovea and AEV fovea are the local errors around the fovea. AEH periphery and AEV periphery are the errors in the periphery. The execution time is expressed as a percentage of the execution time in the Cartesian domain with three spatial scale.

	Number of scales	AEH	AEV	AEH fovea	AEV fovea	AEH periphery	AEV periphery	Execution time (%)
<i>Cartesian domain</i>								
<i>(m × n)</i>								
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
<i>Cortical domain</i>								
<i>R × S, ρ₀, W_{max}</i>								
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

sidered “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 ρ_{max}) 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, i.e. by the lower resolution due to the mapping. The analysis shows that the average error in the region around the fovea is small, i.e. less than 1 pixel in every condition, thus the visual feature is reliable enough to be used in computer vision systems.

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 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.

The results shown in this Section confirm that it is possible to compute the disparity directly in the cortical domain, without modifying the algorithm, by taking advantage of the “horizontal” scale and of the consistent data reduction, and thus of the low execution time, with a tolerable approximation error.

7. Conclusion

In this paper, we have addressed the problem of visual processing through linear spatial filtering directly in the log-polar domain,

by also taking into account the multi-orientation and multi-scale approaches. The extraction of visual features based on spatial filtering has a great importance for many applications of pattern recognition and computer vision, e.g. edge detection, face recognition, object detection, disparity and optic flow estimation. 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 the spatial filters has been carried out.

The major outcome of this analysis is the definition of a set of general design strategies, that allow us to use algorithms, which were originally designed in the Cartesian domain, directly in the log-polar space, without introducing specific modifications. In particular, we have deduced constraints on the aspect ratio of the log-polar pixel and on the spatial support of the filter. Moreover, we have devised a novel rule to efficiently implement a multi-scale analysis, by exploiting the space-variance of the log-polar mapping.

The validity of the proposed design rules has been proved by using an algorithm that computes a complex visual feature (the binocular disparity) through a hierarchical combination of the responses of a bank of spatial filters. The obtained results show that it is possible to recover reliable values of the considered visual features by directly applying the algorithm in the cortical domain, thus achieving the advantages of the space-variant representation, e.g. a consistent reduction in the execution time.

The software libraries and the dataset used for the experimental validation are made publicly available at www.pspc.dibe.unige.it/Research/logpolar.html.

Acknowledgments

This work has been partially supported by EU FP7-ICT 217077 Project “EYESHOTS”, and by national project “PRIN 2008”.

Appendix A. A general approach to spatial filtering in cortical domain

The results of the systematic analysis carried out in Section 4 is not limited to the Gabor filters, since we consider them as an instantiation of the class of band-pass filters used for visual feature extraction (Mallat, 1989; Adelson and Bergen, 1991; Bigun, 2006). The same considerations can be applied to a large set of local spatial operators, since the non-linear mapping of the Cartesian do-

main yields a spatial distortion of the transformed domain. Fig. A.10 shows how the uniform Cartesian domain is represented in the log-polar plane: both the global features, e.g. the red and green stripes, and the local features, e.g. the cyan and magenta boxes, are distorted. We have to take into account how such distortions affect the shape of the filter, and consequently its processing capabilities. To this aim, it is not sufficient to evaluate the magnitude of the distortions in the transformed domain, but how such distortions affect the inner product (see Eqs. (4) and (5)). Since the inner product has no analytical solution for practical use filters, even for a simple Gaussian filter, in this paper it has been solved numerically.

Moreover, it is worth noting that the real implementation of the log-polar mapping relies on a discrete transform, thus we have to take into account the sampling of the Cartesian domain. In partic-

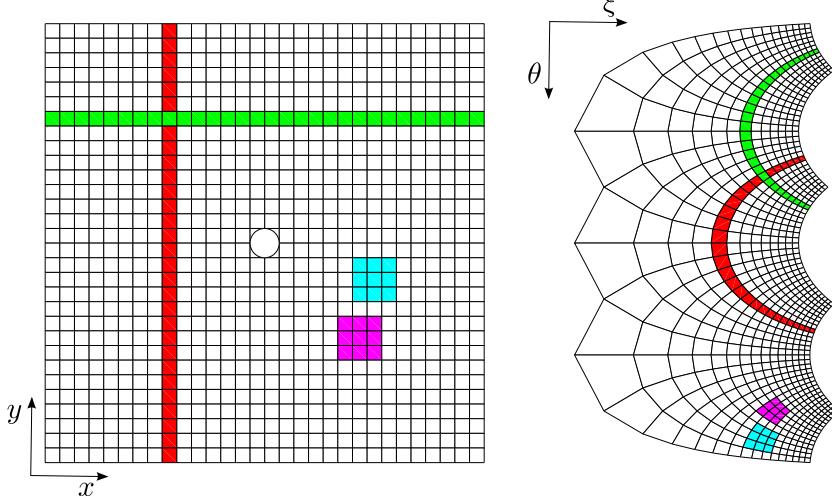


Fig. A.10. Cartesian domain (left) and the corresponding transformed log-polar domain (right).

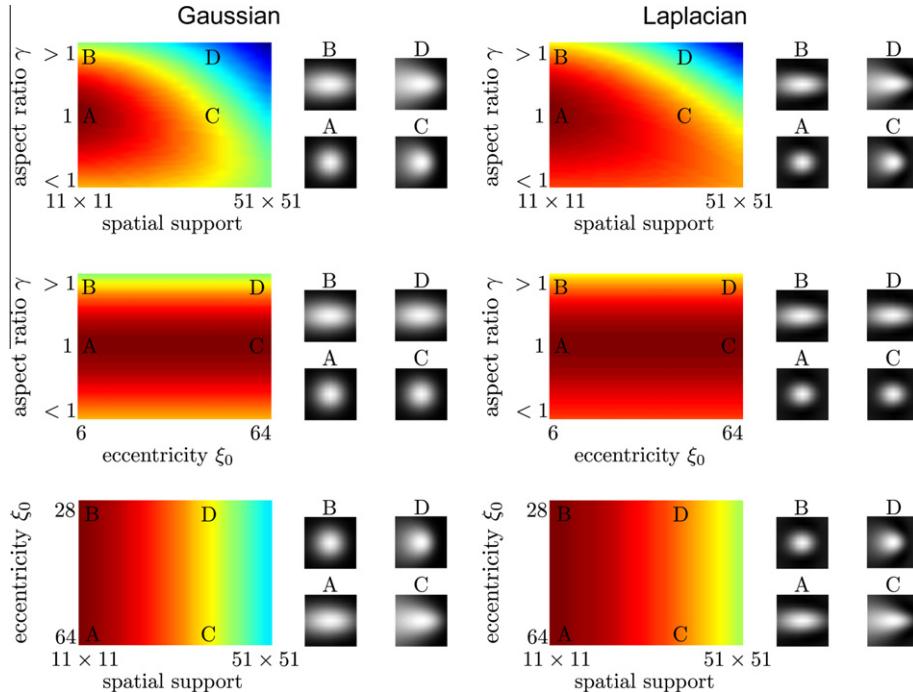


Fig. A.11. Variations of the energy ratio $E_{\text{mapped}}/E_{\text{matched}}$ with respect to the parameters of the log-polar mapping and of the filters (left side of each subfigure). The profiles of the mapped filters for particular choices of parameters are marked by capital letters A–D (right side of each subfigure). Notations and parametric variations as in Fig. 3.

ular, we have to properly choose the sampling of the radial and angular coordinates in order to represent them equally in the cortical domain. This choice constraints the aspect-ratio γ of the log-polar pixel to a value close to 1.

To clarify the generality of the approach, the same analysis of Section 4 is performed for two isotropic spatial filters: a Gaussian and a Laplacian filter. The former is commonly used to preprocess the images, in order to remove noise, but it is also used as a weighting function in the processing stages of several computer vision algorithms. The latter is a two-dimensional differential isotropic filter, used in image processing, in particular for edge detection.

We consider a Gaussian filter normalized by its energy, defined as:

$$g(x, y) = \frac{1}{\sqrt{\pi}\sigma} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right), \quad (\text{A.1})$$

and a Laplacian filter, defined as:

$$g(x, y) = \frac{1}{\sqrt{\pi}\sigma^3} \left(2 - \frac{x^2 + y^2}{\sigma^2}\right) \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right). \quad (\text{A.2})$$

Fig. A.11 shows the results of the analysis described in Section 4.2 for a Gaussian filter (left side) and for a Laplacian filter (right side). As we obtained for the Gabor filters, it is possible to conclude that, to have low distortions, the aspect-ratio γ must be close to 1. Moreover, for a given value of the spatial support, once fixed $\gamma = 1$, the responses of the filters do not depend on the eccentricity in the cortical plane, and a small spatial support produces a more reliable response.

References

- Adelson, E., Bergen, J., 1991. The plenoptic and the elements of early vision. In: Landy, M., Movshon, J. (Eds.), Computational Models of Visual Processing. MIT Press, pp. 3–20.
- Ahrns, I., Neumann, H., 1998. Combining log-polar mapping and gabor filtering for real-time depth estimation. In: Proc. Artificial Intelligence. pp. 119–126.
- Amiri, M., Rabiee, H.R., 2010. A novel rotation/scale invariant template matching algorithm using weighted adaptive lifting scheme transform. *Pattern Recognit.* 43, 2485–2496.
- Bernardino, A., Santos-Victor, J., 1996. Vergence control for robotic heads using log-polar images. In: Intelligent Robots and Systems. pp. 1264–1271.
- Bernardino, A., Santos-Victor, J., 2002. A binocular stereo algorithm for log-polar foveated systems. In: Biologically Motivated Computer Vision. pp. 127–136.
- Bertoni, F., Sandini, G., Metta, G., 2006. Encyclopedia of Sensors. American Scientific Publishers, pp. 1–16, Ch. Anthropomorphic visual sensors.
- Bigun, J., 2006. Vision with Directions: A Systematic Introduction to Image Processing and Computer Vision. Springer-Verlag, Berlin Heidelberg.
- Bolduc, M., Levine, M.D., 1998. A review of biologically motivated space-variant data reduction models for robotic vision. *Comput. Vision Image Understand.* 69 (2), 170–184.
- Bonmassar, G., Schwartz, E., 1997. Space-variant fourier analysis: The exponential chirp transform. *IEEE Trans. Pattern Anal. Mach. Intell.* 19 (10), 1080–1089.
- Burt, P., Adelson, E., 1983. The laplacian pyramid as a compact image code. *IEEE Trans. Commun.* COM-31, 532–540.
- Chan Man Fong, C., Kee, D., Kaloni, P., 1997. Advanced Mathematics For Applied and Pure Sciences. CRC Press.
- Chessa, M., Sabatini, S., Solari, F., 2009a. A fast joint bioinspired algorithm for optic flow and two-dimensional disparity estimation. In: Computer Vision Systems. Lecture Notes in Computer Science, pp. 184–193.
- Chessa, M., Solari, F., Sabatini, S., 2009b. A virtual reality simulator for active stereo vision systems. In: Proc. Fourth Internat. Conf. Computer Vision Theory and Applications. pp. 444–449.
- Colombo, C., Rucci, M., Dario, P., 1996. In: Image Technology: Advances in Image Processing, Multimedia and Machine Vision. In: Sanz, L.C. (Ed.), Ch. Integrating selective attention and spacevariant sensing in machine vision. Springer, pp. 109–127.
- Daugman, J., 1985. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Amer. A* 2, 1160–1169.
- Fischl, B., Cohen, M., Schwartz, E., 1997. The local structure of space-variant images. *Neural Networks* 10 (5), 815–831.
- Fleet, D.J., Jepson, A.D., 1993. Stability of phase information. *IEEE Trans. Pattern Anal. Mach. Intell.* 15 (12), 1253–1268.
- Florack, L., 2007. Modeling foveal vision. In: Scale Space and Variational Methods in Computer Vision. pp. 919–928.
- Gabor, D., 1946. Theory of communication. *J. Inst. Elec. Eng.* 93, 429–459.
- Gomes, H.M., Fisher, R.B., 2003. Primal sketch feature extraction from a log-polar image. *Pattern Recognit. Lett.* 24, 983–992.
- Granlund, G., Knutsson, H., 1995. Signal Processing for Computer Vision. Kluwer Academic Publishers, Dordrecht.
- Grosso, E., Tistarelli, M., 2000. Log-polar stereo for anthropomorphic robots. In: Proc. 6th Eur. Conf. Computer Vision-Part I. pp. 299–313.
- Jerry, A., 1977. The shannon sampling theorem – its various extensions and applications: A tutorial review. *Proc. IEEE* 65 (11), 1565–1596.
- Jurie, F., 1999. A new log-polar mapping for space variant imaging – application to face detection and tracking. *Pattern Recognit.* 32, 865–875.
- Mallat, S.G., 1989. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 11, 674–693.
- Mallot, H.A., Seelen, W., Giannakopoulos, F., 1990. Neural mapping and space-variant image processing. *Neural Networks* 3 (3), 245–263.
- Manzotti, R., Gasteratos, A., Metta, G., Sandini, G., 2001. Disparity estimation on log-polar images and vergence control. *Comput. Vision Image Understand.* 83 (2), 97–117.
- Monaco, J.P., Bovik, A.C., Cormack, L.K., 2009. Active, foveated, uncalibrated stereovision. *Int. J. Comput. Vision* 85, 192–207.
- Nattel, E., Yeshurun, Y., 2002. Direct feature extraction in a foveated environment. *Pattern Recognit. Lett.* 23 (13), 1537–1548.
- Pamplona, D., Bernardino, A., 2009. Smooth foveal vision with gaussian receptive fields. In: 9th IEEE-RAS International Conference on Humanoid Robots.
- Qian, N., Mikaelian, S., 2000. Relationship between phase and energy methods for disparity computation. *Neural Comput.* 12 (2), 279–292.
- Schindler, K., 2006. Geometry and construction of straight lines in log-polar images. *Comput. Vision Image Understand.* 103 (3), 196–207.
- Schwartz, E., 1977. Spatial mapping in the primate sensory projection: Analytic structure and relevance to perception. *Biol. Cybernet.* 25, 181–194.
- Smeraldi, F., Bigun, J., 2002. Retinal vision applied to facial features detection and face authentication. *Pattern Recognit. Lett.* 23, 463–475.
- Traver, V., Bernardino, A., 2010. A review of log-polar imaging for visual perception in robotics. *Robot. Autonom. Syst.* 58 (4), 378–398.
- Traver, V., Pla, F., 2003. Dealing with 2d translation estimation in log-polar imagery. *Image Vision Comput.* 21 (2), 145–160.
- Traver, V., Pla, F., 2008. Log-polar mapping template design: From task-level requirements to geometry parameters. *Image Vision Comput.* 26 (10), 1354–1370.
- Wallace, A., McLaren, D., 2003. Gradient detection in discrete log-polar images. *Pattern Recognit. Lett.* 24 (14), 2463–2470.
- Weiman, C.F.R., 1994. Log-polar binocular vision system. Technical Report, CR-188375, NASA.
- Zhang, X., Tay, L.P., 2011. A spatial variant approach for vergence control in complex scenes. *Image Vision Comput.* 29, 64–77.