Robotics and Autonomous Systems 58 (2010) 378–398

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

A review of log-polar imaging for visual perception in robotics

V. Javier Traver [a,b,*], Alexandre Bernardino [c]

[a] Computer Languages and Systems Dept., Universitat Jaume I, Castellón, Spain

[b] Institute of New Imaging Technologies (INIT), Universitat Jaume I, Castellón, Spain

[c] Instituto de Sistemas e Robótica, Instituto Superior Técnico, Lisbon, Portugal

ARTICLE INFO

Article history:
Received 31 July 2008
Received in revised form
6 October 2009
Accepted 13 October 2009
Available online 30 October 2009

Keywords:
Foveal imaging
Log-polar mapping
Real-time robotics
Active vision

ABSTRACT

Log-polar imaging consists of a type of methods that represent visual information with a space-variant resolution inspired by the visual system of mammals. It has been studied for about three decades and has surpassed conventional approaches in robotics applications, mainly the ones where real-time constraints make it necessary to utilize resource-economic image representations and processing methodologies. This paper surveys the application of log-polar imaging in robotic vision, particularly in visual attention, target tracking, egomotion estimation, and 3D perception. The concise yet comprehensive review offered in this paper is intended to provide novel and experienced roboticists with a quick and gentle overview of log-polar vision and to motivate vision researchers to investigate the many open problems that still need solving. To help readers identify promising research directions, a possible research agenda is outlined. Finally, since log-polar vision is not restricted to robotics, a couple of other areas of application are discussed.

1. Introduction

Both natural and artificial visual systems have to deal with large amounts of information coming from the surrounding environment. When real-time operation is required, as happens with animals or robots in dynamic and unstructured environments, image acquisition and processing must be performed in a very short time (a few milliseconds) in order to provide a sufficiently fast response to external stimuli. Appropriate sensor geometries and image representations are essential for the efficiency of the full visual processing stream. To address this problem it is wise to look for the solutions present in biological systems, which have been optimized by millions of years of evolution. For instance, the visual system of many animals exhibits a non-uniform structure, where the *receptive fields*[1] represent certain parts of the visual field more densely and acutely. In the case of mammals, whose eyes are able to move, retinas present a unique high resolution area in the center of the visual field, called the *fovea*. The distribution of receptive fields within the retina is fixed and the fovea can be redirected to other targets by ocular movements. The same structure is also commonly used in robot systems with moving cameras [2–6].

In the late 70s computer vision researchers broke new ground by considering the foveal nature of the visual systems of primates as an alternative to conventional uniform resolution sensors for artificial perception in computers and robots. Earlier on, biological findings in the visual cortex of monkeys [7] had shown that the displacement of a light stimulus in the retina produces displacements in the cortex that are inversely proportional to the distance to the fovea. This effect, also known as *cortical magnification*, indicates a general scaling behavior by which both receptive field spacing and size increase linearly with eccentricity, i.e. the distance from the fovea [8]. It was found that responses to linear stimuli originating in the fovea lie roughly along lines in the cortex, and circular stimuli centered on the fovea produce linear responses in the cortex at approximately orthogonal orientations [9]. Thus, the information transmitted between the retina and the visual cortex is organized in an approximate logarithmic-polar law [10].

The foveal structure of the retina of some animals is, together with their ability to move the eyes, a fundamental mechanism in the control of visual perception. In the late 80s and early 90s, researchers started exploiting eye movements to achieve complex visual tasks. The paradigm of active vision emerged as a powerful concept to endow an active observer with the ability to find more efficient solutions to problems that, from a passive vision perspective, were ill-posed and non-linear [11]. The idea can be generalized beyond pure perception by including manipulation. The smart usage of robot arms and hands, for instance, opens up many more

* Corresponding author at: Computer Languages and Systems Dept., Universitat Jaume I, Castellón, Spain.
E-mail addresses: vtraver@uji.es (V. Javier Traver), alex@isr.ist.utl.pt (A. Bernardino).

1 Receptive fields are spatially organized biological computational elements and, according to [1], are probably the most prominent and ubiquitous computational mechanism employed by biological information systems. The concept is revisited in Section 2.

0921-8890/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.robot.2009.10.002

possibilities for better visual perception [12]. A great deal of excitement was aroused at that time with regard to the present and future possibilities of active vision [13]. For instance, by purposefully moving the eyes, an observer with a foveal low resolution sensor can acquire a "virtual" high resolution image of its entire field of view [14]. Therefore, following on as the next natural step forward, a new concept appeared —*space-variant active vision* [15].

Since then, efforts have been made to explore the advantages that foveal-like log-polar imaging can bring to robotic applications. However, after almost three decades since those initial studies, no systematic and comprehensive work has been published that attempts to review past research on the topic. While a careful review of log-polar models was conducted ten years ago in [16], its focus was a detailed study and comparison of log-polar mapping templates and models with overlapping receptive fields. More recently, the motivations for retina-like sensors and the properties of the log-polar mapping were nicely considered in [17,18]. Another paper [19] surveyed foveated sensors with a particular emphasis on image processing issues.

Thus, while these few review-like papers did not consider the applications and usages of log-polar images in depth, we feel that such an analysis is needed, to reflect on past achievements, discuss current challenges, and predict future developments. Additionally, this literature overview would be a valuable aid to any researcher interested in approaching the field, particularly to beginners. Finally, another important benefit of such a survey is that of helping to promote further work, both on the theoretical and practical sides of log-polar vision.

Therefore, the present survey aims to complement these previous reviews, by looking further into the variety of applications of log-polar sensing that have been proposed. Furthermore, our analysis pays particular attention to robotic applications. Hence, the usage of the log-polar transform for pattern recognition issues, though important, is not considered here. Studies and reviews on this other perspective can be found elsewhere [20–23].

An overview of the log-polar mapping (Section 2) allows the readers to become familiar with the basics of this transform. There are different ways to obtain log-polar images either from conventional images or directly from a scene, using software- and/or hardware-based solutions (Section 3). In Section 3 we also address issues regarding how the mapping parameters may influence the visual task and whether the selection of optimal parameters can be automated. The area of visual attention and salience computation under foveal vision (Section 4) has not been explored very much, even though it plays a key role in active object search and recognition, in exploratory gaze strategies, and in the proper integration of different visual tasks in practical scenarios. One of the visual processes where log-polar imaging is most suitable is probably active target tracking (Section 5), and substantial research has been devoted to this topic. Some advantages have also been found in estimating the observer's motion using log-polar images (Section 6), basically due to its polar geometric nature which fits particularly well with time-to-collision computation and other navigation tasks in mobile robots. Binocular depth estimation has been considered with a joint usage of log-polar imaging and active vergence movements (Section 7). There are also a number of less conventional sensor arrangements and less known properties of log-polar imaging that deserve some consideration. It is our prediction that many of these issues will open up the door to fascinating new research challenges in automatic foveal vision not only within robotics but also in other fields of application (Section 8).

2. Log-polar mapping

Log-polar mapping is a geometrical image transformation that attempts to emulate the topological reorganization of visual information from the retina to the visual cortex of primates. It can also be found in the literature under different names, such as *log-polar transformation* or the log(z) *model*. The reason for this last denomination comes from the fact that the mapping can be mathematically modeled by the complex logarithmic function log(z), where z is the complex variable representing points on the image plane.

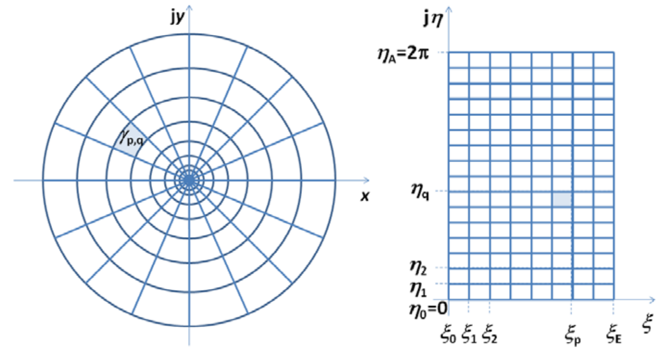

Fig. 1. The log(z) model for retino-cortical mapping, where a central circle of small radius has been left out of the mapping in order to deal with the singularity problem. The retinal plane (left) is mapped onto the cortical plane (right) via $w = \log(z)$. Concentric circumferences and radial lines in the retinal plane become straight lines in the cortical plane. Rectangular cells in the transform domain correspond to sections of concentric annuli in the original domain.

2.1. Definition

Let us consider the complex retinal and cortical (log-polar) planes, represented by the variables $z = x + jy$ and $w = \xi + j\eta$, respectively (j is the complex imaginary unit). The complex log-polar mapping is:

$$w = \log(z) \tag{1}$$

and the log-polar coordinates ξ (eccentricity) and η (angle) are given by:

$$\xi = \log(|z|) = \log\sqrt{x^2 + y^2}$$
$$\eta = \arg(z) = \operatorname{atan2}(y, x)$$

where atan2(y, x) denotes the two-argument arctangent function that considers the sign of x and y in order to determine the quadrant of the resulting angle.

This mapping transforms concentric circumferences and radial lines in the retinal plane into straight lines along the ξ and η directions in the cortical plane, respectively (Fig. 1).

2.2. Properties

The main properties of the mapping and some of their practical implications are as follows:

Conformal mapping: The cortical image (also called *log-polar image*) preserves oriented angles between curves and neighborhood relationships, almost everywhere, with respect to the retinal image. In theory, this property predicts that image processing operations developed for Cartesian images can be applied directly to log-polar images. In practical terms, however, specific algorithms are required in many applications.

Elegant trade-off solution between these three mutually opposing criteria: wide field of view, high visual resolution and little data to process. For robotics applications, there are two significant benefits of this particular sampling. On the one hand, the reduced size of log-polar images (as much as 30 times smaller than uniformly-sampled Cartesian images have been reported) hugely facilitates real-time visual data processing. On the other hand, the radially logarithmic sampling entails that a higher resolution is devoted to the center of the scene (fovea area) which, in turn, means that foveal information is represented by a big number of pixels in the log-polar image. One of the most interesting practical implications of this *foveal predominance*, is that foveated targets can be tracked without being explicitly segmented from the background. Additionally, the segmentation of a verged target in a binocular system becomes easier.

Biological plausibility: It approximates the receptive field distribution and retino-cortical mapping in the visual system of mammals. Depending on the application, this mimicry of biological solutions can be seen as an experimental support tool to neurophysiology, either to validate its findings or to propose new hypotheses. On the other hand, from an engineering point of view, it is a rich source of inspiration for sound strategies developed in the natural animal world.

Rotation and scaling "invariance": When the original image is rotated or scaled with respect to its center, patterns in the log-polar image only undergo translations, thus preserving their shape. This geometric property, also known as *edge* or *shape* invariance, is particularly helpful for rotation- and scale-invariant pattern recognition, but it has also been exploited for motion estimation in active tracking scenarios, as described in Section 5.

The first two properties (conformality and data selection) are common to other space-variant imaging models, while the last two (biological inspiration and edge invariance) are specific advantages of the log-polar mapping.

2.3. Singularity at the origin

One weakness of the $\log(z)$ model is the existence of a singularity in the center of the image, since it is not possible to evaluate $\log(0)$. This means that receptive fields would become infinitely small toward the origin. Therefore, points in the fovea cannot be represented with this mapping. Two solutions are commonly used to overcome this problem: either using a different mapping for the fovea (e.g. the identity mapping or a pure polar structure) or applying the $\log(z + a)$ model. This other model was proposed in [24] as a better approximation to the retino-topic mapping of monkeys and cats. The $\log(z + a)$ model transforms points in the first and fourth quadrants ($x \geq 0$) of the retinal plane via

$$w = \log(z + a), \quad (2)$$

where a is a positive real number. In coordinates, we have:

$$\xi' = \log\sqrt{(x+a)^2 + y^2}$$
$$\eta' = \arctan\left(\frac{y}{x+a}\right).$$

Because x is positive in the first quadrant and a is a positive constant, the minimum value of ξ' is finite ($\min \xi' = \log(a)$), therefore avoiding the singularity present in the $\log(z)$ model (Fig. 2(a,b)).

The mapping for the other quadrants is obtained by symmetry, which saves some computation time. Additionally, pixels in the log-polar plane are rearranged to match the quadrants of the Cartesian plane and to be connected at the origin (Fig. 2(c,d)). The resulting mapping is:

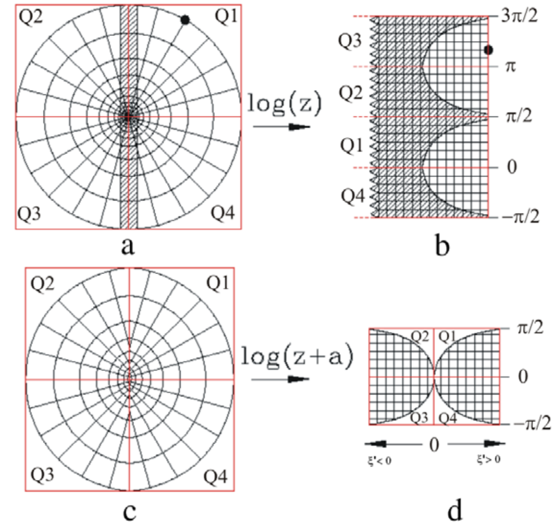

Fig. 2. The $\log(z + a)$ model [25] can be seen as removing the shaded region in (a) from the $\log(z)$ model. The new log-polar coordinates are designed to have connectivity at the origin and regions organized in the usual quadrants ($Q1$, $Q2$, $Q3$, $Q4$). Figure adapted from [25].
Source: With kind permission from Springer Science + Business Media: International Journal of Computer Vision, Volume 13, 1995, pages 75, R. Wallace, P. Ong, B. Bederson, and E. Schwartz, Fig. 3.

- first and fourth quadrants ($x \geq 0$):

$$\xi' = \log\sqrt{(x+a)^2 + y^2} - \log(a)$$
$$\eta' = \arctan\left(\frac{y}{x+a}\right)$$

- second and third quadrants ($x < 0$):

$$\xi' = -\log\sqrt{(-x+a)^2 + y^2} + \log(a)$$
$$\eta' = \arctan\left(\frac{y}{-x+a}\right).$$

The additional term $\log(a)$ produces a shift in the radial coordinates such that points at the origin result in $\xi = 0$, thus forcing the connectivity of the different quadrants. The use of the $\arctan(\cdot)$ function instead of $\text{atan2}(\cdot, \cdot)$ is chosen so that the angular range is $\left[\frac{-\pi}{2}, \frac{\pi}{2}\right]$, which promotes the arrangement of the quadrants in the usual sequence.

The $\log(z+a)$ model lacks the property of exact scale invariance. However, this is often tolerated in practical applications. Other solutions to the singularity problem have also been proposed [16,26–28].

2.4. Discretization

The $\log(z)$ and $\log(z + a)$ are conceptual models defined in continuous coordinates. They tell us how retinal and cortical coordinates are related, but, in practice, the mapping must be discretized. The conventional approach considers the cortical plane to be uniformly discretized as if it was an ordinary Cartesian image, i.e. covered with a dense grid of rectangular regions. Thus, let us consider a grid of $E \times A$ rectangular pixels (also known as *cortical cells*) whose corners are at coordinates $w_{p,q} = \xi_p + j\eta_q$, $p \in \{1, \ldots, E\}$, $q \in \{1, \ldots, A\}$, with E as the number of eccentricities (radial rings) and A the number of angular sectors. In the retinal plane, the corresponding regions, $\gamma_{p,q}$, are shaped like angular sections of concentric annuli. Such regions are called *retinal cells* or *receptive fields* (RFs). Fig. 1 illustrates the partition of the cortical plane with a uniform grid and the corresponding retinal cells.

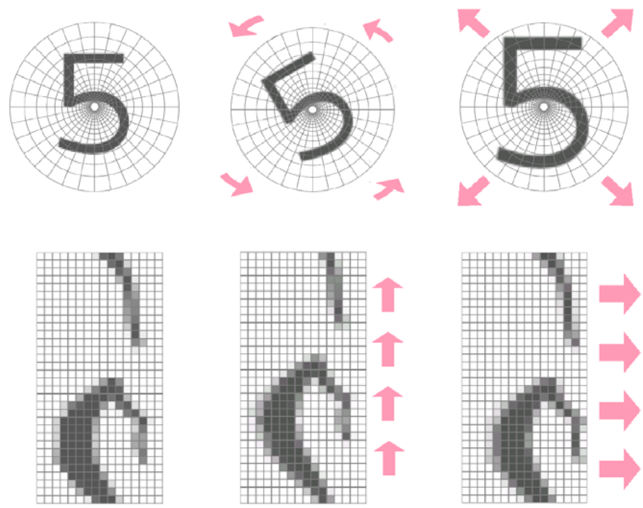

Fig. 3. Top: Retinal images on the log-polar mapping template. Bottom: The corresponding log-polar images. To illustrate the edge invariance property, the image on the left is rotated (middle) and scaled (right), which correspond to approximate translations in the log-polar domain, in the angular (η) and radial (ξ) directions, respectively, as shown with the arrows.

Each retinal cell has a weighing function $\phi_{p,q}(z)$ associated to it that represents the way the value of each cortical pixel is obtained from the information in the retinal array. This can be modeled by

$$c_{p,q} = \langle f, \phi_{p,q} \rangle, \quad (3)$$

where f is the function representing the Cartesian image. With uniform weighting functions, this operation represents the simple averaging of the retinal information within the spatial support of each retinal cell. Neighbor cells in the retinal domain are also neighbors in the cortical domain, except along the angular discontinuity and the radial singularity. Shape invariance to centered rotations and scalings no longer holds perfectly for the discretized $\log(z)$ model. However the approximation is good enough for practical applications, if discretization is not too coarse (Fig. 3).

2.5. Other implementation details

Models following biological data more closely have overlapping cells. They are computationally more expensive than non-overlapping ones but gain in smoothness of the log-polar image pixels. The models presented in [29] and [30] have cells with a log-polar distribution but with circular shapes and a moderate amount of overlap with the neighbors. In [29] a tessellation is proposed with a linear relation between receptive field size and eccentricity, and a cell overlap of 50%, as shown in Fig. 4(a). The proposal in [30] also uses circular receptive fields but attempts to minimize the amount of overlap between them by using a slightly different organization of receptive fields where direct neighbors are not in the same ring (Fig. 4(b)). Besides those seen above, there are other implications of log-polar sampling, as well as less obvious properties. For instance, discrete log-polar images and their internal representation introduce some practical difficulties, as the shapes and motions of the objects are distorted in a complex non-linear and inhomogeneous way. Appropriate algorithms and strategies have to be devised carefully to deal with these issues. Another example is the difference in resolution between the fovea and the periphery which calls for active strategies from the observers so that the gaze can be redirected according to the scene, the goals of the task and the ongoing visual events.

In terms of data structures in computer implementations, a log-polar image may not differ from a conventional image, except for the different meaning of the axes (row and column indices). In this case, one common difficulty to be tackled refers to the circularity of the angular axis, since the two ends of the rectangular log-polar image have to be processed as if they were really connected. To solve these kinds of technical complexities, an alternative representation, the connectivity graph, was proposed [32]. It has the advantage of being very general and able to accommodate arbitrary sensor lattices, as illustrated in Fig. 5. Local image operations, such as edge detection, can therefore be applied without special cases (e.g. image boundaries) in mind. As a disadvantage, graphs have to be defined, built, and processed, which may result in a loss of efficiency in particular circumstances.

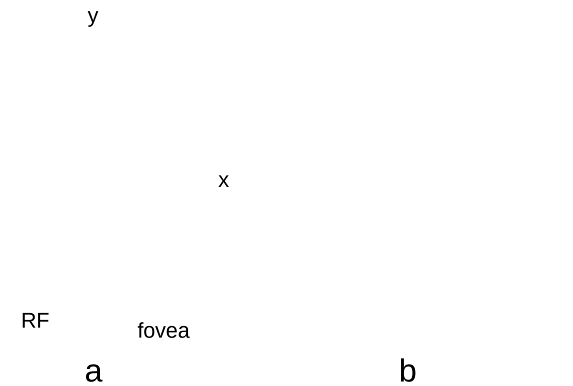

Fig. 4. Overlapping models: (a) the model of [29] implemented in [31], and (b) the model in [30] (figure taken from [16]).[2]

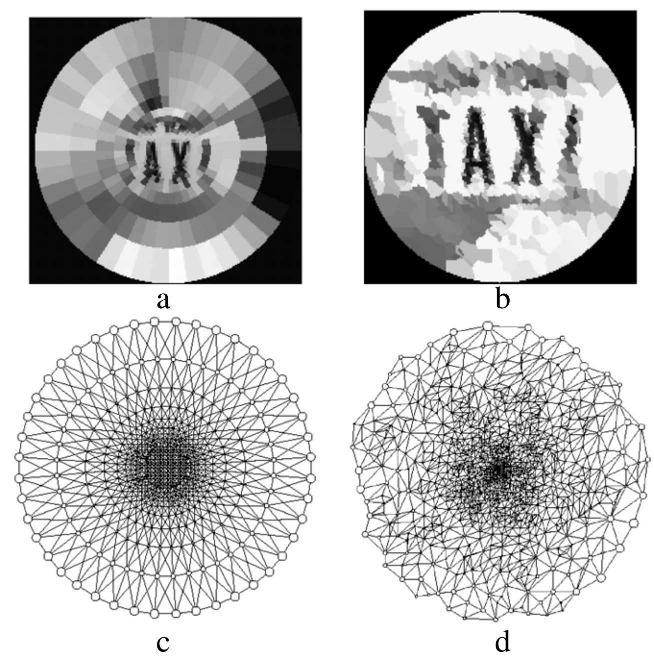

Fig. 5. The connectivity graph (CG) [33,34]: (a, b) sensors with (a) a log-polar and (b) an arbitrary pixel arrangement; (c, d) their corresponding CGs.
Source: With kind permission from Springer Science + Business Media: International Journal of Computer Vision, Volume 13, 1995, pages 77, 79, R. Wallace, P. Ong, B. Bederson, and E. Schwartz, Fig. 4 and Fig. 6.

[2] *Source:* Reprinted from Computer Vision and Image Understanding, Vol. 69, No. 2, M. Bolduc and M. D. Levine, A review of biologically motivated space-variant data reduction models for robotic vision, Pages 170–184, Copyright (1998), with permission from Elsevier.