

Anthropomorphic Visual Sensors

Fabio Berton, Giulio Sandini, Giorgio Metta

LIRA-Lab, DIST, Università di Genova, Genova, Italy

CONTENTS

1. Introduction and Motivations
2. The Log-polar Mapping
3. The Log-Polar Sensor
4. Mathematical Properties of the Log-Polar Transform
5. Applications
 - Glossary
 - References

1. INTRODUCTION AND MOTIVATIONS

The tentative of reproducing a biological eye with artificial electronic devices has always had a major drawback. In fact there is an important difference between how a human eye sees the world and how a standard video camera does: while the common visual sensors generally have constant resolution on each part of the image, in their biological counterparts the picture elements are arranged in order to get a very high amount of information in the central part of the field of view (the so called *fovea*) and a gradually decreasing density of photoreceptors while approaching the borders of the receptive field. There is a practical motivation behind this evolutionary choice: a human being needs both a high resolution in order to distinguish between the small details of a particular object for fine movements (the human eye has actually a maximum resolution of about 1/60 degrees) and, at the same time, a large enough field of view (i.e., about 150 degrees horizontally and about 120 vertically for the human eye) so to have a sufficient perception of the surrounding environment. With a constant resolution array of sensors, this two constraints would have increased the total number of photoreceptors to an incredibly high value, and the consequence of this would have been the need of other unrealistic features such as an optic nerve having a diameter of few centimeters (the actual human optic nerve diameter is about 1.5 mm), in order to transfer this amount of data, and a much bigger brain (weighting about 2300 kg, compared to about 1.4 kg of our brain) in order to process all this information, not considering the huge power requirements of such a big brain.

Since a zooming capability would have implied a renounce to the simultaneity of the two features, the evolution has answered to the question on how to optimally arrange a given number of photoreceptors over a finite small surface. A lot of different eyes evolved with the disposition of the photoreceptors adapted to the particular niche. Examples of this diversity can be found in the eyes of insects (see, for example, [1] for a review) and in those of some birds that have two foveal regions to allow simultaneous flying and hunting ([2, 3]).

In the human eye (Fig. 1.1) we have a very high density of cones (the color sensitive photoreceptors) in the central part of the retina and a decreasing density when we move towards the periphery. The second kind of receptors (the rods, sensitive to luminance), are absent in fovea, but they have a similar spatial distribution. In fact the cone density in the foveola (the central part of the fovea) is estimated at about 150–180,000 cones/mm² (see Fig. 1.1). Towards the retinal periphery, cone density decreases from 6000 cones/mm² at a distance of 1.5 mm from the fovea to 2500 cells/mm² close to the *ora serrata* (the extremity of the optic part of the retina, marking the limits of the percipient portion of the membrane). Rod density peaks at 150,000 rods/mm² at a distance of about 3–5 mm from the foveola. Cone diameter increases from the center (3.3 μm at a distance of 40 μm from the foveola) towards the periphery (about 10 μm). Rod diameter increases from 3 μm at the area with the highest rod density to 5.5 μm in the periphery [4].

Since this sensor arrangement has been proven by the evolution to be an efficient one, we tried to investigate how this higher efficiency could be translated in the world of artificial vision. From the visual processing point of view we asked on one hand whether the morphology of the visual sensor facilitates particular sensorimotor coordination strategies, and on the other, how vision determines and shapes the acquisition of behaviors that are not necessarily purely visual in nature. Also in this case we must note that eyes and motor behaviors coevolved: it does not make sense to have a fovea if the eyes cannot be swiftly moved over possible regions of interest (active vision). Humans developed a sophisticated oculomotor apparatus that includes saccadic movements, smooth tracking, vergence, and various combinations of retinal and extra-retinal signals to maintain

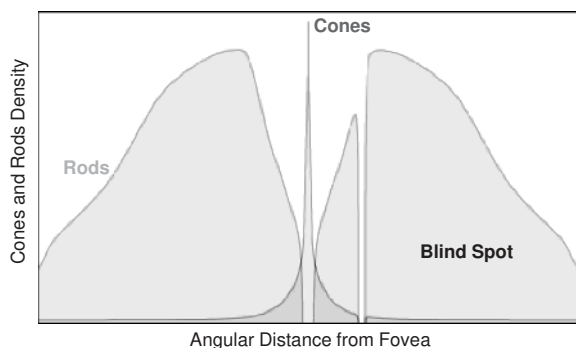


Figure 1.1. Cones and Rods Density—The density of the cones is maximum in the center of the fovea, and rapidly decreases towards the periphery. The rods are absent in fovea, but have a similar decreasing property.

vision efficient in a wide variety of situations (see [5] for a review).

Since the coupling of this space variant structure (Fig. 1.2) with an active vision system gives the capability of seeing the regions of interest always with the best available quality, it is expected that the transposition of this principles to an artificial eye would be extremely efficient as well.

This addresses the question of why it might be worth copying from biology and which are the motivations for pursuing the realization of biologically inspired artifacts. How this has been done is presented in the following sections where we shall talk about the development of a retina like camera. Examples of applications are also discussed in the field of image transmission and robotics. The image transmission problem resembles the issue of the limitation of bandwidth/size of the optic nerve discussed above. The limitations in the case of autonomous robots are in terms of computational resources and power consumption.

2. THE LOG-POLAR MAPPING

While the arrangement of photoreceptors on the human retina is quite complex and irregular, there are some approximations that are able to reproduce well their disposition. The simple mathematical mapping that best fits this space variant sensor structure is the log-polar mapping, that is also called, for this reason, retina like mapping [6–8].

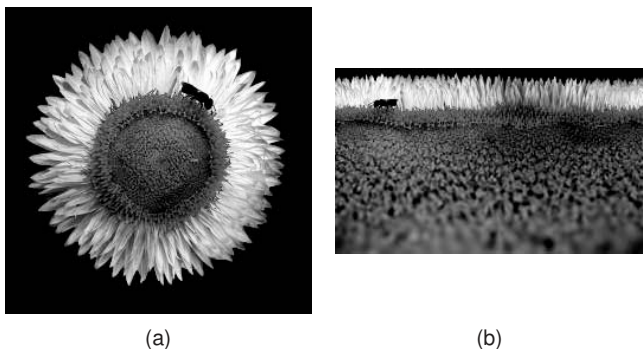


Figure 1.2. Biological eye: (a) The real world (courtesy Tom Dorsey/Salina Journal) and (b) how the human eye perceives it.

The log-polar transform is defined by the function [9]:

$$w = f(z) = \log_{\lambda}(z) \quad (2.1)$$

where both z and w are complex variables.

While

$$z = x + iy = r(\cos \phi + i \cdot \sin \phi) \quad (2.2)$$

is the representation of a point in the cartesian domain,

$$w = \rho(z) + i \cdot \theta(z) \quad (2.3)$$

represents the position of a point on the log-polar plane.

The equation (2.1) can also be written as:

$$\begin{cases} \rho(r, \theta) = \log_{\lambda} r \\ \theta(r, \theta) = h \cdot \phi \end{cases} \quad (2.4)$$

Equations (2.4) point out the *log*-arithmic and the *polar* structure of the mapping. In fact ρ is directly linked to $\log_{\lambda} r$ and θ is proportional to ϕ when considering a traditional polar coordinates system defined by:

$$\begin{cases} r = \sqrt{x^2 + y^2} \\ \phi = \arctan \frac{y}{x} \end{cases} \quad (2.5)$$

In the Fig. (2.1)a it is possible to see how the regions of the rectangular grid in the log-polar domain are arranged, in cartesian coordinates (x, y) , in concentric rings whose width is proportional to their distance from the origin of the mapping. Each ring is then divided in a certain number of receptive fields, each one corresponding to a pixel in the cortical image 2.1b.

The receptive fields that lie on a radius are mapped on vertical lines on the plane (ρ, θ) , while those lying on concentric circumferences centered on the origin are mapped on horizontal lines. The origin of the plane (x, y) has no corresponding points on the plane (ρ, θ) , because of the singularity that the complex logarithm function has in this point and, theoretically, an infinite number of concentric rings exist below any radius having a finite size. So a more

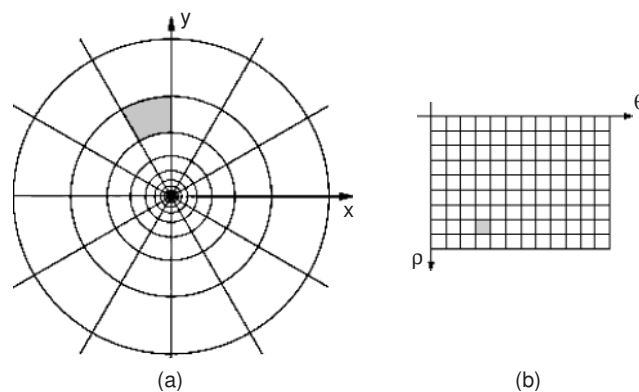


Figure 2.1. Log-polar Transform: (a) The grid in cartesian coordinates and (b) Its transform. The areas marked in gray are each other's transformation.

common expression for the log-polar transform is the following:

$$\begin{cases} \rho(r, \phi) = \log_{\lambda} \frac{r}{r_0} \\ \theta(r, \phi) = h \cdot \phi \end{cases} \quad (2.6)$$

The logarithm base λ is determined by the number of pixels we want to lay on one ring, and by their shape.

While the qualitative shape of a pixel is fixed (it is the intersection between an annulus and a circular sector, see Fig. 2.2, its proportions may vary. If we require the photoreceptor's shape to be approximately square, we can state that the length of the segment **BC** should be equal to the length of the arc **DC** or of the arc **AB**.

In the first case the base of the logarithm will be:

$$\lambda = \frac{2\pi + N}{N} \quad (2.7)$$

while in the second case we will have:

$$\lambda = \frac{N}{N - 2\pi} \quad (2.8)$$

Usually, when handling real situations, due to technological constraints, the size of a pixel in an actual sensor cannot boundlessly decrease, so the mapping has to stop when the size of the receptive field approximates the size of the smallest pixel allowed by the used technology. The consequence is that there is an empty circular area in the center of the field of view where the space variant sensor is blind. This lack of sensibility can be avoided by filling this area with uniformly spaced photoreceptors, whose size is equal to the size of the smallest pixel in the logarithmic part of the sensor. The structure of the whole sensor then will have a central part, i.e., the fovea, where the polar structure is still preserved and the space variant resolution (the logarithmic decreasing) is lost. This choice automatically determines the value of the shift factor r_0 in (2.6). In fact, if we decided to arrange i rings in fovea (with indices $0, \dots, i-1$), we will have to assign an index i to the first *logarithmic* ring. Since varying the value of r_0 we are able to overlap any chosen ring of the *logarithmic* part of the sensor to the i th ring of the fovea, r_0 is then determined.

Then the most general equation which describes the log-polar transform introduces in equation (2.6) a proportional constant k in order to scale the mapping to cover a desired

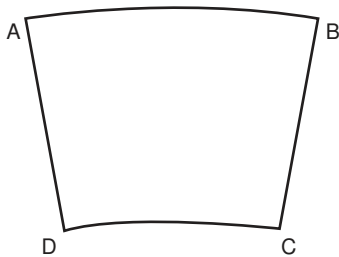


Figure 2.2. Log-polar pixel.

area, while keeping the total number of log-polar pixels, and the previously mentioned shift constant r_0 :

$$\begin{cases} \rho(r, \phi) = k \cdot \log_{\lambda} \frac{r}{r_0} \\ \theta(r, \phi) = h \cdot \phi \end{cases} \quad (2.9)$$

The equations needed to perform a log-polar transform (Eq. 2.10), and its anti-transform (2.11) then will be:

$$\begin{cases} \rho(x, y) = k \cdot \log_{\lambda} \frac{\sqrt{x^2 + y^2}}{r_0} \\ \theta(x, y) = h \cdot \arctan \frac{y}{x} \end{cases} \quad (2.10)$$

and:

$$\begin{cases} x(\rho, \theta) = r_0 \cdot \lambda^{\rho/k} \cdot \cos \frac{\theta}{h} \\ y(\rho, \theta) = r_0 \cdot \lambda^{\rho/k} \cdot \sin \frac{\theta}{h} \end{cases} \quad (2.11)$$

Both the previous equations are valid for the logarithmic part of the mapping, while in fovea the following equations are used:

$$\begin{cases} \rho(x, y) = k \cdot \sqrt{x^2 + y^2} \\ \theta(x, y) = h \cdot \arctan \frac{y}{x} \end{cases} \quad (2.12)$$

and:

$$\begin{cases} x(\rho, \theta) = \frac{\rho}{k} \cdot \cos \frac{\theta}{h} \\ y(\rho, \theta) = \frac{\rho}{k} \cdot \sin \frac{\theta}{h} \end{cases} \quad (2.13)$$

Generally the choice of the various parameters, especially when handling the case of a discrete log-polar transform, is crucial. It deeply affects, in fact, the quality of the images, the compression factor, the size and the relevance of various artifacts and false colors [10]. Peters [11] has investigated a method in order to optimally choose the parameters of the mapping.

3. THE LOG-POLAR SENSOR

In the 1980s various researchers have begun to transfer this efficient idea from the biological world to the artificial one, so to take advantage of the reduced cost of the acquisition of the images. As we stated before, this means lower power consumption, lower number of pixels to be acquired and shorter processing time. Moreover such a family of sensors is important for an artificial being that mimics a complex system such as the human body.

Various approaches have been tried in order to build a visual system capable to produce foveated images, but each choice presented some drawbacks. The first, and more intuitive solution has been a standard CCD video camera connected to a workstation with a frame grabber and a software, which was able to perform the log-polar transform. Unfortunately the technology that was available at that time did

not allow a software simulation of a log-polar mapping or, at best, when this was possible, its performances were not comparable with those of a solid-state device. So the next step has been the design of a dedicated hardware (see, for example, [12–16]) that was able to speed up the remapping process but the maximum resolution allowed was limited by the resolution of the traditional camera, not considering the additional cost of the electronic board and the need of an acquisition of a number of pixel which was much higher than the one in the output image. Another approach that permitted to get a foveated image has been the use of distorting lenses which were able to enlarge the central part of the image while keeping the peripheral region unchanged [17, 18], but in order to get an actual log-polar transform some additional operations were needed. So it has been decided to design an implementation in silicon of a biologically inspired completely new visual system.

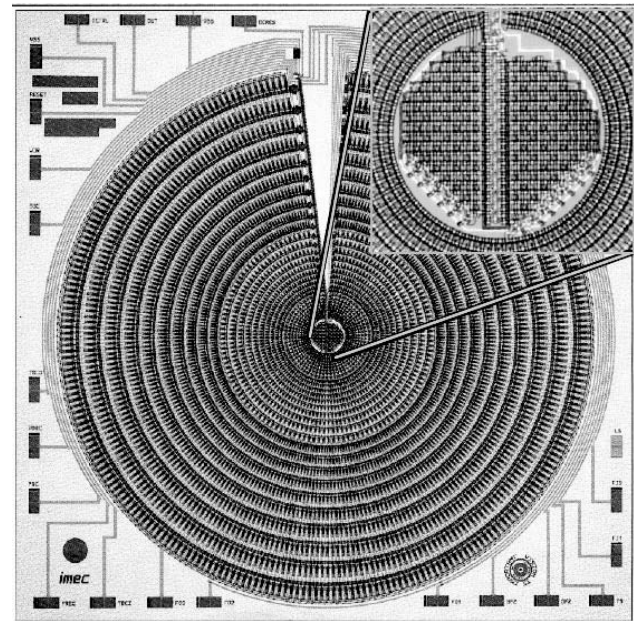
Besides our realizations a few other attempts have been reported in the literature on the implementation of solid-state retina like sensors [19–21]. So far we are not aware of any commercial device, besides those described here, that have been realized based on log-polar retina like sensors.

3.1. The 2000 pixels CCD Sensor

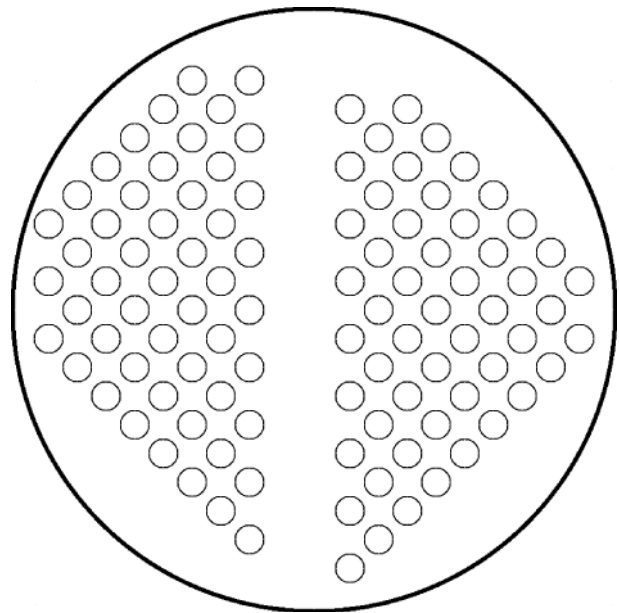
Our first implementation of a solid-state foveated retina like sensor has been realized in the beginning of the 1990s using a $1.5\ \mu\text{m}$ CCD technology [22]. At that time it was the state of the art technology and it allowed a size of the smallest possible pixel (i.e., a pixel in the foveal part of the image) of about $30\ \mu\text{m}$, while the diameter of the whole sensor, for practical reasons was limited to 9.4 mm. This sensor was composed of 30 rings, and each ring was covered by 64 pixels, for a total of 1920 pixels in the log-polar part of the sensor (Fig. (3.1)a), to which we still have to add 102 more pixels covering the fovea, for a total of 2022 elements. The fovea was covered by a uniform grid of square pixels arranged on a pseudo lozenge, roughly an 11×11 cartesian grid with missing corners and one diagonal, like in Fig. (3.1)b. Since the polar structure was not preserved in the fovea, a major discontinuity was present on the border between the two regions of the sensor.

Since the size of the largest pixel in this first CCD implementation was about $412\ \mu\text{m}$, the ratio between the largest and the smallest pixels (R) was about 13.7. This parameter describes the amount of “space variancy” of the sensor and, of course, it is equal to 1 in the standard cartesian sensors with constant resolution.

Another important parameter in space variant sensors is the ratio between the total area of the sensor and the area of the smallest pixel (or, alternatively, between the diameter of a pixel and the diameter of the whole sensor). Rojer and Schwartz [14] defined this value as Q , as a measure unit of the *spatial quality* of a space variant sensor. In order to have a reference term, consider that Q is obviously equal to the size of the sensor (measured in pixels) in a constant resolution array, while it can have values very close to 10,000 in the human retina. The importance of this parameter can be understood by observing that its value represents the square root of the number of pixels we would need to cover the space variant sensor with a constant resolution grid having



(a)



(b)

Figure 3.1. 2000 pixels CCD log-polar sensor: (a) Picture of the whole CCD log-polar sensor. (b) Detail of the fovea.

the same maximum resolution of the log-polar sensor (and the same field of view).

Rojer and Schwartz also proved that Q exponentially increases with the total number of pixels in the log-polar sensor, so the addition of few more rings gives a much better value of Q .

For the retina like CCD sensor, the parameter Q is equal to about 300, meaning that if we want to simulate electronically a log-polar camera starting from a traditional one, the latter should have a sensor which is able to acquire at least a 300×300 square image, in order to obtain the same

amount of information that the solid state retinal sensor could obtain directly.

This first sensor has been the first solid state device of this kind in the world, but presented some drawbacks mostly related to the use of the CCD technology itself, like the difficulty of properly addressing the pixels, which caused the presence of some blind areas on the sensor (one in the fovea, along a diameter and a circular sector in the periphery, which is about 14 degrees wide (or 2.5 pixels), both shown in Fig. 3.1).

3.2. The 8000 pixels CMOS Sensor

The next generation of the sensor had some very important new features. First of all, the evolution of the technology allowed the construction of a much smaller pixel, and consequently a significantly greater number of photoreceptors could be fitted in a sensor having roughly the same size of the CCD version. Moreover, in order to avoid some of the problems that afflicted the previous version, we decided to move to the CMOS technology. Since the addressability in this new sensor was much simpler, no blind areas were present on the surface. The chosen pixel for the 8000 points sensor was the FUGA model from IMEC (now Fill Factory), in Loewen, Belgium. The pixels distinguished themselves from classical CCD or CMOS sensors because of their random addressability and logarithmic photoelectric transfer. The random addressability allowed, in theory, to read just some previously chosen parts of the sensor, while the logarithmic behavior allowed the sensor to correctly acquire images even in extreme light conditions. The logarithmic response yielded a dynamic range beyond six decades (120 dB), by log-compression of the input optical power scale onto a linear output voltage scale: the new pixel could view scenes with vastly different luminances in the same image, without even the need of setting an exposure time. It can be useful to note that this behavior mimics very well that of the eyes in similar conditions.

The major drawback of this technology was the introduction of the so-called *fixed pattern noise* (FPN), which is common to all the CMOS visual sensors. The continuous-time readout of a logarithmic pixel implied the impossibility of the removal of static pixel-to-pixel offsets. As a result, the raw image output of such a sensor contained a large overall non-uniformity, often up to 50 or 100% of the actual acquired signal itself. However, things were not dramatic, since the FPN is almost static in time, so it can be nearly completely removed by a simple first order correction.

The second new feature was the preservation of the polar structure in the fovea. Although the number of pixels per ring was not constant anymore, the presence of a polar arrangement minimized the effect of the discontinuity fovea-periphery. The fovea was structured with a central pixel, then a ring with 4 pixels, one with 8, 2 with 16, 5 with 32 and 10 with 64 pixels per ring, for a total number of 845 pixels on 20 rings. Starting from the 21st ring, the logarithmic increasing was applied (Fig. 3.3).

The third macroscopic difference compared to the CCD sensor was that this time a color version of the chip had been produced. Since a pixel, normally, is sensitive to just one wavelength, the color had to be reconstructed for each

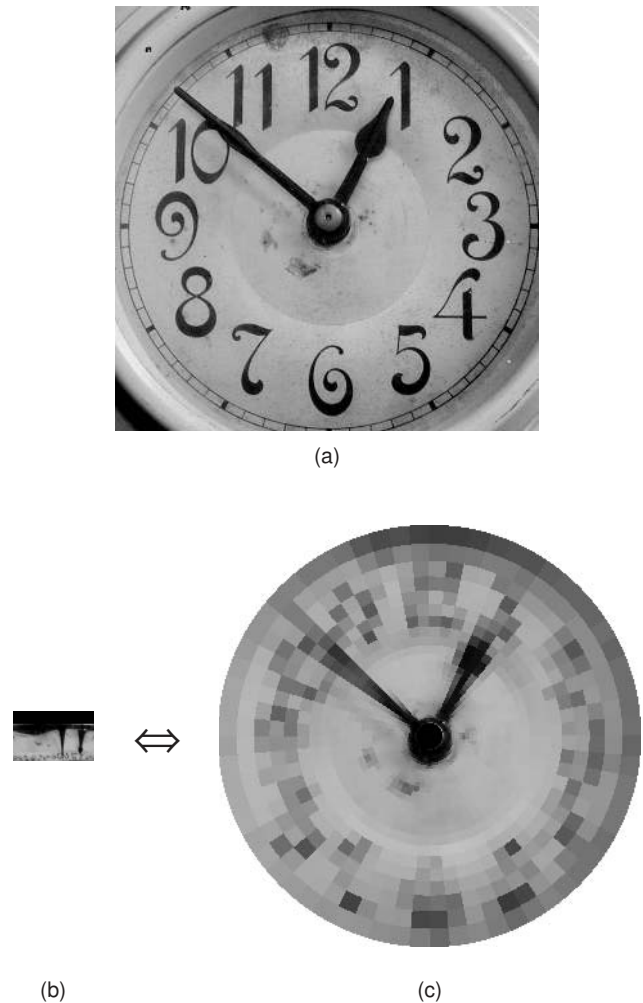


Figure 3.2. 2000 pixels CCD Log-polar Sensor Simulation: (a) A standard cartesian image. (b) Its log-polar transform performed by this sensor. (c) The remapped image. Please note that in this figure, and in the next Figs. 3.4 and 3.6, the log-polar image and its transform are not in scale. In all these images the fovea is not displayed.

photosite by interpolating the outputs of the neighboring pixels. This is a common operation, needed on standard constant resolution arrays as well, and various patterns have been investigated in order to minimize the appearance of false colors, but when dealing with the large pixels close to

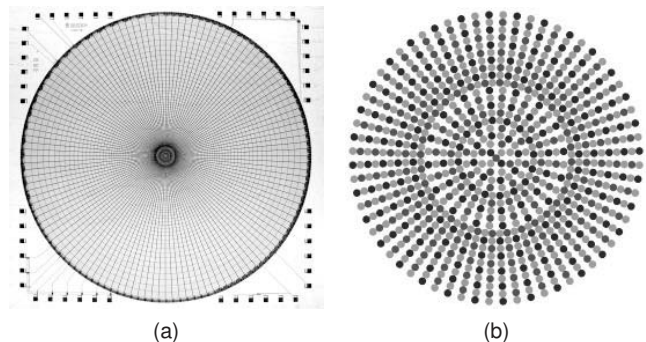


Figure 3.3. CMOS 8000 pixel Log-polar sensor: (a) Picture of the 1st version of the CMOS log-polar sensor. (b) Detail of the fovea.

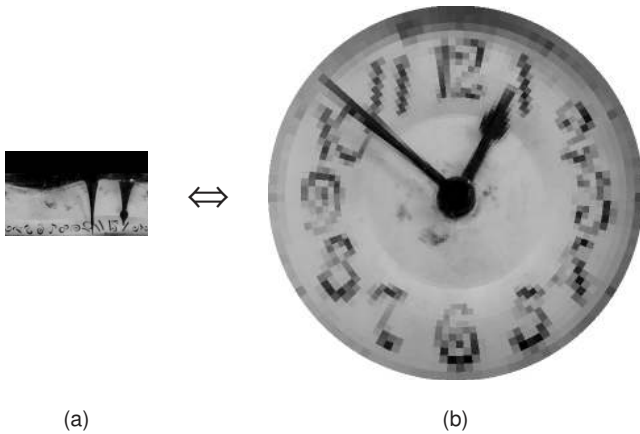


Figure 3.4. 8000 pixels CMOS Log-polar sensor simulation: (a) The log-polar transform of the image 3.2a performed by this sensor. (b) The remapped image.

the external border (low spatial frequency), and when the image presents very high spatial frequencies, it happens that adjoining pixels “see” different elements of the image, and the consequence is that the color components used for the reconstruction belong to uncorrelated object and that causes the false colors.

3.3. The 33000 pixels CMOS Sensor

In the late 1990s the progress in the field of silicon manufacturing allowed us to design a new chip using a $0.35 \mu\text{m}$ technology. This project had been developed within a EU funded research project called SVAVISCAs. The goal of the project was to realize, besides the improved version of the sensor, a micro camera with special purpose lens allowing a 140 degrees field of view. The miniaturization of the camera was possible because some of the electronics required to drive the sensor as well as the A/D converter were included in the chip itself.

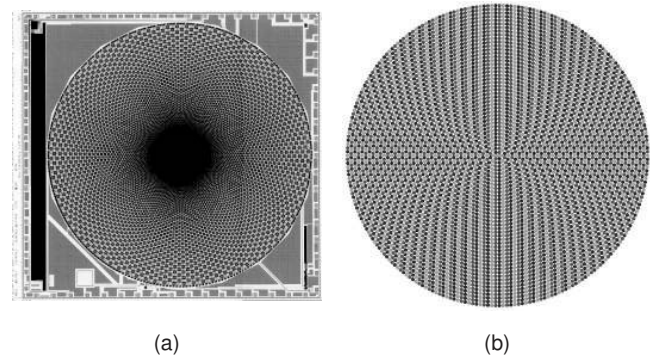


Figure 3.5. CMOS 33000 pixel Log-polar sensor: (a) Picture of the 2nd version of the CMOS log-polar sensor. (b) Detail of the fovea.

The main novelty introduced by this sensor has been the structure of the fovea, which was, this time, completely filled by pixels. The 42 rings inside the fovea, while still having a variable number of pixel each, had a constant decrease of this parameter: so, not considering the ring 0, which was just a single pixel, the other rings had a number of pixel which was proportional to their distance from the center of the sensor (Fig. 3.5).

A minor change in the layout was the adoption of a *pseudo triangular* tessellation, in order to minimize the artifacts introduced by the low spatial frequency of the periphery. It means that each even ring has been rotated by half a pixel (about 0.7 degrees) respect to the odd rings. This had the consequence that the average distance of a single point from the closest red, green and blue pixel was inferior to the one we had with the *square* tessellation, then it allowed a more accurate color reconstruction.

Since the parameter Q in this version is about 1100, for the first time the Log-polar image acquired by the camera was “better” than the images produced by the standard “off the shelf” video cameras available at that time. Better means

Table 1. A comparison between three generations of log-polar sensors.

Sensor Version	Total Number of Pixels	Pixels in Fovea	Pixels in Periphery	Total Number of Rings	Angular Amplitude	Logarithm Base	Technology Used	Radius of the Fovea
CCD	2022	102	1920	30	5.413°	1.094	$1.5 \mu\text{m}$	$317 \mu\text{m}$
CMOS 8k	8013	845	7168	76	2.812°	1.049	$0.7 \mu\text{m}$	$285 \mu\text{m}$
CMOS 33k	33193	5473	27720	152	1.428°	1.02337	$0.35 \mu\text{m}$	$273 \mu\text{m}$

Sensor Version	Rings in Fovea	Rings in Periphery	Angular Amplitude	Logarithm Base
CCD	—	30	5.413°	1.094
CMOS 8k	20	56	2.812°	1.049
CMOS 33k	42	110	1.428°	1.02337

Sensor Version	\varnothing of the Sensor	Size of the Smallest Pixel	R	Q	Technology Used	Radius of the Fovea
CCD	$9400 \mu\text{m}$	$30 \mu\text{m}$	13.7	300	$1.5 \mu\text{m}$	$317 \mu\text{m}$
CMOS 8k	$8100 \mu\text{m}$	$14 \mu\text{m}$	14	600	$0.7 \mu\text{m}$	$285 \mu\text{m}$
CMOS 33k	$7100 \mu\text{m}$	$6.5 \mu\text{m}$	17	1100	$0.35 \mu\text{m}$	$273 \mu\text{m}$

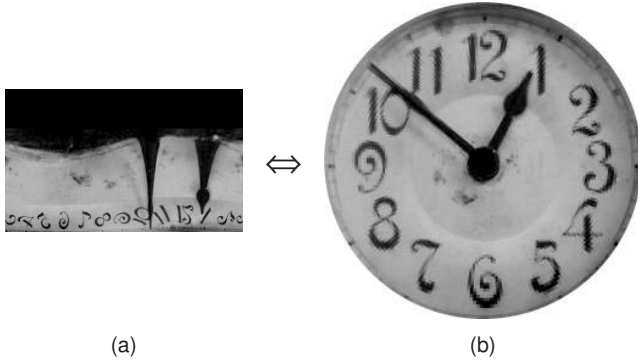


Figure 3.6. 33000 pixels CMOS Log-polar sensor simulation: (a) The log-polar transform of the image 3.2a performed by this sensor. (b) The remapped image. Although the remapped images are not in scale, the log-polar ones are in scale between them.

that the remapped log-polar image had a better maximum resolution than its cartesian counterpart, but using just about one thirtieth of the pixels. This allowed sending video streams on low bandwidth channels at a very high frame rate, using standard existing compression algorithms. It was even possible to transmit a video stream over a GSM cell phone channel, which is a 9600 bit per second channel, achieving a frame rate between 1 and 2 frames per second.

3.4. A Comparison Between the Log-Polar Sensors

In order to have a better understanding of the evolution of the log-polar sensor, it is useful to compare the main characteristics of the three versions.

The increasing of the quality of the image is evident from the Table 1, and it is shown in the simulated images in Figs. 3.2, 3.4, and 3.6: while the numbers in the clock are completely unreadable with the 2000 pixels sensor, they are much more defined in the last versions.

4. MATHEMATICAL PROPERTIES OF THE LOG-POLAR TRANSFORM

The log-polar transform is not only the best compromise between biological motivation and simplicity of the mapping, but it also presents some interesting mathematical features which can be used for a more efficient implementation of many image processing algorithms.

4.1. Conformity

The log-polar mapping is conformal. A conformal mapping, also known as conformal transformation, angle-preserving transformation, or biholomorphic map, is a transformation that preserves local angles [23]. An analytic function is conformal at any point where it has a nonzero derivative. Conversely, any conformal mapping of a complex variable that has continuous partial derivatives is analytic.

The demonstration of the angle preserving property comes from the fact that if we consider a complex function:

$$w = f(z) = \log_{\lambda}(z) \quad (4.1)$$

where z is a complex number in the cartesian domain that is $z = x + jy$, and w is a complex number in the log-polar domain which is $w = \rho + j\theta = |w| \cdot e^{j \cdot \arg(w)}$, then we get:

$$w = \log(|z| \cdot e^{j \cdot \arg(z)}) = \log|z| + j \cdot \arg(z) \quad (4.2)$$

If we set:

$$\begin{cases} \delta w = w - w_0 \\ \delta z = z - z_0 \end{cases} \quad (4.3)$$

and we consider the Taylor expansion of $f(z)$ centered in z_0 , then we get:

$$\begin{aligned} f(z) &= f(z_0) + \sum_{n=1}^{\infty} \frac{f^{(n)}(z_0)}{n!} \cdot (\delta z)^n \\ \Rightarrow f(z) - f(z_0) &= \sum_{n=1}^{\infty} \frac{f^{(n)}(z_0)}{n!} \cdot (\delta z)^n = \delta w \end{aligned} \quad (4.4)$$

If we stop at the first order approximation, then:

$$\delta w \approx f'(z_0) \cdot \delta z \quad \text{when } f'(z_0) \neq 0 \Rightarrow \frac{1}{z_0} \neq 0 \quad (4.5)$$

So, when $\delta \rightarrow 0$, $\arg(\delta w) = \arg(\delta z) + \arg(f'(z_0))$, and if we set $\arg(\delta z) = \theta$ and $\arg(f'(z_0)) = \alpha$, then $\arg(\delta w) = \theta + \alpha$.

This means that the segments in the cartesian plane are just rotated (locally) by an angle α respect to the same segments in the log-polar domain. Furthermore the segment δz is scaled in length by $f'(z_0) = 1/z_0$.

The preservation of angles comes consequently, in fact, if we set:

$$\begin{cases} \delta z_1 = \theta_1 \\ \delta z_2 = \theta_2 \end{cases} \quad (4.6)$$

then:

$$\begin{aligned} \arg(\delta w_2) - \arg(\delta w_1) &= (\theta_2 + \alpha) - (\theta_1 + \alpha) \\ &= \arg(\delta z_2) - \arg(\delta z_1) \end{aligned} \quad (4.7)$$

So, when the derivative in z_0 is not zero, which is always true for the logarithmic function, the angles are preserved (Fig. 4.1) and then the mapping is conformal.

The preservation of angles implies the preservation of the proximity (i.e., a couple of pixels which are close in the cartesian domain are still close in the log-polar one if we do not consider the discontinuities that occur when $\theta = 0$, $\theta = 2\pi$ and $\rho = 0$). These two peculiarities have the consequence that any algorithm involving local operators can be applied to the log-polar image without any significant change (angle detection, edge detection, compression etc.) [8]. In order to avoid the problems introduced by these discontinuities, a graph theory based approach has been investigated, defining a connectivity graph matching the morphology of the sensor [15, 24].

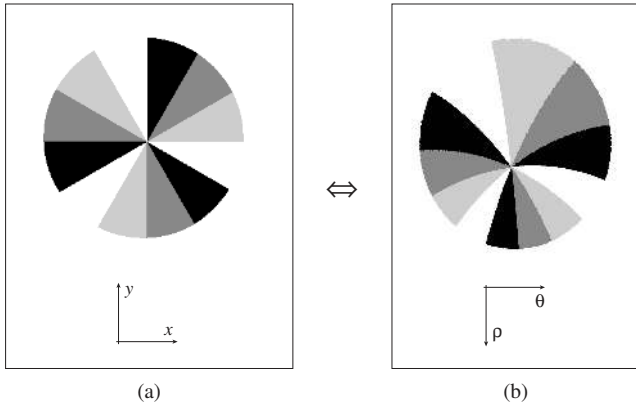


Figure 4.1. Angle Preservation: The angles are locally preserved after a log-polar Transformation. (a) Cartesian domain. (b) log-polar domain. Please note that both images a and b are particulars, so the origin of the mapping falls outside of the cartesian image.

4.2. Scale Change

One interesting and useful property of the log-polar transform is its invariancy to scale change. A pure scale change in the cartesian domain can be seen as a transformation where all the vectors representing each pixel of an object are transformed in another set of vectors each proportional by a common constant k to its pre-transformation counterpart. A scale change in the cartesian domain referred to the center of the log-polar mapping is then denoted by:

$$\begin{aligned} w = \log(kz) &\Rightarrow w = \log(|kz| \cdot e^{j \arg(kz)}) \\ &= \log |z| + \log k + j \arg(z) \quad (4.8) \end{aligned}$$

with $k \in \mathfrak{R}^+$.

This is equal to $\log(z) + K$, with the constant $K = \log_\lambda k$, which is a pure translation along the ρ axis in the log-polar domain (which is vertical in our representation, Fig. 4.2).

The invariance to scale change can be very important and useful in applications where the camera moves along its optical axis, such as the time to impact detection in a mobile vehicle.

4.3. Rotation

Another property is the invariancy to rotation. A pure rotation in the cartesian domain can be seen as a transformation where all the vectors representing each pixel of an object are transformed in another set of vectors whose phase is increased by a common constant ϕ compared to its pre-transformation counterpart. A rotation in the cartesian domain referred to the center of the log-polar mapping is then denoted by:

$$\log(|z| \cdot e^{j \arg(z) + \phi}) = \log |z| + j(\arg(z) + \phi) \quad (4.9)$$

this is equal to $\log(z) + \theta_0$, with the constant $\theta_0 \in \text{Im}$, which is a pure translation along the θ axis in the log-polar domain (horizontal in our representation, Fig. 4.3).

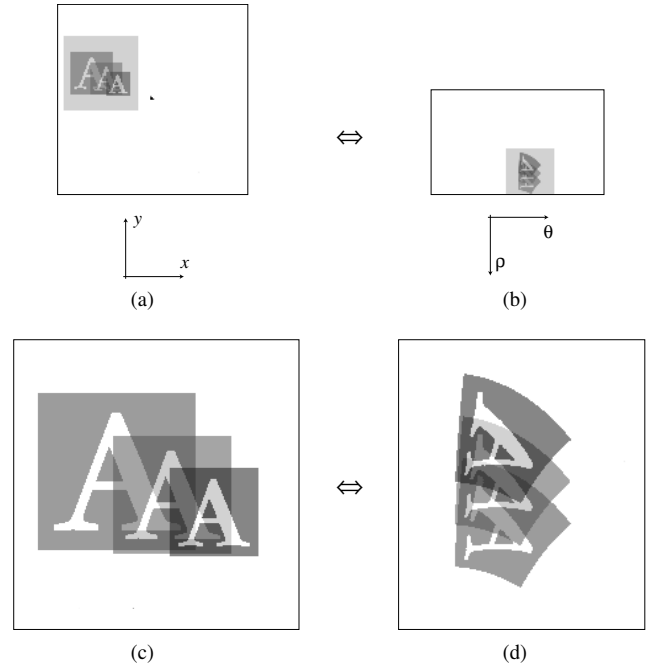


Figure 4.2. Scale Change: A pure scale change referred to the origin of the mapping, with no translational components, becomes a pure translation after the log-polar transform. (a) Cartesian domain. (b) log-polar domain. (c), (d) Enlargement of the shaded areas respectively in (a) and (b).

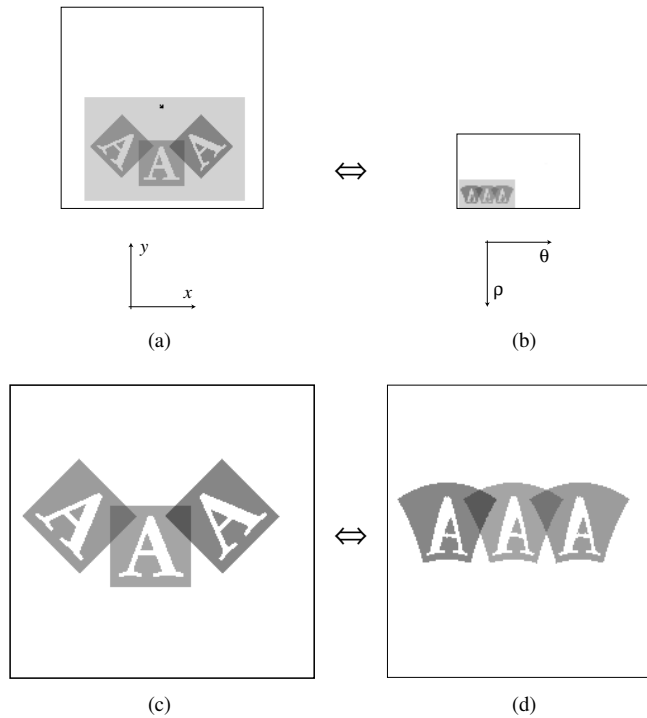


Figure 4.3. Rotation: A pure rotation referred to the origin of the mapping, with no translational components, becomes a pure translation after the log-polar transform. (a) Cartesian domain. (b) log-polar domain. (c), (d) Enlargement of the shaded areas respectively in (a) and (b).

4.4. Translations

While a translation in the cartesian space can be performed without any change in the shape of an object, this is no longer true in the log-polar domain.

In fact, given a point $P(x_0, y_0)$, its transform is

$$P_{LP}(\rho_0, \theta_0) = P_{LP} \left(\frac{1}{2} \log_{\lambda} \left[\frac{x_0^2 + y_0^2}{r_0} \right], \arctan \left[\frac{y_0}{x_0} \right] \right) \quad (4.10)$$

this point, after a translation, becomes:

$$\begin{aligned} P(x_0 + \Delta x, y_0 + \Delta y) \\ &= P_{LP} \left(\frac{1}{2} \log_{\lambda} \left[\frac{(x_0 + \Delta x)^2 + (y_0 + \Delta y)^2}{r_0} \right], \arctan \left[\frac{y_0 + \Delta y}{x_0 + \Delta x} \right] \right) \\ &= P_{LP}(\rho_0 + \xi_{\rho}(x_0 + \Delta x, y_0 + \Delta x), \theta_0 + \xi_{\theta}(x_0 + \Delta x, y_0 + \Delta x)) \end{aligned} \quad (4.11)$$

Since ξ_{ρ} and ξ_{θ} are non-linear functions, it follows that a translation in the cartesian plane becomes a deformation in the log-polar one, Fig. 4.4.

4.5. The Fourier-Mellin Transform

We have seen that the log-polar transform translates scale changes and rotations into vertical and horizontal translations. This property has one important application in the Fourier-Mellin transform [25]. One of the most important properties of the Fourier transform is its magnitude invariance to translations. The idea is to take advantage from the combination between the properties of the Fourier and log-polar transforms in order to get a tool that is invariant to translations, scale changes and rotations.

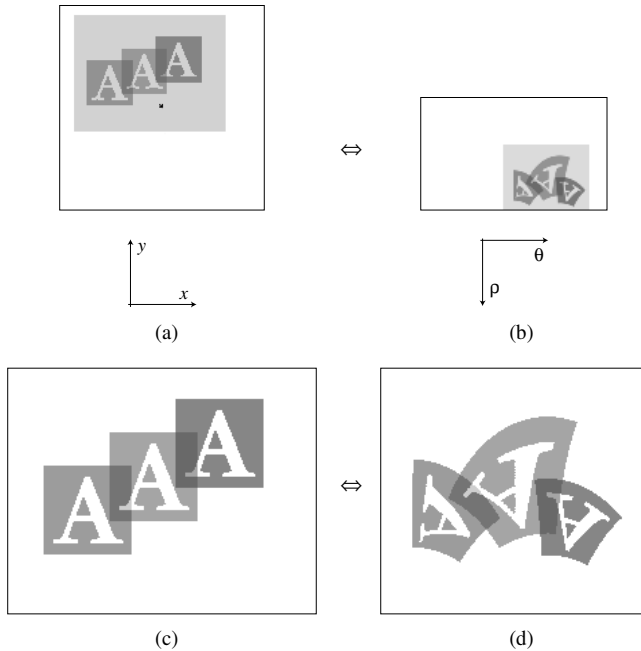


Figure 4.4. Translation: A pure translation implies a deformation of the object after the log-polar transform. (a) Cartesian domain. (b) log-polar domain. (c), (d) Enlargement of the shaded areas respectively in (a) and (b).

The block diagram of the algorithm is described in Fig. 4.5:

If we consider two images $a(x, y)$ and $b(x, y)$, where b is a rotated, scaled and translated copy of a :

$$b(x, y) = a(k[\cos \alpha x + \sin \alpha y] - x_0, k[-\sin \alpha x + \cos \alpha y] - y_0) \quad (4.12)$$

α is the rotation angle, k is the scale factor and x_0 and y_0 are the translation offsets. The Fourier transforms $A(u, v)$ and $B(u, v)$ of a and b respectively are related by:

$$B(u, v) = \frac{e^{-j\phi_b(u, v)}}{k^2} \cdot \left| A \left(\frac{u \cos \alpha + v \sin \alpha}{k}, \frac{-u \sin \alpha + v \cos \alpha}{k} \right) \right| \quad (4.13)$$

where $\phi_b(u, v)$ is the spectral phase of the image $b(x, y)$. This phase depends on the rotation, translation, and scale change, but the spectral magnitude (4.14)

$$|B(u, v)| = \frac{1}{k^2} \cdot \left| A \left(\frac{u \cos \alpha + v \sin \alpha}{k}, \frac{-u \sin \alpha + v \cos \alpha}{k} \right) \right| \quad (4.14)$$

is invariant for translations. Equation (4.14) shows that a rotation of image $a(x, y)$ rotates the spectral magnitude by the same angle α and that a scale change of k scales the spectral magnitude by k^{-1} . However at the spectral origin ($u = 0, v = 0$) there is no change to scale change or rotation. Rotation and scale change can thus be decoupled around this spectral origin by defining the spectral magnitudes of a and b in log-polar coordinates, obtaining:

$$B_{LP}(\rho, \theta) = k^2 A_{LP}(\rho - \gamma, \theta - \alpha) \quad (4.15)$$

where $\rho = \log(r)$ and $\gamma = \log(k)$, while r and θ are the usual polar coordinates. Hence an image rotation (α) shifts the image along the angular axis, and a scale change (k) is reduced to a shift along the radial axis and magnifies the intensity by a constant (k^2).

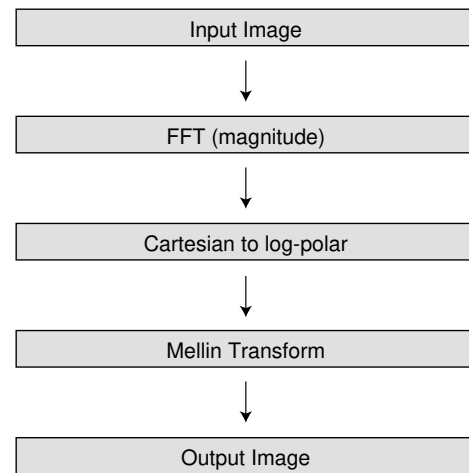


Figure 4.5. The Fourier-Mellin Transform.

This leads to both rotation and scaling now being simple translations, so that taking a Fourier transform of this log-polar representation reduces these effects to phase shifts, so that the magnitudes of the two images are the same.

This is known as a Fourier-Mellin transform and can be used to compare a single template against an unknown image which will be matched even if it has undergone rotation, scaling or translation.

4.6. Straight Line Invariancy

A generic straight line in the cartesian plane can be overlapped to any other one performing a linear combination of a rotation and a translation. If we note that a translation for a straight line of infinite length and zero width can also be seen as a scaling, then the remapped straight line does not change its shape in the log-polar domain.

If we consider a generic straight line in the cartesian domain and its corresponding transform:

$$f_1(x) = \alpha x + \beta \Leftrightarrow f_{1LP}(\theta) = \log_\lambda \left(\frac{\beta}{r_0(\sin \theta - \alpha \cos \theta)} \right) \quad (4.16)$$

we can easily see that another generic straight line is:

$$f_2(x) = \alpha' x + \beta' \Leftrightarrow f_{2LP}(\theta) = \log_\lambda \left(\frac{\beta'}{r_0(\sin \theta - \alpha' \cos \theta)} \right) \quad (4.17)$$

Now we perform a pure rotation, so we set the following values:

$$\begin{aligned} \alpha' &= \frac{\alpha \cos \theta_0 + \sin \theta_0}{\cos \theta_0 + \alpha \sin \theta_0} \\ \beta' &= \frac{\beta}{\cos \theta_0 + \alpha \sin \theta_0} \end{aligned} \quad (4.18)$$

Then we get:

$$\begin{aligned} f_2(x) = \alpha' x + \beta' \Leftrightarrow f_{2LP}(\theta) &= \log_\lambda \left(\frac{\beta'}{r_0(\sin \theta - \alpha' \cos \theta)} \right) \\ &= f_{1LP}(\theta - \theta_0) \end{aligned} \quad (4.19)$$

which is a translation along the θ axis.

Then, considering a pure translation:

$$\begin{aligned} f_3(x) &= \alpha' x + \beta' + k \Leftrightarrow f_{2LP}(\theta) \\ &= \log_\lambda \left(\frac{\beta' + k}{r_0(\sin \theta - \alpha' \cos \theta)} \right) \end{aligned} \quad (4.20)$$

since β' and k are constants, we can set another constant $\xi = \frac{\beta' + k}{\beta'}$ in order to get: $\beta' + k = \xi \beta'$

The equation (4.20) then becomes:

$$\begin{aligned} f_3(x) = \alpha' x + \xi \beta' \Leftrightarrow f_{2LP}(\theta) &= \log_\lambda \left(\frac{\xi \beta'}{r_0(\sin \theta - \alpha' \cos \theta)} \right) \\ &= f_{1LP}(\theta - \theta_0) + \Xi \end{aligned} \quad (4.21)$$

with $\Xi = \log_\lambda \xi$.

This is a pure translation of the original straight-line function, translated by θ_0 horizontally and by $-\Xi$ vertically, Fig. 4.6.

It is interesting to note that if we set in (4.16):

$$\begin{cases} v = \cot^{-1}(-\alpha) \\ w = \log_\lambda \frac{\beta \cdot \sin v}{r_0} \end{cases} \quad (4.22)$$

the equation becomes:

$$f_{1LP}(\theta) = w - \log_\lambda [\cos(\theta - v)] \quad (4.23)$$

where the point $\mathbf{P} = (w, v)$ is the point on the line which is the closest to the origin of the log-polar mapping. This point is unique and uniquely defines the line itself.

This feature is extremely important in various fields. Since it makes easier the detection of straight lines, it can be used in object recognition application, when the object of interest has straight borders. Bishay [26] has taken advantage of this property in order to detect objects edges with an unknown orientation in an indoor scene.

Another application where the invariancy of the straight lines is helpful is, in stereo vision, the detection of the epipolar lines in a stereo pair [27].

4.7. Circumference Invariancy

In general, the shape of a circumference is not invariant in the log-polar domain, but it is interesting to note that if we change the sign of the equation of a generic straight line that we saw in the previous section:

$$\begin{aligned} f_{LP}(\theta) &= -\log_\lambda \left(\frac{\beta}{r_0(\sin \theta - \alpha \cos \theta)} \right) \\ &= \log_\lambda \left(\frac{r_0(\sin \theta - \alpha \cos \theta)}{\beta} \right) \end{aligned} \quad (4.24)$$

and then we transform it back to the cartesian domain, we get:

$$f(x, y) = \beta x^2 + \beta y^2 + r_0^2 \alpha x + r_0^2 y \quad (4.25)$$

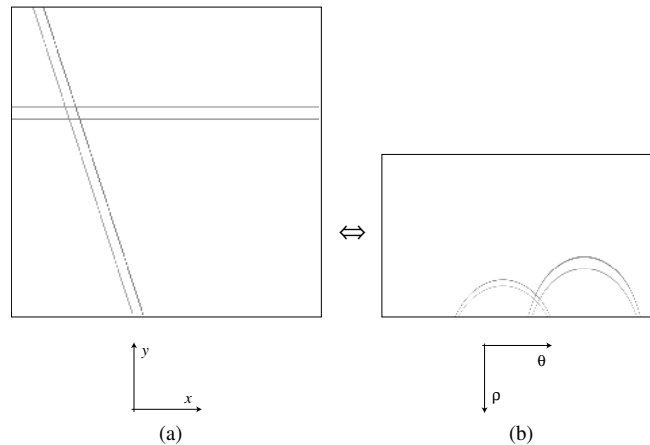


Figure 4.6. Straight Line Invariancy: The shape of a straight line after a log-polar transformation is invariant to its position and orientation. (a) Cartesian domain. (b) Log-polar domain.

which is the equation of a circumference passing through the center of the mapping, centered in:

$$C_x = \frac{r_0^2 \alpha}{2\beta}; \quad C_y = \frac{-r_0^2}{2\beta} \quad (4.26)$$

with radius:

$$R = \frac{r_0^2}{2\beta} \sqrt{\alpha + 1} \quad (4.27)$$

as shown in Fig. 4.7.

Another significant family of circumferences is the set of all the circles centered on the origin of the mapping. Since ρ is constant for every value of θ , if we transform the equation:

$$f(x, y) = x^2 + y^2 - R^2 \quad (4.28)$$

in log-polar coordinates, we get:

$$f(\rho, \theta) = \log_{\lambda} \frac{R^2}{r_0} \quad (4.29)$$

Obviously, this is a constant expression, which is mapped in a horizontal straight line, as in Fig. 4.7. For details about the straight line and the circumference invariancy see [28].

4.8. The Log-Polar Hough Transform

As we showed before, both straight lines and the circumferences passing through the origin are invariant in the log-polar space. A very useful tool that can be used to detect these families of lines is the Hough transform. The Hough Transform [29] is a method for finding simple shapes (generally straight lines) in an image. It sums the number of pixels elements that are in support of a particular line, defined in a uniformly discretized parametric Hough space. The transform is a map from points in an image to sinusoidal curves in Hough space. In the ideal case, this curve represents the family of straight lines that could pass through the point.

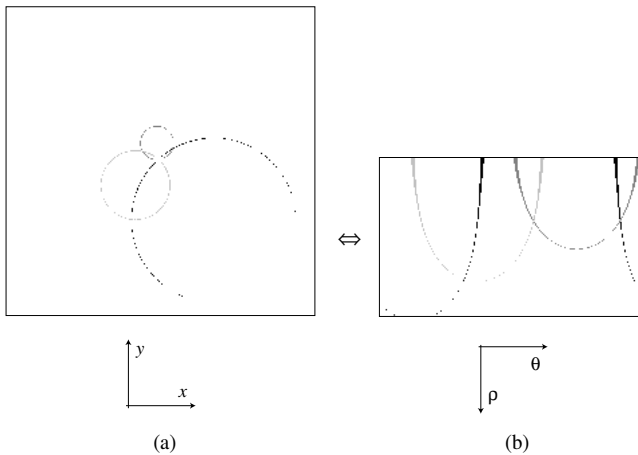


Figure 4.7. Circumference Invariancy: The shape of a circumference, passing through the origin of the mapping, after a log-polar transformation is invariant to its position and orientation. (a) Cartesian domain. (b) Log-polar domain.

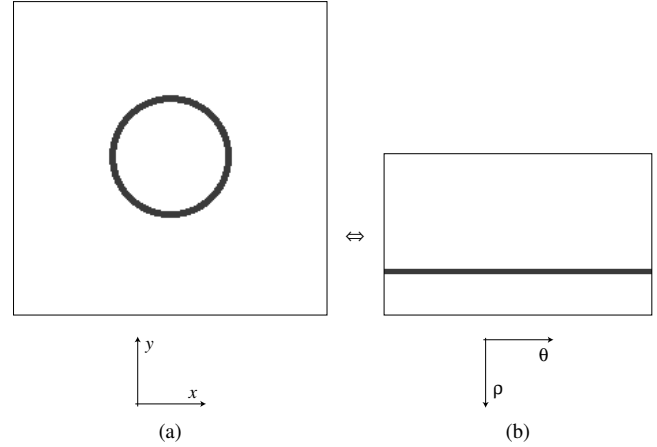


Figure 4.8. Circumference Centered in the origin: The shape of a circumference, centered in the origin of the mapping, after a log-polar transformation is a horizontal straight line. (a) Cartesian domain. (b) Log-polar domain.

The standard form of the transform is the equation of a straight line, parameterized in polar coordinates:

$$\rho(\theta) = x_i \cos \theta + y_i \sin \theta = d_1 \cos(\alpha_1 - \theta) \quad (4.30)$$

where $(x_i; y_i)$ are the pixel cartesian coordinates, $(x_i; y_i)$ are the pixel polar coordinates, ρ is the closest distance to the origin, and θ the slope of the normal to the line.

Using this transform, any edge point that appears in the image votes for all lines that could possibly pass through that point. In this way, if there is a real line in the image, by transforming all of its points to Hough space, we will accumulate a large number of votes for the actual line, and only one for any other line (assuming no noise) in the image.

The vast majority of research on the Hough transform has treated sensors that use a uniform distribution of sensing elements. For space-variant sensors various approaches have been tried. Weiman [30] has proposed a log-Hough transform, where equation (4.30) becomes:

$$\rho_{\log}(\theta) = \log \frac{\rho(\theta)}{\rho_0} = \log d - \log \rho_0 + \log(\cos(\theta - \alpha)) \quad (4.31)$$

Barnes [31] has investigated the log Hough transform in the real (discrete) case of images acquired by a log-polar sensor.

4.9. Spirals

The last geometrical figure with some interesting characteristics is the logarithmic spiral. The parametric equations for such a spiral is:

$$\begin{cases} x = \alpha \cos(\theta) \cdot e^{\beta\theta} \\ y = \alpha \sin(\theta) \cdot e^{\beta\theta} \end{cases} \quad (4.32)$$

with α and β arbitrary parameters.

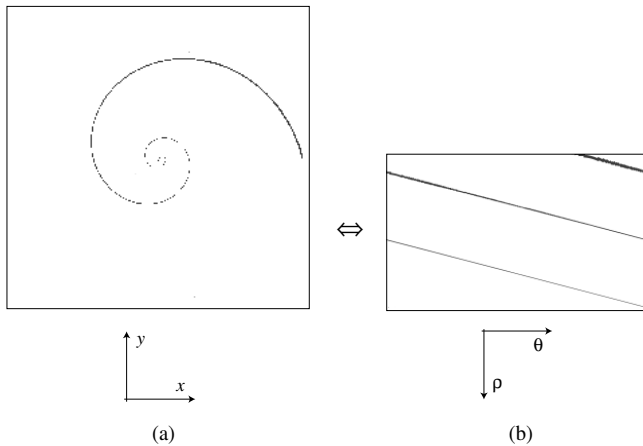


Figure 4.9. Logarithmic Spiral: Such a spiral, centered in the origin of the mapping, after a log-polar transformation becomes a straight line. (a) Cartesian domain. (b) Log-polar domain. Please note that in our representation we display the set of points $\theta_0 + 2k\pi$ all on θ_0 .

The log-polar transform of (2.32) gives:

$$\rho = \log_{\lambda} \frac{\alpha \cdot e^{\beta\theta}}{r_0} \quad (4.33)$$

If we set in (4.33):

$$\begin{cases} k_1 = \ln \frac{a}{r_0} \\ k_2 = \frac{1}{\ln \lambda} \end{cases} \quad (4.34)$$

then we get:

$$\rho = k_2(k_1 + \beta\theta) \quad (4.35)$$

which is the equation of a straight line (Fig. 4.9).

5. APPLICATIONS

5.1. Panoramic View

Traditionally the acquisition of real time panoramic images has been performed by the usage of lenses or mirrors coupled with standard image sensors, but this solution presents the problems of having different resolutions on different part of the resulting images and the need of both the acquisition of a huge amount of pixels and the processing of these acquired pixels.

The main objective of the OMNIVIEWS project [32] was to integrate optical, hardware, and software technology for the realization of a smart visual sensor, and to demonstrate its utility in key application areas. In particular the intention was to design and realize a low-cost, miniaturized digital camera acquiring panoramic (360°) images and performing a useful low-level processing on the incoming stream of images, Fig. 5.1.

The solution proposed in OMINVIEWS was to integrate a retina like CMOS visual sensor, with a mirror with a specially designed, matching curvature. This matching, if feasible, provides panoramic images without the

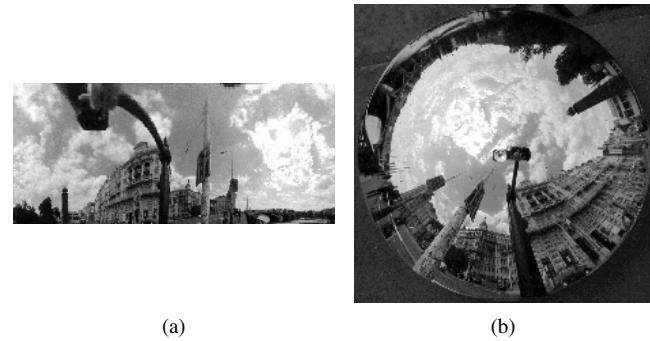


Figure 5.1. Panoramic View: (a) Image acquired by an OMNIVIEWS log-polar camera (software simulation). (b) Image acquired by a conventional omnidirectional camera. Note that the image from OMNIVIEWS camera is immediately understandable while the image from a conventional camera requires more than 1.5 million operations to be transformed into something similar with no added advantage.

need of computationally intensive processing and/or hardware remapper as required by conventional omnidirectional cameras. Therefore reducing overall cost, size, energy consumption and computational power with respect to the currently used devices. The panoramic images obtained with a log-polar technology are not only equivalent to the ones obtained with conventional devices but also these images can be obtained at no computational cost. For example with our current prototype a panoramic image composed of about 27,000 pixels is obtained by simply reading out the pixels (i.e., 27,000 operations) while with a conventional solution the same image would required more than 1.7 million operations (about 50 times more). Besides that, unlike with a warped traditional image, we get the interesting side effect of a uniform resolution along the entire panoramic image.

The guiding principle is to design the profile of the mirror so that if the camera is inserted inside a cylinder, the direct camera output provides an undistorted, constant resolution image of the internal surface of the cylinder. The advantage of such an approach lays in providing the observer with a complete view of its surrounding in one image which can be refreshed at video rate.

For technical details about the design of the mirror, in Fig. 5.2, see the OMNIVIEWS project report [33].

The panoramic vision has a lot of different applications, with or without the presence of a space variant sensor but, as we described above, the utilization of a log-polar sensor greatly improves the performances of the system.

The most common usage of this technology (there are in fact several commercial products already available based on a camera and a panoramic mirror) is in the field of remote surveillance, where it is very useful to have a complete vision of the whole environment at once, maybe coupled with a motorized standard camera in order to zoom on the objects of interest.

Other applications involve biological vessels (panoramic endoscopies) or industrial pipes investigation, where it is important to have a simultaneous view of the whole internal surface of the cylinder, or robotic navigation, where the advantage is represented by the need of just one camera in order to have a representation of the whole surrounding environment [34].

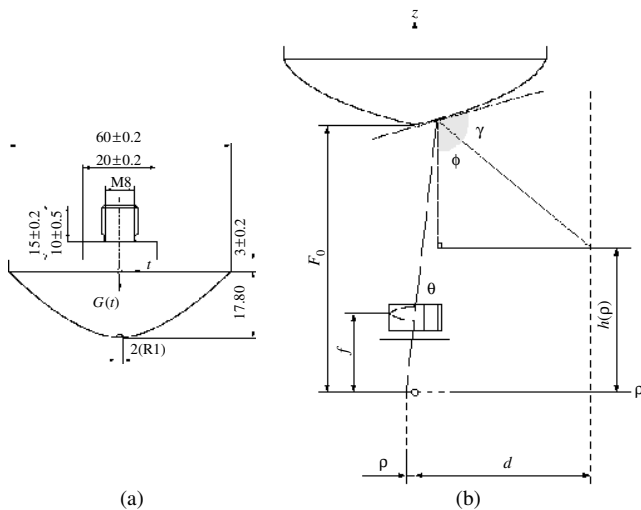


Figure 5.2. Mirror design: (a) Profile of the mirror. (b) The mirror is designed so that vertical resolution of a cylindrical surface is mapped into constant radial resolution in the image plane.

5.2. Robotic Vision

The need of real-time image processing speed is particularly relevant in visually guided robotics. It can be achieved both by increasing the computational power and by constraining the amount of data to be processed. We have seen that the log-polar mapping is able to limit the number of pixels while keeping both a wide field of view and a high resolution in the fovea. But this is not the only reason for such a choice: the biological motivation is very important too when dealing with human-like robots.

Over the last few years, we have studied how sensorimotor patterns are acquired in a human being. This has been carried out from the unique perspective of implementing the behaviors we wanted to study in artificial systems.

The approach we have followed is biologically motivated from at least three perspectives: the morphology of the artificial being should be as close as possible to the human one, i.e., the sensors should approximate their biological counterparts; its physiology should be designed in order to have its control structures and processing modeled after what is known about human perception and motor control; its development, i.e., the acquisition of those sensorimotor patterns, should follow the process of biological development in the first few years of life.

The goal has been that of understanding human sensorimotor coordination and cognition rather than building more efficient robots. So, instead of being interested in what a robot is able to do, we are interested in how it does it. In fact some of the design choices might even be questionable from a purely engineering point of view but they are pursued nonetheless because they improved the similarity with, hence the understanding of, biological systems. Along this feat of investigating the human cognitive processes, our group implemented various artifacts both in software (e.g., image processing and machine learning) and hardware (e.g., silicon implementation of the visual system, robotic heads, and robotic bodies).

The most part of the work has been carried out on a humanoid robot called *Babybot* [35, 36] that resembles the human body from the waist up although in a simplified form, Fig. 5.3. It has eighteen degrees of freedom overall distributed between the head, the arm, the torso and the hand. Its sensory system consists of cameras (the eyes), gyroscopes (the vestibular system), microphones (the ears), position sensors at the joints (proprioception) and tactile sensors on the palm of hand. During the year 2004 a new humanoid is being developed. It will have 23 degrees of freedom, it will be smaller and all its parts will be designed with the final goal of having the best possible similarity with their biological counterparts.

Investigation touched aspects such as the integration of visual and inertial information [37], and the interaction between vision and spatial hearing [38].

So, it is very important to understand why the visual system of the robot has to be as similar as possible to human vision: we have to find out whether the morphology of the visual sensors can improve particular sensorimotor strategies and how vision can have consequences on the behaviors which are not purely visual in nature. We should note again that the eyes and motor behaviors coevolved: it is useless to have a fovea if the eyes cannot be swiftly moved over possible targets. Humans developed a sophisticated oculomotor apparatus that includes saccadic movements, smooth tracking, vergence, and various combinations of retinal and extra-retinal signals to maintain vision efficient in a wide variety of situations.

As stated above, handling with log-polar images can be sometimes uneasy and time consuming, due to the deformation of the plane, but this drawbacks are always balanced

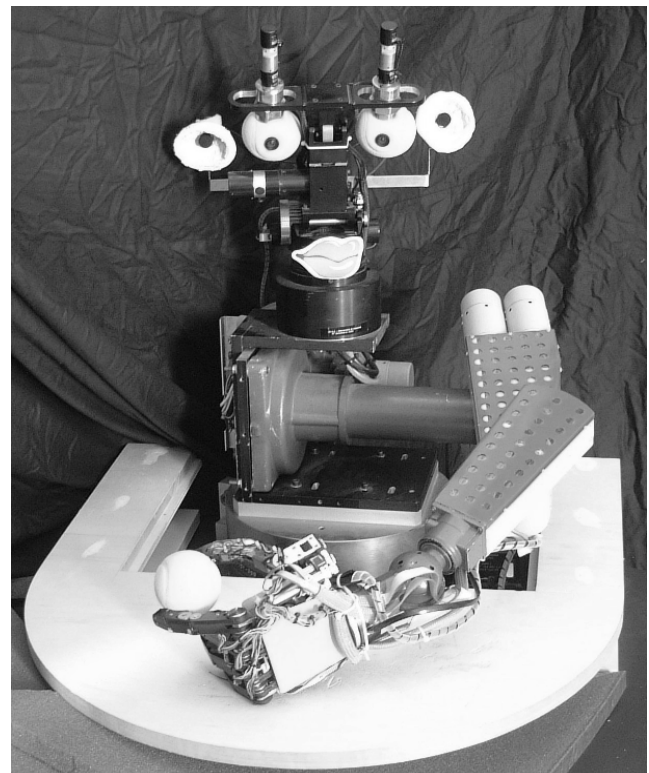


Figure 5.3. LIRA-Lab Humanoid: Babybot

by the highly reduced number of pixels that have to be processed.

Various algorithms have been developed to perform common tasks in log-polar geometry, like vergence control and disparity estimation [39]. In this last case, grossly simplifying, by employing a correlation measure, the algorithm evaluates the similarity of the left and right images for different horizontal shifts. It finally picks the shift relative to the maximum correlation as a measure of the binocular disparity. The log-polar geometry, in this case, weighs differently the pixels in the fovea with respect to those in the periphery. More importance is thus accorded to the object being tracked.

Positional information is important but for a few tasks optic flow is a better choice. One example of use of optic flow is for the dynamic control of vergence as in [40]. We implemented a log-polar version of a quite standard algorithm [41]. The details of the implementation can be found in [42]. The estimation of the optic flow requires taking into account the log-polar geometry because it involves non-local operations.

5.3. Video Conferencing, Video Telephony and Image Compression

The log-polar transform of an image can be seen as a lossy image compression algorithm, where all the data loss is concentrated where the information should be less attracting for the user. With our family of sensor the compression rate for a single frame is between 30 and 40 to 1, which is (see Table 1) given by:

$$\frac{Q^2}{N_{LP}} \quad (5.1)$$

where N_{LP} is the total number of pixels in the log-polar image. As we have seen, some algorithms can still be applied with no relevant change to the log-polar images, so we can process the image, or the video stream with virtually all the compression algorithms commonly available. This impressive compressing factor can be used to send video streams on very narrow band channels.

One possible application has been investigated in the past few years with the EU funded project IBIDEM [43]. This project has been inspired by the fact that hearing impairment prevents many people from using normal voice telephones for obvious reasons. A solution to this problem for the hearing impaired is the use of videophones.

At that time available videophones working on standard telephone lines (Public Switched Telephone Network) did not meet the dynamic requirements necessary for lip reading, finger spelling and signing. The spatial resolution was also too small.

The main objective of IBIDEM was to develop a videophone useful for lip reading by hearing impaired people based on the space variant sensor and using standard telephone lines. The space variant nature of the sensor allowed to have high resolution in the area of interest, lips or fingers, while still maintaining a wide field of view in order to perceive, for example, the facial expression of the interlocutor,

but reducing drastically the amount of data to be sent on the line (see Fig. 5.4).

After the IBIDEM project, its natural continuation was to test the performances of the log-polar cameras (which have been called *Giotto*, see Fig. 5.5) on even narrower bandwidths. Extensive experiments on wireless image transmission were conducted with a set-up composed of a remote PC running a web server embedded into an application that acquires images from the a retina like camera and compress them following one of the recommendations for video coding over low bit rate communication line (H.263 in our case).

The remote receiving station was a palmtop PC acting as a client connected to the remote server through a dial-up GSM connection (9600 baud). Using a standard browser interface the client could connect to the web server, receive the compressed stream, decompress it and display the resulting images on the screen. Due to the low amount of data to be processed and sent on the line, frame rates of up to four images per second could be obtained. The only special-purpose hardware required is the log-polar camera; coding/decoding and image remapping is done in software on a desktop PC (on the server side), and on the palmtop PC (on the client side). The aspect we wanted to stress in these experiments is the use of off-the-shelf components and the overall physical size of the receiver. This performance in terms of frame rate, image quality and cost cannot clearly be accomplished by using conventional cameras. More recently (within a project called AMOVITE) we started realizing a portable camera that can be connected to the palmtop PC allowing bi-directional image transmission through GSM or GPRS communication lines. The sensor itself is not much different from the one previously described apart from the adoption of a companion chip allowing a much smaller camera.

The same principle has been adopted in other experiments of image transmission: Comanicu [44] has added a face-tracking algorithm to the software log-polar compression for remote surveillance purposes. Weiman [45], combining the log-polar transform with other compression algorithms has reached a compression factor of 1600:1.

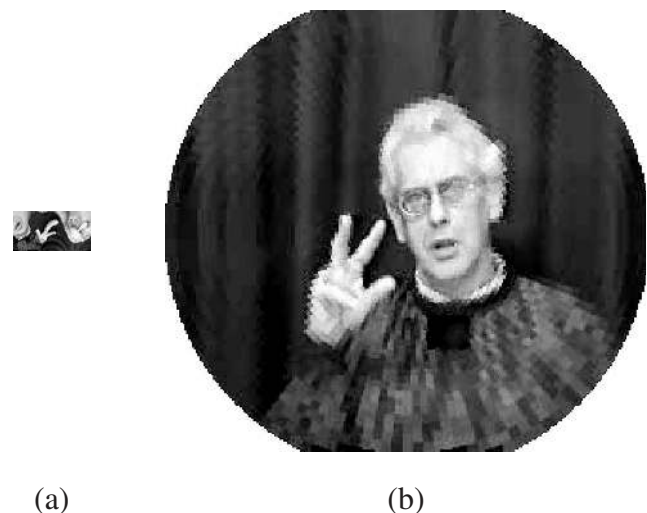


Figure 5.4. IBIDEM project finger spelling experiment: (a) Log-polar and (b) remapped images.



Figure 5.5. Giotto Camera.

GLOSSARY

Active Vision The control of the optics and the mechanical structure of cameras or eyes to simplify the processing for vision.

Cone A specialized nerve cell in the retina, which detects color.

Field Of View (FOV) An angle that defines how far from the optical axis the view is spread.

Fovea A high-resolution area in the retina, usually located in the center of the field of view.

Foveola The central part of the fovea.

Optical Axis An imaginary line that runs through the focus and center of a lens.

Photoreceptor A mechanism that emits an electrical or chemical signal that varies in proportion to the amount of light that strikes it.

Receptive Field The portion of the sensory surface where stimuli affect the activity of a sensory neuron.

Rod A light-detecting cell in the retina, detects light and movement, but not color.

Saccadic Movement A rapid eye movement used to alter eye position within the orbit, causing a rapid adjustment of the fixation point to different positions in the visual world.

Vergence The angle between the optical axes of the eyes.

Retina The light-sensitive layer of tissue that lines the back of the eyeball, sending visual impulses through the optic nerve to the brain.

CCD (Charge-Coupled Device) A semiconductor technology used to build light-sensitive electronic devices such as cameras and image scanners. Such devices may detect either colour or black-and-white. Each CCD chip consists of an array of light-sensitive photocells. The photocell is sensitized by giving it an electrical charge prior to exposure.

CMOS (Complementary Metal Oxide Semiconductor) A semiconductor fabrication technology using a combination of n- and p-doped semiconductor material to achieve low power dissipation. Any path through a gate through which current can flow includes both n and p type transistors. Only one type is turned on in any stable state so there is no

static power dissipation and current only flows when a gate switches in order to charge the parasitic capacitance.

REFERENCES

1. M. V. Srinivasan and S. Venkatesh, "From Living Eyes to Seeing Machines." Oxford University Press, Oxford, 1997.
2. P. M. Blough, "Neural Mechanisms of Behavior in the Pigeon." Plenum Press, New York, 1979.
3. Y. Galifret, *Z. Zellforsch* 86, 535 (1968).
4. J. B. Jonas, U. Schneider, and G. O. Naumann, *Graefes Arch. Clin. Exp. Ophthalmol.* 230, 505 (1992).
5. R. H. S. Carpenter, "Movements of the Eyes," 2nd Edn., Pion Limited, London, 1988.
6. G. Sandini and V. Tagliasco, *Comp. Vision Graph.* 14, 365 (1980).
7. E. L. Schwartz, *Biol. Cybern* 37, 63 (1980).
8. C. F. R. Weiman and G. Chaikin, *Computer Graphic and Image Processing* 11, 197 (1979).
9. E. L. Schwartz, *Biol. Cybern* 25, 181 (1977).
10. C. F. R. Weiman, *Proc. SPIE* Vol. 938 (1988).
11. R. A. Peters II, M. Bishay, and T. Rogers, Tech Rep, Intelligent Robotics Laboratory, Vanderbilt Univ, 1996.
12. T. Fisher and R. Juday, *Proc. SPIE* Vol. 938 (1988).
13. G. Engel, D. N. Greve, J. M. Lubin, and E. L. Schwartz, *ICPR* (1994).
14. A. Rojer and E. L. Schwartz, *ICPR* (1990).
15. R. S. Wallace, P. W. Ong, B. B. Bederson, and E. L. Schwartz, *Int. J. Comput. Vision* 13, 71 (1994).
16. C. F. R. Weiman and R. D. Juday, *Proc. SPIE*, Vol. 1192 (1990).
17. R. Suematsu and H. Yamada, *Trans. of SICE*, Vol. 31, 10, 1556 (1995).
18. Y. Kuniyoshi, N. Kita, S. Rougeaux, and T. Suehiro, ACCV (1995).
19. T. Baron, M. D. Levine, and Y. Yeshurun, ICPR (1994).
20. T. Baron, M. D. Levine, V. Hayward, M. Bolduc, and D. A. Grant, Proc. CASI (1995).
21. R. Wodnicki, G. W. Roberts, and M. D. Levine, *IEEE J. Solid-State Circuits* 32, 8, 1274 (1997).
22. J. Van der Spiegel, G. Kreider, C. Claeys, I. Debusschere, G. Sandini, P. Dario, F. Fantini, P. Bellutti, and G. Soncini, in "Analog VLSI and Neural Network Implementations" (C. Mead and M. Ismail, Eds.), DeKluwer Publ, Boston, 1989.
23. E. W. Weisstein. "Conformal Mapping." From MathWorld—A Wolfram Web Resource. <http://mathworld.wolfram.com/Conformal-Mapping.html>.
24. L. Grady, Ph.D. Thesis, 2004.
25. R. Schalkoff, "Digital Image Processing and Computer Vision." Wiley and Sons, New York, 1989.
26. M. Bishay, R. A. Peters II, and K. Kawamura, ICRA (1994).
27. K. Schindler and H. Bischof, ICPR (2004).
28. D. Young, BMVC (2000).
29. Paul V. C. Hough, U.S. Patent No. 3,069,654 (1962).
30. C. F. R. Weiman, Phase I SBIR Final Report, HelpMate Robotics Inc. 1994.
31. N. Barnes, Proc. OMNIVIS'04 Workshop at ECCV (2004).
32. G. Sandini, J. Santos-Victor, T. Pajdla, and F. Berton, *IEEE Sensors* (2002).
33. S. Gächter, FET Project No: IST-1999-29017 (2001).
34. J. Santos-Victor and A. Bernardino, "Robotics Research, 10th International Symposium." (R. Jarvis and A. Zelinsky, Eds.), Springer, 2003.
35. G. Metta, Ph.D. Thesis, 2000.
36. G. Metta, G. Sandini, and J. Konczak, *Neural Networks* 12, 1413 (1999).
37. F. Panerai, G. Metta, and G. Sandini, *Robot Auton. Syst.* 30, 195 (2000).
38. L. Natale, G. Metta, and G. Sandini, *Robot Auton. Syst.* 39, 87 (2002).

39. R. Manzotti, A. Gasteratos, G. Metta, and G. Sandini, *Comput. Vis. Image Und.* 83, 97 (2001).
40. C. Capurro, F. Panerai, and G. Sandini, *Int. J. Comput. Vision* 24, 79 (1997).
41. J. Koenderink and J. Van Doorn, *J. Optical Soc. Am.* 8, 377 (1991).
42. H. Tunley and D. Young, ECCV (1994).
43. F. Ferrari, J. Nielsen, P. Questa, and G. Sandini, *Sensor Review* 15, 17 (1995).
44. D. Comaniciu, F. Berton, and V. Ramesh, *Real Time Imaging* 8, 427 (2002).
45. C. F. R. Weiman, *Proc. SPIE* 1295, 266 (1990).