

# 18. Retina-Like Sensors: Motivations, Technology and Applications

Giulio Sandini and Giorgio Metta

## Abstract

- I. Introduction
  - II. Space-Variant Imaging
  - III. Technology of Solid State Log-Polar Sensors
    - A. Parameters of Log-Polar Sensors
    - B. CMOS Implementations
    - C. The Fovea of Log-Polar Sensors
  - IV. Applications
    - A. Image Transmission
    - B. Robotics
  - V. Conclusions
- References

## Abstract

Retina-like visual sensors are characterized by space-variant resolution mimicking the distribution of photoreceptors of the human retina. These sensors, like our eyes, have a central part at highest possible resolution (called fovea) and a gradually decreasing resolution in the periphery. We will present a solid-state implementation of this concept. One attractive property of space-variant imaging is that it allows processing the whole image at frame rate while maintaining the same field of view of traditional rectangular sensors. The resolution is always maximal if the cameras are allowed to move and the fovea placed over the regions of interest. This is the case in robots with moving cameras. As an example of possible applications, we shall describe a robotic visual system exploiting two retina-like cameras and using vision to learn sensorimotor behaviors.

## I. Introduction

This chapter describes the physical implementation in silicon of a biologically inspired (retina-like) visual sensor. It is important also to understand the context and the motivations for undertaking such a research. Over the past few years, we have studied how sensorimotor patterns are acquired in a complex system such as the human body. This has been carried out from the unique perspective of implementing the behaviors we wanted to study in artificial systems. The approach we followed is biologically motivated from at least three perspectives: i) morphology (sensors are shaped as close as possible to their biological counterparts); ii) physiology (control structures and processing are modeled after what is known about human perception and motor control); iii) development (the acquisition of those sensorimotor patterns follows 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. In fact, some of the design choices might even be questionable on a purely engineering ground but they

are pursued nonetheless because they improved the similarity with biological systems. Along this feat of understanding human cognition our group implemented various artifacts both in software (e.g. image processing, machine learning) and hardware (e.g. silicon implementation, robot heads).

The most part of the work has been carried out on a humanoid robot we call the *Babybot* (Metta 2000, Metta et al. 1999). It resembles the human body from the waist up although in a simplified form. It has twelve degrees of freedom overall distributed between the head, arm and torso. Its sensory system consists of cameras (eyes), gyroscopes (vestibular system), microphones (ears) and position sensors at the joints (proprioception). *Babybot* cannot grasp objects but it can touch and poke them around. Investigation touched aspects such as the integration of visual and inertial information (vestibulo-ocular reflex) (Panerai et al. 2000), and the interaction between vision and spatial hearing (Natale et al. 2002).

In this paper we will focus on the eyes of *Babybot* that mimic the distribution of photoreceptors of the human retina – we call these *retina-like cameras*. Apart from the purely computational aspects, they are best understood within the scientific framework of the study of biological systems. In our view, the retina-like camera truly represents such a thought-provoking mix of technological and biologically inspired work. Retina-like cameras have a non-uniform resolution with a high resolution central part called the *fovea* and a coarser resolution periphery.

As we hope to clarify in this chapter, the uniform resolution layout, common to commercial cameras, did not survive the evolutionary pressure. Evolution seems to be answering the question of how to optimally place a given number of photoreceptors over a finite small surface. Many different eyes evolved with the disposition

of the photoreceptors adapted to the particular ecological niche. Examples of this diversity can be found in the eyes of insects (see for example (Srinivasan and Venkatesh 1997) for a review) and in those of some birds which have two foveal regions to allow simultaneous flying and hunting (Blough 1979, Galifret 1968). There is clearly something to be learned by optimizing the placement of photosensitive elements. In a constrained problem with limited computational resources (i.e. the size of the brain), limited bandwidth (i.e. the diameter of the nerves delivering the visual information to the brain), and limited number of photoreceptors, nature managed to obtain a much higher acuity than what can be achieved with uniform resolution.

From the visual processing point of view, we asked on the 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 which 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. Humans developed a sophisticated oculomotor apparatus which 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 (see Carpenter 1988 for a review).

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 the retina-like camera. Examples of applications are also discussed in the field of

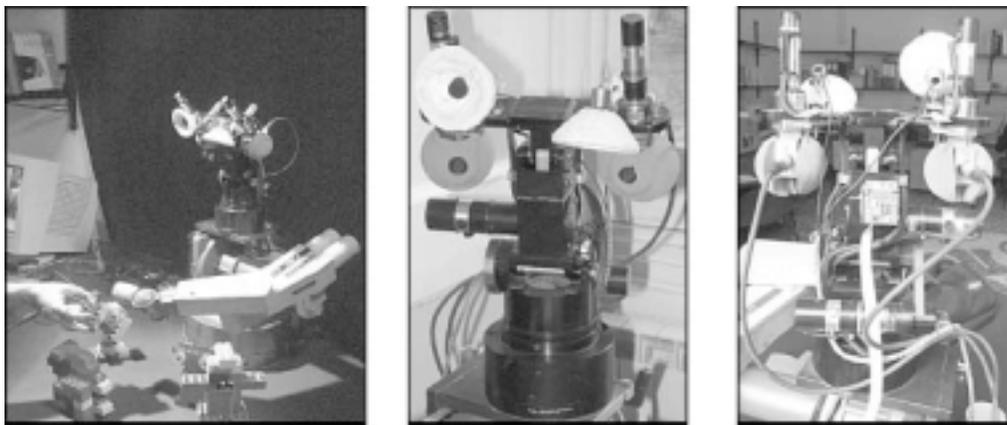
image transmission and robotics. The image transmission problem is alike the limitation of bandwidth/size of the optic nerve mentioned above. The limitations in the case of autonomous robots are in terms of computational resources and power consumption.

## II. Space-Variant Imaging

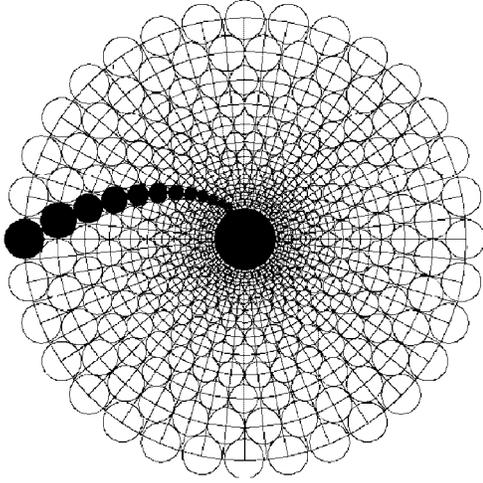
Babybot, shown in Fig. 1, relies on a pair of retina-like sensors for its visual processing. These sensors are characterized by a space-variant resolution mimicking the distribution of photoreceptors of the human retina. The density of photoreceptors is highest in the center (limited by the particular technology used) and decreases monotonically as the eccentricity – the distance of the photosite from the center of the sensory surface – increases. The resulting image is, consequently a compromise between resolution, amplitude of the field of view (FOV), and number of pixels. This space-variant imaging is unique because it enables high-resolution tasks

using the central region while maintaining the lower resolution periphery providing relevant information about the background. The arrangement is advantageous, for example, for target tracking: the wide peripheral part is useful for detection while the central part takes over during the tracking and performs with the highest accuracy.

Of all possible implementations of space-variant sensors what is described here is the so-called log-polar structure (Sandini and Tagliasco 1980, Schwartz 1980, Weiman and Chaikin 1979). The log-polar geometry models accurately the wiring of the photoreceptors from the retina to the geniculate body and the primary visual cortex (area V1). In this schema a constant number of photosites is arranged over concentric rings (the polar part of the representation) giving rise to a linear increase of the receptor's spacing with respect to the distance from the central point of the structure (the radius of the concentric rings). A possible implementation of this arrangement is shown in Fig. 2. Because of the polar structure



**Fig. 1.** The Babybot. Left: the complete setup. Middle: detail of the head, the tennis-like balls cover the eyes for esthetic purpose, the microphones and ear lobes are mounted on top of the head. Right: back view of the head showing the inertial sensors in its center



**Fig. 2.** Layout of receptor's placing for a log-polar structure composed of 12 rings with 32 pixels each. The pixels marked in black follow a logarithmic spiral

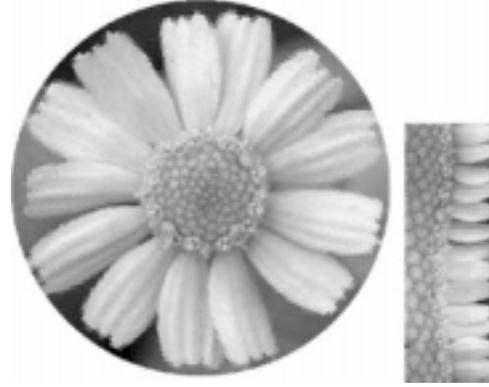
of the photosensitive array and the increasing size of pixels in the periphery, retina-like sensors, do not provide images with a standard topology. In formulas mapping from the retina  $(\rho, \vartheta)$  into the cortical plane  $(\eta, \xi)$  accounts to:

$$\begin{cases} \eta = q \cdot \vartheta \\ \xi = \log_a \frac{\rho}{\rho_0} \end{cases} \quad (1)$$

where  $\rho_0$  is the radius of the innermost circle, and  $1/q$  is the minimum angular resolution. Equation (1) is easily understood by observing that the angular variable  $\eta$  is linearly mapped from its polar representation  $\vartheta$ , and the eccentricity  $\rho$  is scaled logarithmically (with basis  $a$ ). Equation (1) is related to the traditional rectangular coordinate system by:

$$\begin{cases} x = \rho \cos \vartheta \\ y = \rho \sin \vartheta \end{cases} \quad (2)$$

It is worth noting that the mapping is obtained at no computational cost, as it is a direct consequence of the arrangement of the photosites and the read-out sequence.



**Fig. 3.** Left: space variant image obtained by re-mapping an image acquired by a log-polar sensor. Note the increase in pixels size with eccentricity. Right: Log-polar image acquired by a retina-like sensor. Horizontal lines in the log-polar image are mapped into rings to obtain the re-mapped image shown on the right

The topology of a log-polar image is shown in Fig. 3. Note that radial structures (the petals of the flower) correspond to horizontal structures in the log-polar image. In spite of this seemingly distorted image, the mapping is conformal and, consequently, any local operator used for standard images can be applied without changes (Weiman and Chaikin 1979).

In the Babybot, for example, images coming from the left and right channels were processed to recover the location, in retinal coordinates, of possibly interesting objects. Color is a good candidate for this purpose and it is extracted after mapping the RGB components into the Hue Saturation and Value (HSV) space. This is not normalized or scaled as for example in (Darrell et al. 2000), but proved to be enough for our experiments. The processing extracts and labels the most salient image regions. Regions are successively combined through a voting mechanism and finally the coordinates of the most "voted" one are selected as the position

of the object to look at. This fusion procedure provides the attentional information to generate tracking behavior. The error – i.e. the distance of the projection of the target on the image plane from the center of the image – is used to trigger saccade-like movements or, in case of smoothly moving targets, as feedback signal for closed loop control.

The way the different visual cues are combined does not provide information about binocular disparity. In fact the procedure described above does not perform any stereo matching or correspondence search – regions are treated as 2D entities. For the purpose of measuring depth we used a different algorithm whose details can be found in (Manzotti et al. 2001). This, too, uses log-polar images and provides a measure of the global disparity, which is used to control vergence. 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 (Capurro et al. 1997). We implemented a log-polar version of a quite standard algorithm (Koenderink and Van Doorn 1991). According to the choice of the algorithm we defined the affine model as:

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} + \begin{bmatrix} D + S_1 & S_2 - R \\ R + S_2 & D - S_1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} \quad (3)$$

where  $\dot{x}$  and  $\dot{y}$  are the optic flow, and  $x$  and  $y$  the image plane coordinates (with

origin in the image center). Equation (3) depends on four quantities: translation, rotation, divergence and shear. The first two components  $u_0$  and  $v_0$  represent a rigid 2D translation, and  $D$ ,  $R$ ,  $S_1$ ,  $S_2$  are the four first-order vector field differential invariants: divergence, curl, and shear respectively. The details of the implementation can be found in (Tunley and Young 1994). The estimation of the optic flow requires taking into account the log-polar geometry because it involves non-local operations.

### III. Technology of Solid State Log-Polar Sensors

Traditionally, the log-polar mapping has been obtained in two different ways: by means of electronic boards transforming, in real-time, standard images into log-polar ones or by building sensors with the photosites arranged according to the log-polar mapping. Electronic boards were employed first; they were used to generate log-polar images for real-time control and image compression (Engel et al. 1994, Rojer and Schwartz 1990, Wallace et al. 1994, Weiman and Juday 1990). The advantage is the use of standard off-the-shelf electronic components. The main disadvantage is the constraint introduced by the size of the original image limiting the potential advantages of the log-polar structure. This point will be clarified in the following sections. Whatever the approach used, the design of the sensor has to start from the technological limitations: the most important of which are the minimum pixel size and the maximum sensor size. Besides our realizations a few other attempts have been reported in the literature on the implementation of solid-state retina like sensors (Baron et al. 1995, Baron et al. 1994). So far we are not aware of any commercial device, besides those

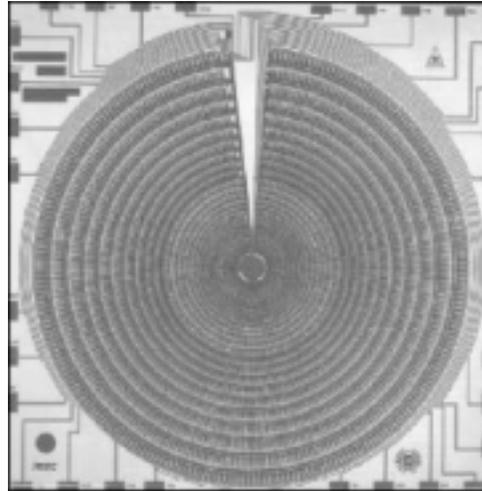
described here, that have been realized based on log-polar retina-like sensors.

#### A. Parameters of Log-Polar Sensors

Starting from the technological constraints (minimum pixel size and size of the sensor), the most important sensor's parameter is the total number of pixels. The total number of pixels is directly related to the amount of information acquired. For constant resolution devices this parameter is fixed by the technological constraints in a simple way. In the case of log-polar sensors the relationship between these two parameters is not a simple one (see Sandini and Tagliasco 1980, Wallace et al. 1994, Weiman 1988) for more details). The second important parameter, unique to log-polar sensors, is the ratio between the largest and the smallest pixels – we shall call it  $R$ .

For example, our first solid-state implementation was realized at the beginning of the 90s using CCD technology (Van der Spiegel et al. 1989). At that time, with the technology available, the size of the smallest possible pixel was about  $30\ \mu\text{m}$  and for practical limitations the overall sensor diameter was limited to 94 mm. A picture of the layout is shown in Fig. 4. This sensor is composed of 30 rings and each ring is covered by 64 pixels. Altogether the sensor had 2022 pixels, 1920 of which in the log-polar part of the sensor. In the CCD implementation  $R$  was about 13.7 (the largest pixel was  $412\ \mu\text{m}$ ). This parameter describes the amount of space variance of the sensor and is, of course, equal to 1 in standard constant resolution sensors.

The third important parameter is the ratio between size of the sensor and the size of the smallest pixel. We shall call it  $Q$ . The importance of  $Q$  can be understood by observing that its value is equal to the size of a constant resolution image



**Fig. 4.** Structure of the first log-polar sensor realized with CCD technology

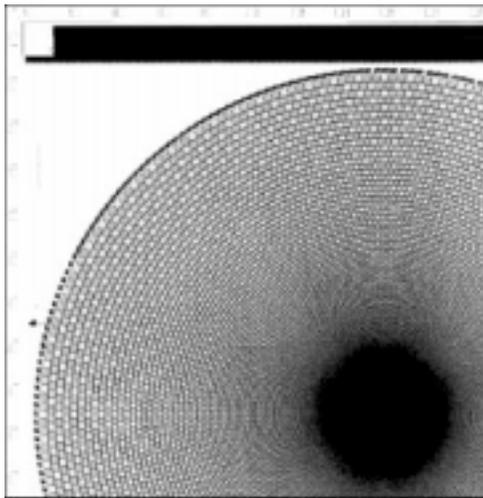
with the same field of view and the same maximum resolution of the corresponding log-polar sensor. For example for the CCD sensor shown in Fig. 4  $Q$  is equal to about 300 meaning that if we want to electronically remap a constant resolution image and obtain the same amount of information obtained from our log-polar sensor, the original image must be at least  $300 \times 300$  pixels.

#### B. CMOS Implementations

The CCD implementation described earlier, even if it was the first solid-state device of this kind in the world, had some drawbacks mostly related to the use of CCD technology itself. In our more recent implementations, the CMOS technology was used. A first version of the sensor was realized using a  $0.7\ \mu\text{m}$  technology allowing a minimum pixel size of  $14\ \mu\text{m}$ . Later, improvement of technology enabled a further reduction of the minimum pixel size to  $7\ \mu\text{m}$ . In fact, our most recent CMOS implementation uses a  $0.35\ \mu\text{m}$

technology. This sensor – realized at Tower in Israel – has been developed within a European Union-funded research project (SVAVISCA). The goal of the project was to build, beside the sensor, a microcamera with a special-purpose lens with a 140 degree field of view. The miniaturization of the camera is now possible because some of the electronics required to drive the sensor as well as the analog-to-digital converter is included on the chip itself. A picture of part of the layout of the sensor is shown in Fig. 5. Table 1 summarizes the main parameters of the sensor.

To compare, the first implementation had a  $Q$  equal to 300. Even if the total number of pixels of the  $300 \times 300$  image was 40 times larger than the 2000 pixel retina-like image, its size was still well within the limits of standard computer hardware (e.g. bus bandwidth, memory, etc.). The latest sensor, if simulated using a software or hardware remapper would require the storage and processing of an image with a number of pixels exceeding



**Fig. 5.** Layout of the latest CMOS implementation

the current standard dimensions and, consequently, the design of special purpose hardware. The silicon solution not only requires a much smaller number of pixels but, more importantly, it also requires a lower consumption and has much faster read-out times – 33,000 pixels can be read about 36 times faster than 1,200,000 pixels. This advantage is bound to increase even more in the future when higher integration will be available.

### C. The Fovea of Log-Polar Sensors

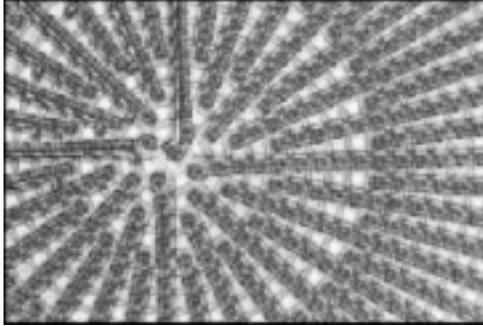
The equations of the log-polar mapping have a singularity at the origin where the size of the individual photoreceptors would theoretically go to zero. In practical terms this is not possible because – for silicon as well as biological photoreceptors – the technology limits the size of the smallest realizable photoreceptor<sup>1</sup>. For the sensors described here, therefore, the radius of the innermost ring of the log-polar representation is constrained by the size of the smallest pixel while the region inside this circle – the fovea – cannot follow this representation. With the CCD sensor this inner circular region was covered by a square array of pixels, all of the same size. The solution had two drawbacks. First, the square array inside the circular fovea left a portion of the visual scene unsampled; second, not only the logarithmic part of the representation was broken but also its polar components.

In the successive implementation we tried to improve on the design of the first sensor: the resulting geometry is shown in Fig. 6. The solution was to reduce the number of pixels in the foveal rings by

<sup>1</sup>In humans the smallest diameter of a photoreceptor in the fovea is of the order of  $1.5 \mu\text{m}$  while in the sensors realized so far the size of the smallest realizable photoreceptor was  $30 \mu\text{m}$ ,  $14 \mu\text{m}$ ,  $7 \mu\text{m}$  for the 2000, 8000 and 33,000 pixel sensors respectively.

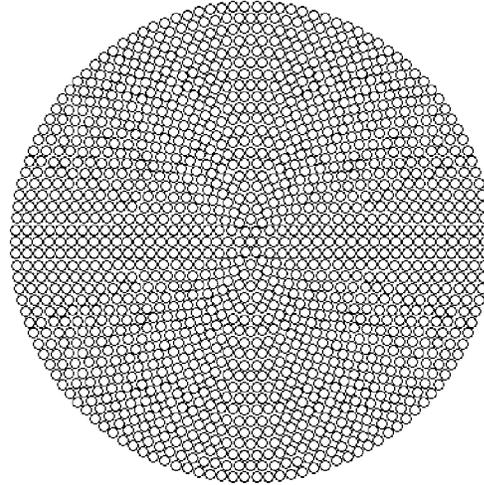
**Table 1.** Descriptive Parameters of the Latest CMOS Sensor

Peripheral Pixels	Foveal Pixels	Total Pixels	R	Q	Size (diameter)
27720 or $252 \times 110$	5473	33193	17	1100	7.1 mm

**Fig. 6.** Layout of the fovea of the IBIDEM CMOS sensor – 0.7  $\mu\text{m}$  technology

halving their number when necessary. For example, as in the periphery the structure has 128 pixels per ring and the size decreases toward the center, when the size of the pixels can no longer be reduced, the successive ring only accommodates 64 pixels until the technological limit is reached again and the number of pixels per ring is halved one more time. In summary, as it can be observed in Fig. 6, the fovea contains 10 rings with 64 pixels, 5 rings with 32 pixels, 2 rings with 16 pixels and 1 ring with 8, 4, and 1 pixel respectively. It is worth noting that by employing this arrangement the polar geometry is preserved and there is no empty space between the periphery and the fovea. However, continuity in terms of spatial resolution is not preserved because, whenever the number of pixels per ring is halved, the size of the pixels almost doubles.

In our latest implementation we adopted a different solution which is optimal in the sense that it preserves the

**Fig. 7.** Solution adopted for the arrangement of the photoreceptors in the fovea of our latest realization. With this topology the centermost region is covered uniformly with pixels of the same size

polar structure and covers the fovea with pixels of the same size. This solution is graphically shown in Fig. 7.

## IV. Applications

### A. Image Transmission

Built around a retina-like camera 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 (Giotto – see Fig. 8) and compress them following one of the recommendations for video coding over low bit rate communication line (H.263 in our case).



**Fig. 8.** Two different prototypes of the latest version of the Giotto camera. Both cameras are shown with a wide field of view lens ( $140^\circ$ ). The rightmost model has a standard C-Mount for the lens

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 Giotto camera; coding/decoding and image remapping is done in software on a 200-MHz 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.

### B. Robotics

As far as robotics is concerned the main advantage of log-polar sensors is related to the small number of pixels and the comparatively large FOV. For some important visual tasks images can be used as if they were high resolution. Typically, for example, in target tracking the fovea allows precise positioning and the periphery allows the detection of moving targets. This property is exploited to control the robot head of the Babybot (see (Metta 2000, Sandini 1997)). In this implementation a log-polar sensor with less than 3000 pixels is used to control in real-time the direction of gaze of a 5 degree-of-freedom head. Aspects of the image processing required were outlined in section II.

Particularly relevant is the possibility to use vision to learn a series of diverse behaviors ranging from multisensory integration to reaching for visually identified targets. Fundamental to any other visual task is the possibility of moving the cameras so that the fovea moves and explores interesting regions of the visual field. We implemented a series of eye-head visuo-motor behaviors including saccades, vergence, smooth pursuit and vestibulo-ocular reflexes (VOR). For example vergence is controlled by relying on a measure of the binocular disparity (Manzotti et al. 2001), and dynamically by using the  $D$  component of the optic flow (see equation (3)); see (Capurro et al. 1997) for more details.

In the case of the VOR the appropriate command, using a combination of visual and inertial information, is synthesized

by measuring the stabilization performance. The latter measure is an example of the kind of parameters which is best estimated by means of the optic flow. In this case the translational component provides information to the robot about the performance of the controller; a neural network can be trained on the basis of this information (Panerai et al. 2000, Panerai et al. 2002). Saccades and other visual control parameters are also learnt as shown in (Metta et al. 2000). Visual information about suitable targets is acquired in terms of eccentricity (the angular position with respect to the current gaze direction), binocular disparity, and optic flow as described in section II.

We also investigated the relationship between vision and the acquisition of an acoustic map of the environment (in neural network terms). In this case the robot, equipped with the appropriate learning rules, learnt autonomously the relationship between vision, sound, and saccade/pursuit behaviors (Natale et al. 2002). After a certain amount of training, the Babybot is able to orient towards a visual, acoustic, or visuo-acoustic target (a colored object emitting sound), thus bringing two non-homogeneous quantities (vision and sound) into the same reference frame.

Finally, eye-head-arm coordination was investigated. Under certain biological compatible hypotheses, we were able to show that the robot could learn how to transform visual information (coding the position of a target) into the sequence of motor commands to reach the target (Metta et al. 1999) and touch it. Also in this case, vision was the organizing factor to build autonomously and on-line an internal motor representation apt to control reaching.

It is worth stressing that the whole system runs on a network of a small number of PCs (four Pentium-class processors) at frame-rate and carries out on-line

learning – neural network weights are changed on the fly during the operation of the robot. While the reason this is possible is related to the overall number of pixels (a limiting factor in most implementations), it was not clear beforehand whether this same visual information was enough to sustain a broad range of behaviors. We believe that this implementation provides a proof by existence that indeed the amount of information required can be obtained visually, with a moderately computational burden, if the visual scene is sampled appropriately. Log-polar, in this sense, proved to be a good enough sampling strategy.

## V. Conclusions

The main advantage of retina-like sensors is represented by the compromise between resolution and field of view allowing high resolution as well as contextual information to be acquired and processed with limited computational power. The advantages of a silicon realization with respect to hardware and software remappers has been discussed showing that, as technology progresses, the retina-like approach will not only maintain its current advantages, but it is bound to become even more interesting in the application areas described – and possibly in others. Different realizations of log-polar sensors have been illustrated as well as two key applications. In particular the peculiarities of the retina-like sensors for real-time control of gaze and, in general, for robotics have been presented. This is certainly the most obvious use of a sensor topology that has been shaped by evolution to support the control of behavior while maintaining the balance with energy consumption and computational requirements. In spite of the great variety of eyes found in nature, the conventional camera solution based either on

increasing simultaneously the resolution and field of view or on the use of interchangeable or variable focal length lenses, has not survived. There is no doubt, in our view, that in the future, adaptable robots will have space-variant eyes or, conversely, the design will trade autonomy for batteries and computational power. The image transmission application demonstrates empirically that, in order to fully exploit a communication channel, it is better to eliminate useless information at the sensor's level than compressing information which is not used.

### Acknowledgements

The work described in this paper in relation to Babybot has been supported by the EU Project COGVIS (IST-2000-29375), MIRROR (IST-2000-28159). The research described on the retina-like CMOS technology has been supported by the EU Projects SVAVISCA (ESPRIT 21951) and AMOVITE (IST-1999-11156). The research on the CMOS sensors has been carried out in collaboration with IMEC and FillFactory (Lou Hermans and Danny Scheffer), AITEK S.p.A. (Fabrizio Ferrari, Paolo Questa) and VDS S.r.l. (Andrea Mannucci e Fabrizio Ciciani). Features of the retina-like sensor topologies described are covered by patents. Giorgio Metta is Postdoctoral Associate at MIT, AI-Lab with a grant by DARPA as part of the "Natural Tasking of Robots Based on Human Interaction Cues" project under contract number DABT 63-00-C-10102.

### References

- Baron T, Levine MD, Hayward V, Bolduc M, Grant DA (1995) A biologically-motivated robot eye system. Paper presented at the 8th Canadian Aeronautics and Space Institute (CASI) Conference on Astronautics, Ottawa, Canada
- Baron T, Levine MD, Yeshurun Y (1994) Exploring with a foveated robot eye system. Paper presented at the 12th International Conference on Pattern Recognition, Jerusalem, Israel
- Blough PM (1979) Functional implications of the pigeon's peculiar retinal structure. Granda AM, Maxwell JM (eds) *Neural Mechanisms of Behavior in the Pigeon* (pp. 71–88). New York, NY: Plenum Press
- Capurro C, Panerai F, Sandini G (1997) Dynamic vergence using log-polar images. *Int J Comput Vision* 24: 79–94
- Carpenter RHS (1988) *Movements of the Eyes* (Second ed.). London: Pion Limited
- Darrell T, Gordon G, Harville M, Woodfill J (2000) Integrated person tracking using stereo, color, and pattern detection. *Int J Comput Vision* 37: 175–185
- Engel G, Greve DN, Lubin JM, Schwartz EL (1994) Space-variant active vision and visually guided robotics: design and construction of a high-performance miniature vehicle. Paper presented at the International Conference on Pattern Recognition, Jerusalem
- Galifret Y (1968) Les diverse aires fonctionelles de la retine du pigeon. *Z Zellforsch* 86: 535–545
- Koenderink J, Van Doorn J (1991) Affine structure from motion. *J Optical Soc Am* 8: 377–385
- Manzotti R, Gasteratos A, Metta G, Sandini G (2001) Disparity estimation in log polar images and vergence control. *Comput Vis Image Und* 83: 97–117
- Metta G (2000) Babybot: a study on sensori-motor development. Unpublished Ph.D. Thesis, University of Genova, Genova
- Metta G, Carlevarino A, Martinotti R, Sandini G (2000) An incremental growing neural network and its application to robot control. Paper presented at the International Joint Conference on Neural Networks, Como, Italy
- Metta G, Sandini G, Konczak J (1999) A developmental approach to visually-guided reaching in artificial systems. *Neural Networks* 12: 1413–1427
- Natale L, Metta G, Sandini G (2002) Development of auditory-evoked reflexes: visuo-acoustic cues integration in a binocular head. *Robot Auton Syst* 39: 87–106
- Panerai F, Metta G, Sandini G (2000) Visuo-inertial stabilization in space-variant binocular systems. *Robot Auton Syst* 30: 195–214
- Panerai F, Metta G, Sandini G (2002) Learning stabilization reflexes in robots with moving eyes. *Neurocomputing*, In press

- Roger A, Schwartz EL (1990) Design considerations for a space-variant visual sensor with complex-logarithmic geometry. Paper presented at the 10th International Conference on Pattern Recognition, Atlantic City, USA
- Sandini G (1997) Artificial systems and neuroscience. Paper presented at the Otto and Martha Fischbeck Seminar on Active Vision, Berlin, Germany
- Sandini G, Tagliasco V (1980) An anthropomorphic retina-like structure for scene analysis. *Comp Vision Graph* 14: 365–372
- Schwartz EL (1980) A quantitative model of the functional architecture of human striate cortex with application to visual illusion and cortical texture analysis. *Biol Cybern* 37: 63–76
- Srinivasan MV, Venkatesh S (1997) *From Living Eyes to Seeing Machines*. London: Oxford University Press
- Tunley H, Young D (1994) First order optical flow from log-polar sampled images. Paper presented at the Third European Conference on Computer Vision, Stockholm
- Van der Spiegel J, Kreider G, Claeys C, Debusschere I, Sandini G, Dario P, Fantini F, Bellutti P, Soncini G (1989) A foveated retina-like Sensor using CCD technology. In: Mead C, Ismail M (eds), *Analog VLSI Implementation of Neural Systems* (pp. 189–212). Boston: Kluwer Academic Publishers
- Wallace RS, Ong PW, Bederson BB, Schwartz EL (1994) Space variant image processing. *Int J Comput Vision* 13: 71–91
- Weiman CFR (1988) 3-D Sensing with polar exponential sensor arrays. Paper presented at the SPIE – Digital and Optical Shape Representation and Pattern Recognition
- Weiman CFR, Chaikin G (1979) Logarithmic spiral grids for image processing and display. *Computer Graphic and Image Processing* 11: 197–226
- Weiman CFR, Juday RD (1990) Tracking algorithms using log-polar mapped image coordinates. Paper presented at the SPIE International Conference on Intelligent Robots and Computer Vision VIII: Algorithms and Techniques, Philadelphia (PA)