

A RETINA-LIKE CMOS SENSOR AND ITS APPLICATIONS

Giulio Sandini (1), Paolo Questa (2), Danny Scheffer (3), Bart Dierickx (3), Andrea Mannucci (4).*

(1)LIRA-Lab - DIST, University of Genova

(2) Aitek S.r.l., Genova - Italy

(3) IMEC – Belgium

(4) Video Display Systems S.r.l. - Italy

sandini@dist.unige.it

ABSTRACT

This paper describes the implementation in silicon of a retina-like sensor characterized by a space-variant resolution similar to that of the human retina. Besides the actual implementation using CMOS technology, the design peculiarities of retina-like sensors are compared with the use of software or hardware electronic remappers. The paper also describes the realization of a retina-like digital camera and its applications for real-time control and image transmission.

1. INTRODUCTION

Retina-like visual sensors are characterized by a space-variant resolution mimicking the distribution of photoreceptors in the human retina. Density of photoreceptors is highest in the center (limited by the particular technology used) and decreases monotonically toward the periphery of the Field of View (FOV). The resulting image is, consequently a compromise between resolution, amplitude of the FOV, and number of pixels. Such space-variant image is peculiar because it allows high-resolution tasks using the central region while maintaining a lower resolution part providing relevant information about the background. Such 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 the one described in this paper is the so-called *log-polar* structure [1-3]. In this schema a constant number of photosites are 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 Figure 1. Because of the polar structure of the photosensitive array and the increasing size of pixels in the periphery, retina-like sensor, do not provide images with a *standard* topology.

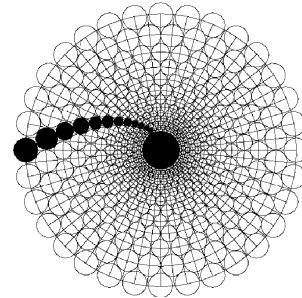
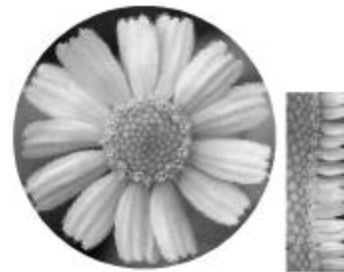


Figure 1: Layout of receptor's placing for a log-polar structure composed of 12 rings with 32 pixels each. The pixels marked in black follow a logarithmic spiral.

In fact in the log-polar format a point in the image plane at polar coordinates (r, θ) is mapped into a Cartesian plane $(\log(r), \theta)$. It is worth noting that the mapping is obtained at no computational cost, as it is a direct consequence of the photosites array and the read-out sequence.



*The research described here is supported by a grant of the European Union (FET project SVAVISA). Danny Scheffer and Bart Dierickx are presently at Fillfactory- Belgium

Figure 2: 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 Figure 2. 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[1].

2. SOLID STATE LOG-POLAR SENSORS

Log-polar mapping has been obtained in two different ways. The first is by means of electronic boards transforming in real-time standard images into log-polar ones. Historically this has been the first way implemented for real-time control and image compression [4-7]. 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 of which the most important are the minimum pixel size and the maximum sensor size.

Our first solid-state realization of a log-polar sensor was realized at the end of '90 using CCD technology [8]. At that time with the technology available to us the size of the smallest possible pixel was about 30 μm and for practical limitations the overall sensor size (diameter) was limited to 94 mm. A picture of the layout is shown in Figure 3.

This sensor is composed of 30 rings and each ring is covered by 64 pixels. Figure 3 will help clarifying a few peculiarities of log-polar sensors. First of all the fact that CCD technology is not really suited to realize polar structures because part of the sensing area has to be covered by the read-out shift register and, consequently, the sensor has a blind "wedge". Secondly, the increase of pixel's size with eccentricity¹ posed some problems because of the relationship between the sensitivity of pixels and their size [8]. Third is the solution adopted for the centermost part of the array. In this initial implementation the central part, which we call *fovea* in analogy with the centermost, high-resolution part of the human retina, was partially covered by a diamond-shaped

¹ Eccentricity is defined as the distance between a pixel and the center of the polar structure.

constant resolution array. Altogether the sensor had 2022 pixels, 1920 of which in the log-polar part of the sensor.

2.1 Parameters of log-polar sensors

Starting from the above-mentioned technological constraints (minimum pixel size and size of the sensor), the most obvious sensor's parameter is the total number of pixels, which is directly related to the amount of information acquired.

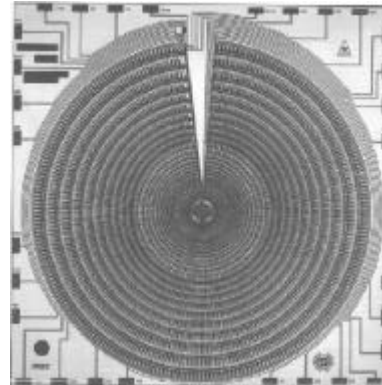


Figure 3: Structure of the first log-polar sensor realized with CCD technology.

For constant resolution sensors this parameter is fixed by the technological constraints in a straightforward way. Also for log-polar sensors these two parameters are linked even if not in a straightforward way (see [2, 7, 9] 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 the CCD implementation this ratio was about 13.7 (the largest pixel was 412 μm). This parameter illustrates the amount of "space variability" of the sensor and is, of course, equal to 1 in standard constant resolution sensors. The third important parameter is the ratio between sensor's size and 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 with the same field of view and the same maximum resolution of the corresponding log-polar sensor. For example for the CCD sensor Q is equal to about 300 meaning that if we want to electronically re-map a constant resolution image and obtain the same amount of information obtained by our log-polar sensor, the original image must be at least 300x300 pixels.

2.2 8,000 pixels log-polar CMOS sensors.

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 CCD technology adopted. In our recent implementations, a CMOS technology was adopted based on the FUGA pixel implemented at IMEC. This pixel has a logarithmic response to illumination and, consequently, the sensitivity of the largest and smaller pixels is comparable. The first CMOS implementation was carried out within a research project supported by the European Union under the TIDE initiative (Technology Initiative for Disabled and Elderly). The objective of the project was to demonstrate that using a camera with a retina-like sensor it was possible to acquire and transmit images that could be used by deaf and hard of hearing persons to understand signing and finger-spelling. The constraint was to use standard telephone lines. The rationale of the project was that a retina-like sensor provides, with a small number of pixels, both the resolution required to understand, for example, the shape of the lips, as well as a wide field of view to understand the facial expression and/or gestures. The small number of pixels, in addition, allows the high frame rate required to convey the dynamic information contained in the stream of images. The resulting structure is shown in Figure 4.

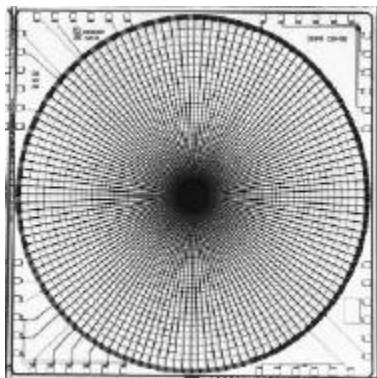


Figure 4: Layout of the IBIDEM sensor realized with CMOS technology.

The important parameters of the sensor are:

Peripheral Pixels	Foveal Pixels	Total Pixels	R	Q	Size (diameter)
7168 (128x56)	845	8013	14	600	8.1 mm

The sensor itself was realized using a 0.7 μ m technology allowing the minimum pixel size to be 14 μ m. From our preliminary user's test, the requirement in terms of frame-rate was to be able to obtain about 15 frames/s. For this reason a compression engine was realized based on the H263 standard. The results of the final tests were published in [10] and demonstrated that using our retina-like technology we were able to obtain the same results on

analog telephone lines, as it was possible over ISDN lines. A sample image acquired by the sensor in its log-polar and remapped version is shown in Figure 5.

The following technological step was to transform this CMOS sensor into a color one [11]. Two patterns for microfilters were tested (see Figure 6). From a qualitative point of view, the patterns shown on the left gave better results particularly in regions with color discontinuities.

A peculiarity of log-polar sensors is related to the distribution of photoreceptors in the central, high resolution, part of the sensor. For this sensor the minimum pixel size was 14 μ m.



Figure 5: Sample image acquired during a finger-spelling test. Note that the log-polar image is shown enlarged to present its topology. The image is composed of 8013 pixels.

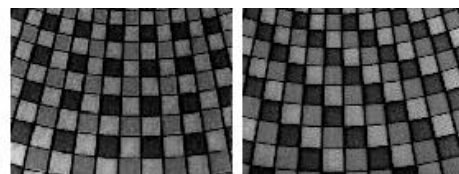


Figure 6: Patterns tested for the deposition of micro filters. Even in this black and white picture the difference should be evident. On the left green pixels are more than blue and red ones while on the right an equal number of pixels is used for red, green and blue filters.

In this implementation we tried to solve two of the major drawbacks of the first sensor. In particular what we wanted to remove was the "un-covered" area resulting from the square shape of the fovea and the round shape of the innermost circle of the peripheral part. Furthermore we wanted to maintain, if not the log-polar structure, at least the polar arrangement of the pixels also in the fovea. The resulting geometry is shown in Figure 7. The solution was to reduce the number of pixels in the foveal rings by 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 cannot be reduced any more¹, 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 should be clear in Figure 7, 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 using this arrangement the polar geometry is preserved and there is no "empty" space between the periphery and the fovea.

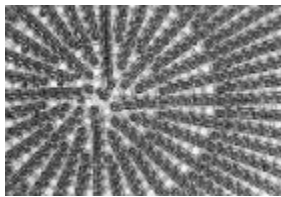


Figure 7: Layout of the fovea of the IBIDEM CMOS sensor.

However, it is fair to say that 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.

2.3 33,000 pixels log-polar sensor

Our most recent CMOS implementation was realized using a 0.35 μ m technology (realized at Tower in Israel). This sensor is being developed within a recent EU-funded research project called SVAVISA. The goal of the project is to realize, besides the sensor, a microcamera with a special-purpose lens allowing 140° field of view. The miniaturization of the camera is possible because some of the electronics required to drive the sensor as well as the A/D converter is now included on the chip itself. A picture of part of the layout of the sensor is shown in Figure 8. The following table summarizes the main parameters of the sensor.

Peripheral Pixels	Foveal Pixels	Total Pixels	R	Q	Size (diam.)
27720 or 252x110	5473	33193	17	1100	7.1 mm

The fourfold increase in number of pixels was obtained without increasing the size of the sensor because of the updated technology used and the structure of the fovea allows now a better coverage of the central part removing some of the drawbacks of the previous implementation. It

¹ This represents the boundary between the fovea and the periphery.

is worth noting that the **Q** parameter is now 1100 (compared to 300 and 600 of our previous implementation) without reducing the figure **R** measuring the "space-variacy" of the sensor. In practical terms what this parameter says is that if one wants to produce the same retina-like image by re-mapping a standard constant-resolution image, it is necessary to start from an image of size 1100x1100 which corresponds to more than 35 times more pixels. We believe this observation to be particularly relevant because it demonstrates that a silicon realization of the retina-like structure is becoming more and more advantageous with respect to other methods based on re-mapping standard images.

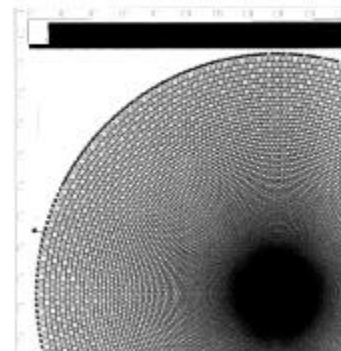


Figure 8: Layout of the latest CMOS implementation.

A sample image acquired by the new sensor is shown in Figure 9. In fact, if we consider the first implementation with a **Q** equal to 300, even if the total number of pixels of the 300x300 image was still 40 times larger than the 2000 pixels retina-like image, its size was still manageable. The current sensor, if simulated using a software or hardware remapper would require an image with a number of pixels exceeding the current standard dimensions and, consequently, the design of a special purpose sensor. The silicon solution not only requires a much smaller number of pixels but, consequently, requires a lower consumption and much faster read-out times (reading 33,000 pixels takes about 36 times less time than reading 1,200,000 pixels). This advantage is bound to increase even more in the future when higher integration will be possible.



Figure 9: Sample image "Danny" acquired by the latest CMOS sensor.

3. APPLICATIONS

Log-polar cameras have been used, mainly, within two application areas: visually driven real-time control and image transmission. It is outside the scope of this paper to describe in details the advantages of the log-polar geometry for image processing. The interested reader is referred to the following publications such as, for example: [1, 4, 7, 12-20].

3.1 Robotics and real-time control

As far as robotics is concerned the main advantage related to the small number of pixels and the comparatively large FOV, is the fact that for some important visual tasks images can be used as if the all image was high resolution and, at the same time the lower resolution in the periphery provides vital information to understand what is happening in the field of view. Typically, for example, in target tracking where the fovea allows precise positioning of the tracking device and periphery allows the detection of moving targets. This peculiarity is used to control the robot head of our *babybot* humanoid (see [21, 22]). 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 degrees of freedom head using a Pentium PC. Particularly relevant is the possibility of controlling in a simple way the vergence angle of the stereo-head [23] as well as the integration of visual and inertial information for image stabilization [24, 25].

3.2 Image transmission

As a result of the IBIDEM project a prototype camera based on the 8,000 pixels CMOS sensor was realized (Giotto¹). A picture of the camera is shown in Figure 10.



Figure 10: Picture of the Giotto Camera. The camera is interfaced directly to the parallel (printer) port of a PC in EPP mode.

Built around the Giotto camera extensive experiments on wireless image transmission were conducted. The set-up was composed of a remote PC running a web server embedded into an application that acquires images from the Giotto camera and compress them following the H.263 recommendation for video coding over low bit rate communication line. The remote station was, in our case, 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 can 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, up to 4 images per second can be transmitted of a traffic scene or a speaking person. Moreover it is worth noting that the only special-purpose hardware device is the Giotto Camera and all coding/decoding and a Pentium 200 MHz PC does image remapping in software. The GSM connection is obtained with an off-the-shelf NOKIA Card connected to the PCMCIA interface of a Toshiba *Libretto* accessing the Omnitel Pronto Italia voice communication GSM network. This performance in terms of frame rate, image quality and cost cannot be afforded with systems using conventional cameras. Figure 11 shows some sample images as they appear at the remote site after compression, transmission, decompression [26].

¹ More information about Giotto can be found at: www.aitek.it. Giotto has won the European IST Prize for innovation in 1999 and is among the finalists of the EDN 1999 Innovation Awards

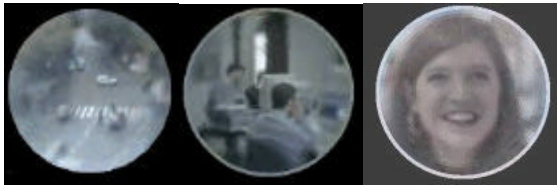


Figure 11: Image acquired at a remote site after transmission over a GSM line. Coding and decoding using H263 standard is done by the host PC with no additional hardware.

4. CONCLUSIONS

Different realizations of log-polar sensors have been described as well as some applications. 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 remapper has also been presented 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).

5. REFERENCES

- [1] Weiman, C.F.R. and G. Chaikin, *Logarithmic Spiral Grids for Image Processing and Display*. Comp. Graphic and Image Process., 1979. **11**: p. 197--226.
- [2] Sandini, G. and V. Tagliasco, *An Anthropomorphic Retina-like Structure for Scene Analysis*. CVGIP, 1980. **14**(3): p. 365-372.
- [3] Schwartz, E.L., *A Quantitative Model of the Functional Architecture of Human Striate Cortex with Application to Visual Illusion and Cortical Texture Analysis*. Biological Cybernetics, 1980. **37**: p. 63-76.
- [4] Weiman, C.F.R. and R.D. Juday. *Tracking Algorithms Using Log-Polar Mapped Image Coordinates*. in *SPIE Int. Conf. on Intelligent Robots and Computer Vision VIII: Algorithms and Techniques*, 1990. Philadelphia (PA),.
- [5] Engel, G., et al., *Space-Variant Active Vision and Visually Guided Robotics: Design and Construction of a High-Performance Miniature Vehicle*, . 1994, IEEE Computer Society, p. 487-490.
- [6] Rojer, A.S. and E.L. Schwartz, *Design Considerations for a Space-Variant Visual Sensor with Complex-Logarithmic Geometry*, . 1990, IEEE Computer Society, p. 278--285.
- [7] Wallace, R.S., et al., *Space Variant Image Processing*. International Journal of Computer Vision, 1994. **13**(1): p. 71-91.
- [8] Spiegel, J.V.d., et al., *A Foveated Retina-like Sensor Using CCD Technology*, . 1989, De Kluwer,.
- [9] Weiman, C.F.R. *Exponential sensor array geometry and simulation*. in *Proc. SPIE, Vol 938*. 1988.
- [10] Woelders, W.W., et al., *New Developments in Low-bit Rate Videotelephony for People Who are Deaf*. Journal of Speech, Language and Hearing Reserach,, 1997.
- [11] Sandini, G., et al. *The Project SVAVISCA: a Space-Variant Color CMOS Sensor*. in *Advanced Focal Plane Arrays and Electronic Cameras II*. 1998: SPIE.
- [12] Questa, P., E. Grossmann, and G. Sandini. *Camera self orientation and docking maneuver using normal flow*. in *Proc. SPIE AeroSense95*,. 1995. Orlando, U.S.A.,.
- [13] Weiman, C.F.R. *3-D Sensing with polar exponential sensor arrays*. in *Proc. SPIE, Vol 938*. 1988: SPIE.
- [14] Weiman, C.F.R. *Polar Exponential Sensor Arrays Unify Iconic and Hough Space Representation*. in *SPIE Int. Conf. on Intelligent Robots and Computer Vision VIII: Algorithms and Techniques*,. 1990.
- [15] Massone, L., G. Sandini, and V. Tagliasco, *Form-Invariant Topological Mapping Strategy for 2D Shape Recognition*. Computer Vision Graphics and Image Processing,, 1985. **30 No.2**: p. 169--188.
- [16] Braccini, C., et al., *A Model of the Early Stages of the Human Visual System: Functional and Topological Transformation Performed in the Peripheral Visual Field*. Biological Cybernetics,, 1982. **44**: p. 47--58.
- [17] Tistarelli, M. and G. Sandini, *On the advantages of polar and log-polar mapping for direct estimation of time-to-impact from optical flow*. IEEE Transactions on PAMI,, 1993. **14**(4): p. 401--410.
- [18] Jain, R.C., S.L. Bartlett, and N.O. Brian, *Motion Stereo Using Ego-Motion Complex Logarithmic Mapping*. IEEE Trans. on PAMI,, 1987. **PAMI-9**(3): p. 356--369.
- [19] Yeshurun, T. and E.L. Schwartz, *Shape Description With a Space-Variant Sensor: Algorithms for Scan-Path, Fusion, and Convergence Over Multiple Scans*. IEEE Trans. PAMI,, 1989. **11**(11): p. 1217--1222.
- [20] Fischl, B., M.A. Cohen, and E.L. Schwartz, *The Local Structure of Space-variant Images*. Neural Networks,, 1997. **10**(5): p. 815-831.
- [21] Sandini, G., G. Metta, and J. Konczak. *Human Sensori-motor Development and Artificial Systems in AIR&IHAS '97*. 1997.
- [22] Metta, G., G. Sandini, and J. Konczak, *A Developmental Approach to visually-guided reaching in artificial systems*. Neural Networks, 1999. **12**(10): p. 1413-1427.

- [23]Capurro, C., F. Panerai, and G. Sandini, *Dynamic Vergence using Log-polar Images*. International Journal of Computer Vision., 1997. **24**(1): p. 79--94.
- [24]Panerai, F. and G. Sandini, *Oculo-Motor Stabilization Reflexes: Integration of Inertial and Visual Information*. Neural Networks, 1998. **11**: p. 1191-1204.
- [25]Panerai, F., G. Metta, and G. Sandini, *Visuo-inertial Stabilization in Space-variant Binocular Systems*. Robotics and Autonomous Systems, 2000. **30**(1-2): p. 195-214.
- [26]Sandini, G., *et al. Image Transmission with a Retina-like CMOS Camera*. in *IEEE Workshop on Charged-Coupled Devices and Advanced Image Sensor*. 1999. Nagano.