

# Adaptive Fovea Structures for Space-Variant Sensors

*Pelegrín Camacho , Fabián Arrebola and Francisco Sandoval*  
Dpto. Tecnología Electrónica - E.T.S. Ingenieros de Telecomunicación  
Universidad de Málaga - Campus de Teatinos, 29071- Málaga - SPAIN  
E-mail : pelegrin@dte.uma.es

## Abstract

In this paper we describe the architecture and data structure of space-variant sensors with reconfigurable cartesian geometries. The ability of these sensors to change the position and size of their high resolution regions or electronic foveas, makes them suitable to compensate the limited performance or coarse fixation characteristics of the mechanical systems utilized for gaze tasks in active vision applications where size, weight or cost could be conditioning factors to the performance or feasibility of the whole system. An alternative to the implementation of these sensors is based on off-the-shelf CCD cameras and devices with reconfiguration capabilities, such as FPGAs. In this way, besides the multiresolution data output, sensor reconfiguration systems let generate additional data adapted to the functions of the higher level modules of the active vision systems. As a result of this computing capability at the sensor level, it is possible to unload the processing stages of certain tasks without penalty in time or significant addition of hardware. An approach to selective foveation tasks and motion detection is presented.

## 1. Introduction

There are technical applications based on active vision systems requiring wide field of view -FOV- and high resolution capability. Facing the implementation of such a systems, several research groups focused their efforts to the development of spatially variant sensors having structures with non-uniform resolution across polar geometries, resembling the retinotopology of the human eye [1][2][3]. An approach based on cartesian topologies was proposed by Bandera and Scott [4], having fovea centered lattices surrounded by sets of concentric rings. The advantages of multiresolution sensing -selective data reduction in space, resolution and time- are obtained at the expense of complex gaze fixation mechanisms to place the sensor fovea onto the regions of interest, considering both the position and size of the objects. Thus, in relation to the development of multiresolution sensors two aspects must be considered: firstly the number of elements, their size and placement on the sensor structure, since they will determine the number and the acuity of the different resolution cells, the levels of the sensor and its data compression and, secondly, the dimensional ratio of the resolution levels or resolution profile because, depending on the size and singularities of the objects, multiple foveations could be required for recognition, leading to complex gaze control procedures. Within the architectures of vision systems [5], our approach has been oriented to space-variant sensors with reconfigurable structures, looking for adaptability to the size and position of the objects of interest, in emulation of saccadic eye movements with an electronic fovea.

## 2. Rectangular Shifted-fovea Structures for Space-variant Sensors

Cartesian multiresolution geometries [4] consist of concentric structures with a fovea surrounded by rings with decreasing resolution. They are characterized with the following parameters:

$m$  : Number of concentric rings surrounding the central fovea.

$h, v$  : Horizontal and vertical subdivision factors, giving the number of resolution cells -*rexels*- or subrings found in the directions of the cartesian axes within any of the  $m$  resolution rings. Fovea size is fixed to  $4h \times 4v$  pixels.

Using off-the-shelf CCD cameras, we have applied reconfigurability of non-concentric structures [6], to obtain rexels by pixel averaging on rectangular FOVs sized  $H$  and  $V$  pixels. Eccentricity is fixed by shifting all rings  $s_h, s_v$  rexels, as seen in Fig. 2.a. The minimum jump,  $J_N$ , for any ring  $N$ , corresponding to  $|s_h|$  or  $|s_v| = 1$ , is the shift of that ring across the FOV, caused by the simultaneous and relative shifts of all rings within the rings surrounding them. The jump, in pixels, is given by

$$J_N = 2^{N+1} + 2^{N+2} + \dots + 2^m = \sum_N^{m-1} 2^{i+1} \quad (1)$$

valid for all rings, except the peripheral one delimiting the FOV. The effects can be seen on the foveal image of Fig.1.b obtained with a structure having  $m = 3$  and whose fovea was shifted onto the car at the right in the uniresolution image of Fig.1.a. Resolution levels associated to the foveal image for  $m = 2$  are shown in Fig. 1.c. Shifted fovea sensors allow data structures with an equal number of elements on all resolution levels. The size of the fovea and upper levels is fixed, as seen in Fig. 1.c., because once the number of rings has been defined, the values of  $h$  and  $v$ , their ratio and the fovea size  $-4h \times 4v-$  are uniquely determined by the expressions

$$h = \frac{H}{2^{m+2}} \quad ; \quad v = \frac{V}{2^{m+2}} \quad (2)$$

The fact of having a fovea with fixed dimensions implies that targets with a size greater than fovea could only be recognized by refoveating onto the targets, unless the details leading to recognition could be identified in a higher level because of its wider solid angle, and in spite of its lower resolution. In case that neither of those options would be feasible on the system, the alternative would be to reduce  $m$ , i.e. to increase the fovea dimensions. However, the convenience of reducing the number of resolution levels must be considered in relation to the increase of data involved: Structures with low  $m$ , i.e. having a smooth resolution profile on a fixed FOV, have a reduced compression factor which is equivalent to have a foveal sensor with greater number of elements. Considering the expressions in (2) and the relative sizes of fovea and rings, the compression factor  $F$ , defined as the ratio of the number of pixels on the FOV to the number of elements in sensor structures as that of Fig.2.a, will be given by

$$F = \frac{H \cdot V}{4h \cdot 4v \cdot (1 + \frac{3m}{4})} = \frac{h \cdot 2^{m+2} \cdot v \cdot 2^{m+2}}{h v \cdot (16 + 12m)} = \frac{2^{2m+4}}{16 + 12m} \quad (3)$$

where it can be seen that the higher the value of  $m$ , the higher the compression factor.

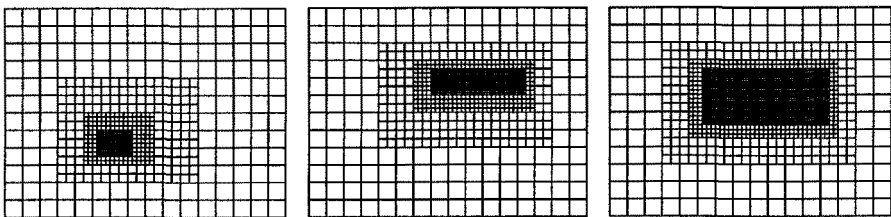


a) Original image 640x480 pixels

b) Foveal image ( $m=3$   $h, v, s_h, s_v = 20, 15, 11, 1$ )c) Prism resolution levels : Fovea, L1, L2, for  $m=2$ . (Expanded  $\approx 3 \times 3$ )**Fig. 1 Foveal image and resolution levels of a shifted fovea structure.**

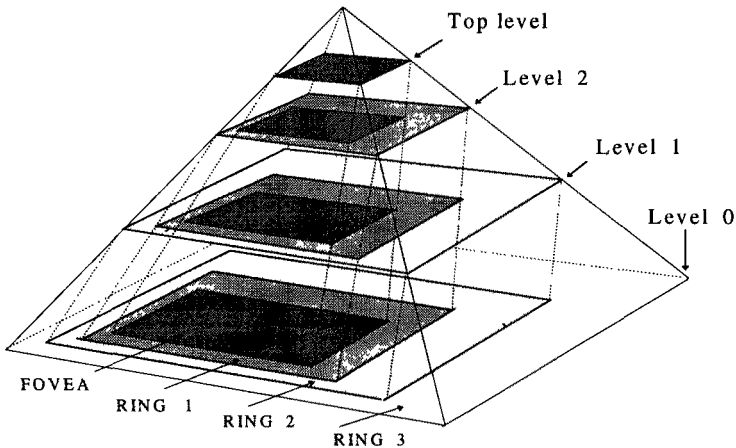
### 3. Rectangular Lattices with Adaptive Data Structure

To avoid the inconveniences of a low  $m$ , without losing the advantages of greater fovea, the approach here proposed has the double effect of expanding the fovea and the reconfigurability of shifted fovea geometries, without increasing the data volume involved, as will be shown later. The procedure consists in making different resolution profiles at each side of the fovea and combining them horizontal and vertically. To do that, as seen in Fig.2.b and 2.c, instead of parameters  $h, v, s_h, s_v$ , we introduce a new set of four parameters  $l, r, t, b$  to define subdivision factors at each side of fovea. Resolution profiles increase or decrease inversely with lateral subdivision factors. Thus, in Fig. 2.b, at the right of fovea, the gradient is twice the one at the left. Departing from structure in Fig.2.b, the resolution gradients at the left and bottom regions could be increased in one step, leading to the structure of Fig. 2.c, with a greater fovea, which is placed according to the set of parameters  $l, r, t, b$ .

a)  $h, v, s_h, s_v = 4, 3, -1, 1$ b)  $l, r, t, b = 4, 2, 2, 4$ c)  $l, r, t, b = 3, 2, 2, 3$ **Fig. 2 Rectangular structures with  $m = 3$ . a) Shifted fovea, b) and c) Adaptive fovea.**

The variation of lateral resolution profiles is shown in Fig.3, where it can be seen a tilted and truncated inner pyramid sharing its top level with the coarsest one of the outer pyramid, containing the compressed FOV. Projection of that top level onto the FOV, performed through the truncated pyramid, will determine the size and location of fovea in relation to the projecting angles at each side of inner pyramid. Parallel projections of the outer pyramid levels onto the FOV, will determine lateral resolution gradients and the hierarchical data structure of the sensor. At the limit, both pyramids become one, when adaptive fovea is expanded up to cover the whole FOV.

The sensors so obtained will have asymmetric structures whose highest resolution levels can be sized to cover the shape and position of any region of interest or target on the FOV. Fovea adaptation improves greatly the performance of the vision system by eliminating the refoveation processes required when fovea size is smaller than target and, therefore, optimizing data handling and processing time for recognition purposes. Thus, structures of Fig.2 have the same number of resolution levels, but changing the lateral gradients in a regular manner, we are able to fix the asymmetry of the structure and the ratio of each level  $N$  to the  $N+1$ . That ratio and the data structure associated is regularly maintained in an homothetic way which resembles the outer pyramid data compression.



**Fig. 3 Merged pyramids structure for adaptive fovea lattices with  $m = 3$ .**

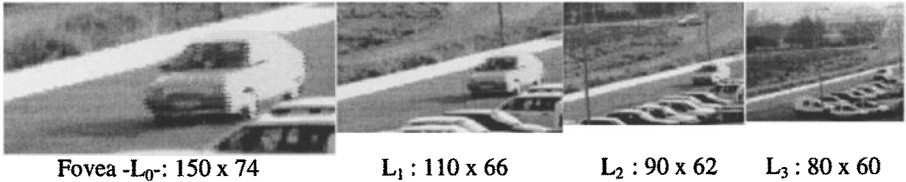
To configure the structure to the size and location of the target, it is assumed that target has an associated bounding box whose coordinates  $X_{min}$ ,  $X_{max}$ ,  $Y_{min}$ ,  $Y_{max}$  were previously determined after applying the appropriate algorithms for objects detection. According to these set of coordinates it is inferred the minimum fovea dimensions as well as its location on the FOV. The remaining steps to apply to the algorithm determining the structure consist on the determination of the maximum subdivision factors to have on each side of the fovea, without overskipping the bounding box coordinates. Thus, lateral subdivision factor will be obtained as :

$$\begin{aligned}
\text{Left subdivision factor} \quad l &= INT \left[ \frac{X_{min}}{J_0} \right] \\
\text{Right subdivision factor} \quad r &= INT \left[ \frac{H - X_{max}}{J_0} \right] \\
\text{Top subdivision factor} \quad t &= INT \left[ \frac{Y_{min}}{J_0} \right] \\
\text{Bottom subdivision factor} \quad b &= INT \left[ \frac{V - Y_{max}}{J_0} \right]
\end{aligned} \tag{4}$$

where  $J_0$ , as defined in (1), refers to the minimum jump to apply on fovea sides to alter the resolution gradient one step. After obtaining  $l$  and  $t$ , all inner rings and fovea can be referred to the origin or upper left corner of the FOV. The origin,  $O$ , and the opposite corner,  $Z$ , of any level  $N$  are thus related to the jump,  $J_N$ , and their coordinates given by

$$O_{xN} = J_N.l \quad O_{yN} = J_N.t \quad Z_{xN} = H - J_N.r \quad Z_{yN} = V - J_N.b \tag{5}$$

After applying reconfiguration to image in Fig.1.a, resolution levels are obtained as shown in Fig.4, with lateral dimensions given in number of elements, fovea in pixels and upper levels,  $L_i$ , in their rexels equivalent to  $2^i$  pixels. It can be seen the relative eccentricity of levels, their different sizes as well as the homothecy relating them.



**Fig. 4 Adaptive fovea levels obtained with  $m = 3, l, r, t, b = 32, 3, 18, 11$ .**

To evaluate the effectiveness of adaptive fovea structures -AFSS- the number of cells in these lattices is compared to that in shifted fovea structures. That number is obtained by adding the expressions giving the cells in fovea and rings to derive the expression for the whole structure, considering that the number of subrings on each side of fovea is determined by  $l, r, t, b$ . Fovea and rings sizes are given by the differences between the lengths of the FOV and the respective widths of the rings to each side of the fovea, or ring considered, therefore, the magnitudes, in pixels, are

$$\begin{aligned}
\text{Horizontal width of level } N \quad W_{hN} &= H - J_N.(l + r) = H - J_N.x \\
\text{Vertical width of level } N \quad W_{vN} &= V - J_N.(t + b) = V - J_N.y \\
\text{Pixels related to } L_N \text{ or Size of level } N \quad S_N &= (H - J_N.x).(V - J_N.y)
\end{aligned} \tag{6}$$

where  $x$  and  $y$  refer to total resolution steps horizontal and vertically.

Therefore, the number of elements -rexels- on ring  $N$ ,  $R_N$ , will be given by

$$R_N = \frac{\text{Size of level } N - \text{Size of level } (N-1)}{\text{Size of rexel in ring } N} \tag{7}$$

The size, in pixels, of one rexel in ring  $N$  is given by the exponential  $2^{2N}$ . From (6) and (7), the following expressions are obtained

$$\begin{aligned} \text{Elements in Fovea} \quad R_0 &= HV - J_0 \cdot (x + y) + x \cdot y \cdot J_0^2 \tag{8} \\ \text{Elements in Ring } N \quad R_N &= \frac{2^N \cdot (Hy + Vx) + x \cdot y \cdot (J_N^2 - J_{N-1}^2)}{2^{2N}} \\ \text{Elements in Ring } m \quad R_m &= \frac{2^m \cdot (Hy + Vx) - x \cdot y \cdot J_{m-1}^2}{2^{2m}} \end{aligned}$$

and the total number of elements in any adaptive sensor lattice will be given by

$$R_T = \sum_0^m R_i = H \cdot V - (H \cdot y + V \cdot x) \cdot \sum_1^m \left( 2^i - \frac{1}{2^i} \right) + \frac{3 \cdot x \cdot y}{4} \cdot \left[ J_0^2 + \sum_1^{m-1} (J_0 - J_i)^2 \right] \tag{9}$$

from which it can be obtained, as a particular case for  $x = 2h$  and  $y = 2v$ , the expression for the number of elements in fixed size fovea lattices, given by

$$R_i = h v ( 16 + 12m ) \tag{10}$$

Table I shows the number of elements in AFSs for some pairs of  $x$  and  $y$  in lattices with  $m = 3$ . Sensor elements are compared with those of shifted fovea structures having  $m = 2$  and related to the number of pixels on their respective foveas.

**TABLE 1: Comparison of foveal structures (H, V = 640 x 480 pixels)**

Structure	m	x	y	Sensor elements	% elements	Fovea dimensions	Fovea pixels	% pixels
Shifted	2	80	60	48000	100,0	160 x 120	19200	100
adaptive 1	3	32	24	39936	83,2	192 x 144	27648	144
adaptive 2	3	30	23	47130	98,2	220 x 158	34760	181
adaptive 3	3	31	22	47814	99,6	206 x 172	35432	184
adaptive 4	3	30	22	50220	104,6	220 x 172	37840	197

It is worth to note that adaptive structures 1, 2 and 3 above, in spite of having a lower number of elements, which implies a lower data handling on the system, have foveas with higher size than the one of shifted structure with  $m = 2$ . Thus, wider regions of the FOV are covered with the highest resolution, eliminating the need to refoveate when target singularities and size require examination with fovea acuity. All structures cover the same FOV, but adaptive fovea enhances high resolution levels, looking for better performance on recognition tasks, at the cost of reducing acuity of coarse resolution levels, which are mainly used for object detection.

#### 4. Motion Detection with Adaptive Fovea Structures

Application of adaptive fovea sensors for motion detection is based on the temporal differences,  $D_N(x, y, t)$ , between corresponding elements of consecutive images of the  $I_N(x, y, i)$  input sequence :

$$D_N = | I_N(x, y, t+1) - I_N(x, y, t) | \quad (11)$$

Temporal differences are measured for all sensor cells,  $x$  and  $y$  being their coordinates within the resolution level  $N$  containing the cell. Differences are compared to preset thresholds and transformed to binary masks determining a rough estimate of the shape and position of moving objects. Differences on the coarsest resolution level, because of the cell averaging performed, can be considered as the result of applying a low-pass filter to the original images and, therefore, binary masks obtained at this level are used as bounding box delimiters, indicating dynamic regions within the static FOV.

From sequences as in Fig. 5, binary masks sets are obtained as shown in Fig. 6 and 7.



Fig. 5 Motion sequence.



Fig. 6. Binary masks set of first two images of sequence, applied to both images.



Fig. 7. Binary masks set of last two images of sequence, applied to last image.



Fig. 8. AND of Binary masks in Figs. 6 and 7, applied to second image.

Both masks are ANDed to form the second level binary mask. Application of this mask to the image in common to both first level masks, lets extraction of the moving objects on any resolution level, as seen in Fig. 8. Foveations to track the object of interest can be based on predictions obtained from bounding boxes of coarsest levels, which are continuously corrected after obtaining parameters  $l$ ,  $r$ ,  $t$ ,  $b$  as given in (4), to extract the segmented regions covering moving objects at the coarsest level.

## 5. Conclusions

We have described a new methodology to obtain adaptive multiresolution structures for visual sensors. Their main advantage is the possibility to adapt the size and position of their foveas to areas of interest, minimizing the overall data volume related to the resolution levels as well as the number of foveations required to examine a region and, therefore, the time involved on those tasks. Pixel averaging, to obtain rexels from the uniresolution camera data, as well as rexel or pixel additions and subtractions, are easily performed on FPGAs, applying their capability to perform basic arithmetic operations. Similarly, the In-circuit reconfiguration feature of these devices is well suited to select the areas of the FOV where resolution must be changed or brightness values of the image sequence subtracted for motion detection. Adaptive fovea structures allow coverage of wide FOVs, where several moving objects could appear but, at the same time, selected areas of image sequences can be observed in a closer detail, emulating saccadic eye movements for tracking tasks on low cost active vision systems. Mechanical camera orientations, as coarse saccadic head movements, are used to change the FOV whenever the targets get out of scenes being examined, but the stringent mechanical requirements associated to non-reconfigurable sensors are eliminated with the fast and low cost electronic foveation system.

## Acknowledgements

The present work has been partially supported by the Spanish Comisión Interministerial de Ciencia y Tecnología, ( CICYT ), Project No. TIC095 - 0589.

## References

- [1] J.Van der Spiegel, G.Kreider, C.Claeis, I.Debusschere, G.Sandini, F. Fantini, G.Soncini, "A foveated retina-like sensor using CCD technology" in *Analog VLSI implementation of Neural Systems*, C.Mead, M.Ismail, Eds., KluwerAcad. Publ, 1989.
- [2] M.Tistarelli, G.Sandini, " Estimation of depth from motion using anthropomorphic visual sensor", *Image and Vision Computing*, vol.8, No. 4, pp. 271- 278, 1990.
- [3] A.S.Rojer, E.L.Schwartz, "Design considerations for a space-variant visual sensor with complex logarithmic geometry", in *Proceedings of the 10th. International Conference on Pattern Recognition*, pp.278 - 285, 1990.
- [4] C.Bandera, P. Scott, "Foveal Machine Vision Systems" in *IEEE International Conf. on Sys., Man and Cybernetics*, Cambridge, MA, pp.596-599. November 1989.
- [5] E. Schwartz, D.N Greve, G.Bonmassar, "Space-variant Active Vision: Definition, Overview and Examples". *Neural Networks*, Vol.8, No. 7/8, pp.1297 - 1308, 1995.
- [6] P.Camacho, F.Arreola, F.Sandoval, "Shifted Fovea Multiresolution Geometries" in *Proceedings of the IEEE International Conference on Image Processing*, Vol.I pp.307- 310. Lausanne, Switzerland. 1996.