Efficient Retina-like Resampling from Cartesian Images

Hugo Vieira Neto, Diogo Rosa Kuiaski and Gustavo Benvenutti Borba
Graduate School of Electrical Engineering and Applied Computer Science
Federal University of Technology - Paraná, Brazil
hvieir@utfpr.edu.br, diogo.kuiaski@gmail.com, gustavobborba@gmail.com

Abstract

This paper describes a novel method to resample cartesian images to a retina-like hexagonal tessellation, consisting of foveal and peripheral regions. We introduce the inflated-hexagons model for the foveal region resampling, which has an efficient data structure for storage and mapping of cortical coordinates. The integral image representation is used as intermediate step for the peripheral region log-polar resampling, which is less computationally demanding than the simulation of Gaussian receptive fields with varying widths. An important feature of the resulting model is that a gap-free transition is obtained between the near-uniformly sampled fovea and the space-variant periphery. Experiments with Principal Component Analysis (PCA) compression of the ORL Database of Faces show that the proposed retina-like resampling, in spite of massively reducing the amount of image data, still retains most of the information contents of conventional cartesian image sampling.

1. Introduction

The human retina can be subdivided topologically into foveal and peripheral regions. The fovea is the region at the centre of the retina, where the highest visual acuity is obtained. Surrounding the fovea there is a peripheral region whose visual acuity is increasingly lower with eccentricity [9]. In biological vision, retinal photoreceptors are arranged in a nearly hexagonal tessellation [3], while in computer vision image representation is often made using an uniformly sampled 2D cartesian coordinate system.

The use of the log-polar mapping for retina-like peripheral resampling is well-known for its advantages in achieving rotation- and scale-invariant representations [4, 8]. A review of the methods available in the literature can be found in [3]. This paper presents a new method to resample cartesian images into a full retina-like representation that differs from known models, such as the ones presented in [2] and [5], in the sense that it has an efficient data structure for storage and also that there is no sampling gap in the transition between fovea and periphery.

The proposed model is completely described by simple equations and does not resort to more complex methods, such as self-organisation [1], in order to generate the tessellation. The resulting data structure allows accurate and fast mapping between retinal and cortical coordinates. Cartesian image reconstructions from the retina-like resampling and quantitative assessment using incremental PCA [11] compression of the ORL Database of Faces [10] demonstrate the model's ability to preserve visual information contents.

2. The retinal model

To remap a conventional cartesian image that uses a uniform square tessellation to a log-polar image representation that uses a space-variant hexagonal tessellation, a data structure consisting of two regions – the fovea and the periphery – is needed. This happens because in the foveal region there is a large amount of evenly distributed photoreceptors, while in the peripheral region the amount of photoreceptors decreases with eccentricity. So, the model needs distinct data structures for the fovea, which is a small area with high photoreceptor density, and for the peripheral region, which is a larger area but with sparser photoreceptor density.

2.1. Foveal region

In the foveal region of our model, resampling is made with a series of concentric rings of photoreceptors, as shown in figure 1. The first internal ring has the shape of an hexagon and contains one photoreceptor in each of its vertexes. The second ring, which surrounds the first one, mostly maintains the hexagonal shape, but contains an additional photoreceptor between vertexes, along each side of the hexagon. As the rings grow farther from the centre of the fovea, the number of additional photoreceptors increases between the vertexes of the hexagon, while the hexagonal

rings themselves start to lose more and more their original shape and tend to become more and more circular in shape. We have called this "the inflated-hexagons model of the fovea".

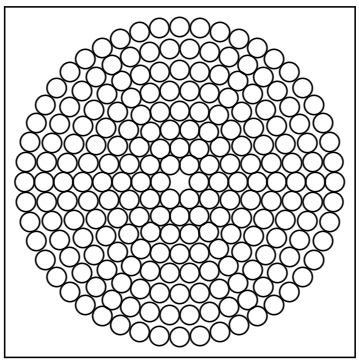


Figure 1. Spatial distribution of photoreceptors in the inflated-hexagons model of the fovea. Outer rings are more circular in shape, while inner rings are more hexagonal.

The high density of sampling in the foveal region leads to each photoreceptor having the same size of a cartesian pixel in every ring. The foveal region of the model is specified by the desired number of photoreceptor rings (M_f) given by the user. Ring radii increase in arithmetic progression and are given by $R_i = i + 1$, for $0 \leq i < M_f$. In each ring i there are $6 \times R_i$ photoreceptors.

The centre of each inflated-hexagons model photoreceptor is given in (x, y) cartesian coordinates, which are computed according to equation 1:

$$\begin{aligned} x &= x_a + (x_b - x_a)\,\tfrac{i}{M_f} \\ y &= y_a + (y_b - y_a)\,\tfrac{i}{M_f} \end{aligned}, \tag{1}$$

where (x_a, y_a) and (x_b, y_b) are given by

$$\begin{aligned} x_a &= x_c + R_i \left[\cos\theta_j + (\cos\theta_{j+1} - \cos\theta_j)\,\tfrac{k}{R_i}\right] \\ y_a &= y_c + R_i \left[\sin\theta_j + (\sin\theta_{j+1} - \sin\theta_j)\,\tfrac{k}{R_i}\right] \end{aligned} \tag{2}$$

and

$$\begin{aligned} x_b &= x_c + R_i \left\{\cos\left[\theta_j + (\theta_{j+1} - \theta_j)\,\tfrac{k}{R_i}\right]\right\} \\ y_b &= y_c + R_i \left\{\sin\left[\theta_j + (\theta_{j+1} - \theta_j)\,\tfrac{k}{R_i}\right]\right\} \end{aligned}, \tag{3}$$

for $0 \leq k < R_i$.

Equations 2 and 3 compute intermediate coordinate pairs for hexagonal (x_a, y_a) and circular (x_b, y_b) ring shapes, respectively, which are later composed into a single coordinate pair (x, y) by the weighted sums in equation 1. Equation 2 yields coordinates for a perfect hexagonal ring, while equation 3 yields coordinates for a perfect circular ring. When these coordinates are composed in equation 1, rings near to the centre of the fovea appear more like hexagons and rings far from the centre appear more like circles. In these equations, (x_c, y_c) represents the centre of the fovea, and θ_j and θ_{j+1} are angles between the x axis and lines connecting neighbour vertexes of the base hexagon to the centre, which are given by:

$$\theta_j = \frac{\pi}{3} j, \tag{4}$$

with $0 \leq j < 6$.

Cartesian visual data is remapped through bilinear interpolation and stored in memory as a 2D array – the cortical image – with the first column corresponding to the first photoreceptor ring, the next two columns corresponding to the second ring, the next three to the third and so forth, following an arithmetic progression, as indicated in figure 2. This topological arrangement also results in each row of the storage array representing wedges of $\pi/3$ rad of the foveal region.

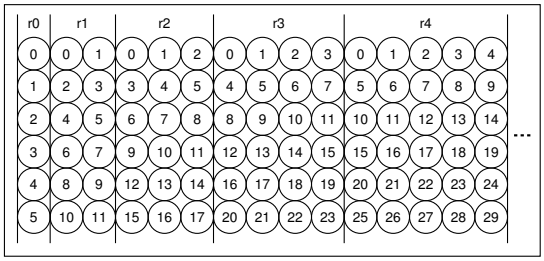


Figure 2. Corresponding cortical image for the inflated-hexagons model of the fovea. Column sections r_i represent photoreceptor rings increasingly far from the foveal centre and rows represent photoreceptor wedges of $\pi/3$ rad.

Coordinates in the foveal image are given in terms of ring and sector (r, s) and have corresponding coordinates (m, n) in the cortical image, where m is the cortical column coordinate and n is the cortical row coordinate. Retrieving visual data from the cortical array while searching the foveal image is straightforward. Equation 5 shows how foveal coordinates (i, s) can be computed from cortical coordinates (r, c):

$$\begin{matrix} i = \left\lceil \frac{\sqrt{8c+1}-1}{2} \right\rceil \\ s = i\left(r - \frac{1}{2}\right) - \frac{i^2}{2} + c \end{matrix}, \quad (5)$$

where $(\lceil . \rceil)$ denotes the ceiling function for rounding.

Equation 5 is derived from the arithmetic progression resulting from the cortical image arrangement, as shown in figure 2. The value of i is given by the largest value of c that satisfies the condition $i - \sum_{j=0}^{c} j \geq 0$. Therefore, solving Gauss's equation for the sum of the elements of an arithmetic progression yields the relationship between i and c.

Equation 6 shows how to compute the inverse mapping, obtaining (r, c) as a function of (i, s):

$$\begin{matrix} c = \frac{i^2-i}{2} + \mathrm{mod}\{s-1, i\} \\ r = \mathrm{mod}\{3(i^2-i) + \left\lfloor \frac{n-1}{m} \right\rfloor, 6\} \end{matrix}, \quad (6)$$

where $\mathrm{mod}\{a, b\}$ is the remainder of the integer division a/b and $\lfloor . \rfloor$ denotes the floor function for rounding.

2.2. Peripheral region

In the periphery, where a log-polar model is used, photoreceptor rings have a perfect circular shape and a constant number of photoreceptors per ring, but photoreceptor size starts to increase as a function of the distance to the foveal centre. The total size of the peripheral region is provided by the user in terms of the desired number of rings (M_p).

In order to keep coherence at the border between the fovea and the periphery, the number of sectors in the peripheral region is the same as in the last ring of the foveal region. Therefore, the number of sectors in the periphery is given by $N = 6 \times M_f$ and the half angular resolution can be computed as:

$$\theta = \frac{\pi}{N}. \quad (7)$$

The photoreceptor radius r_i is related to the ring radius R_i according to:

$$r_i = R_i \sin\theta. \quad (8)$$

As the minimum desired size for a photoreceptor is equivalent to one cartesian pixel ($r_0 = 1/2$), the innermost peripheral ring radius R_0 is given by:

$$R_0 = \frac{1}{2\sin\theta} \quad (9)$$

Ring radii increase in a geometric progression given by:

$$R_i = bR_{i-1}, \quad (10)$$

where b is computed as follows:

$$b = \frac{\sin\theta(\sin\theta + \sqrt{2\cos\theta + 1}) + \cos\theta}{(\cos\theta)^2}. \quad (11)$$

Finally, the cartesian coordinates of each photoreceptor centre is given by:

$$\begin{matrix} x_{i,j} = x_c + R_i \cos[\theta(i+2j)] \\ y_{i,j} = y_c + R_i \sin[\theta(i+2j)] \end{matrix}, \quad (12)$$

for $0 \leq i < M_p$ and $0 \leq j < N$.

Figure 3 shows the resulting space-variant tessellation with $M = 18$ rings and $N = 48$ sectors. Notice that the final organisation consists of spiral sectors and rings.

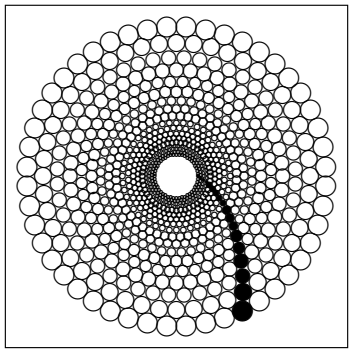


Figure 3. Peripheral log-polar hexagonal tessellation with $M = 18$ rings and $N = 48$ sectors. The first spiral sector is indicated by black circles.

One of the most important features of the proposed retina-like model is that there is no sampling gap between fovea and periphery, unlike other models available in the literature [5]. Notice that the centre of the image is lost if a purely log-polar structure is used, as shown in figure 3.

2.3. Efficient implementation

Photoreceptor values in the peripheral region can be approximated by the average of cartesian pixel values within the square-shaped areas shown in figure 4.

An efficient implementation to compute the peripheral log-polar mapping can be made using the integral image representation introduced in [12]. The integral image at location (x, y) contains the sum of the pixel values above and to the left of (x, y), as follows:

$$ii(x, y) = \sum_{m \leq x, n \leq y} i(m, n), \quad (13)$$

where $ii(x, y)$ is a pixel of the integral image and $i(m, n)$ is a pixel of the original image.

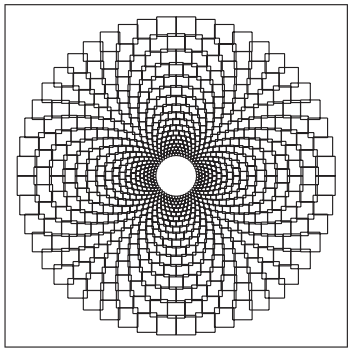


Figure 4. Square photoreceptor approximation for the peripheral log-polar tessellation in figure 3.

The integral image can be computed in a single pass over the original image with the following recurrences [12]:

$$\begin{aligned} s(x,y) &= s(x,y-1) + i(x,y) \\ ii(x,y) &= ii(x-1,y) + s(x,y) \end{aligned}, \tag{14}$$

where $s(x,-1) = 0$ and $ii(-1,y) = 0$.

The most relevant property of the integral image representation is that the sum of pixel values within a rectangle of any size in the original image can be computed in constant time using the values of the corresponding four corners in the integral image, as shown in figure 5.

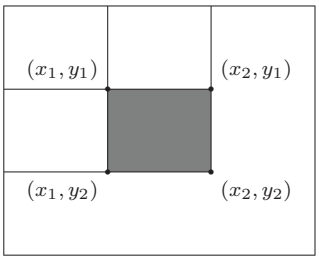


Figure 5. Integral image: the sum of the pixel values within the grey rectangle can be computed as $ii(x_1,y_1) - ii(x_2,y_1) - ii(x_1,y_2) + ii(x_2,y_2)$.

The sum of pixel values in the square areas shown in figure 4 can be very efficiently computed with the use of look-up tables containing their corner coordinates:

$$\begin{aligned} x_1 &= x_{i,j} - r_i \\ y_1 &= y_{i,j} - r_i \\ x_2 &= x_{i,j} + r_i \\ y_2 &= y_{i,j} + r_i \end{aligned} \tag{15}$$

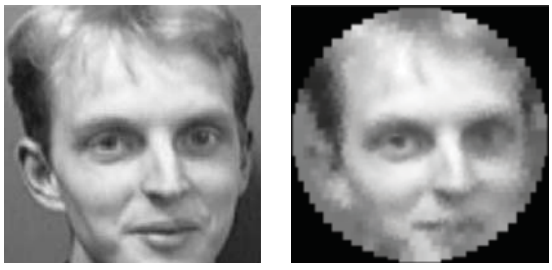
Therefore, average photoreceptor values are obtained by:

$$p = \frac{ii(x_1,y_1) - ii(x_2,y_1) - ii(x_1,y_2) + ii(x_2,y_2)}{(x_2 - x_1)(y_2 - y_1)} \tag{16}$$

3. Experiments

In our initial experiments, we have resampled cartesian images of 100×100 pixels from the ORL Database of Faces [10] with our retina-like model. Ten photoreceptor rings in the fovea and 16 in the periphery were used in the structure of the model, resulting in a total of 1290 photoreceptors. When compared to the original cartesian representation, this number of photoreceptors represents a reduction in storage space in excess of 87%.

Figure 6 shows the example of an original image and its respective retina-like resampled output reconstruction for an image from the ORL Database of Faces. The data reduction seems extremely lossy when a comparison is to be made between the original and reconstructed images. However, the reconstructed image obtained is still reasonably intelligible, demonstrating that the "essential information" of the original image was preserved. The main advantage of this fact is that image processing methods can be executed faster in the cortical map domain because the amount of data to be processed is reduced.



(a) (b)

Figure 6. ORL Database of Faces example: (a) original image; (b) retina-like resampled image reconstruction.

It is important to keep in mind that such retina-like representation is useful in applications that involve robotic vision, in which cameras are able to move with respect to the environment. In these applications, it is necessary to centre the high-resolution fovea of the model at the current object of interest that needs to be analysed. The low-resolution periphery in this kind of application is used mostly to shift the focus of attention of the system [7].

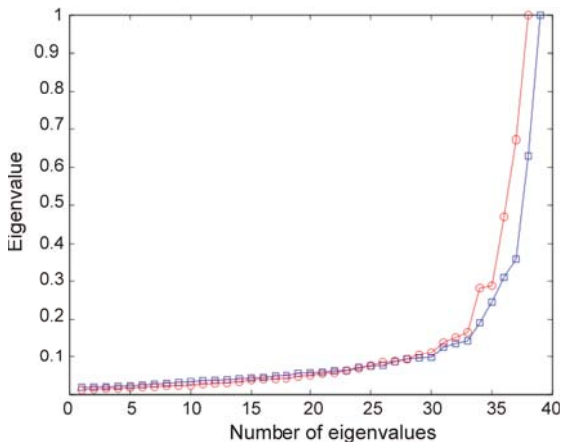


Figure 7. PCA of the ORL Database of Faces: (a) eigenvalue plot; (b) cumulative eigenvalue plot. The blue plots correspond to the cartesian image domain and the red plots correspond to the retinal image domain.

In order to verify if the main information contents was preserved after the retina-like resampling, we performed incremental PCA [11] on the complete ORL Database of Faces. PCA – also known as the Hotteling Transform or discrete Karhunen-Loeve Transform in the context of image processing – provides optimal image reconstruction in the least-squares sense [6]. Therefore, we refer here to "amount of information" or "main information" as the eigenvalues obtained through PCA.

In figure 7a, it is possible to visualise that the amount of information for the retina-like representation is similar to that of the cartesian representation, as the number of eigenvalues obtained is almost the same. Figure 7b shows a graph for the cumulative eigenvalue plot, where one can observe that information contents grows even faster for the retina-like mapping.

An additional experiment was conducted in order to demonstrate some of the topological properties of the proposed retina-like resampling. Figure 8 shows the cortical data structure and the reconstructed image in contrast with the original cartesian image. In this example, it is possible to notice that circular regions are remapped to horizontal lines in the cortical periphery and that corners with at least $\pi/3$ rad are also remapped to horizontal lines in the cortical fovea of the data structure.

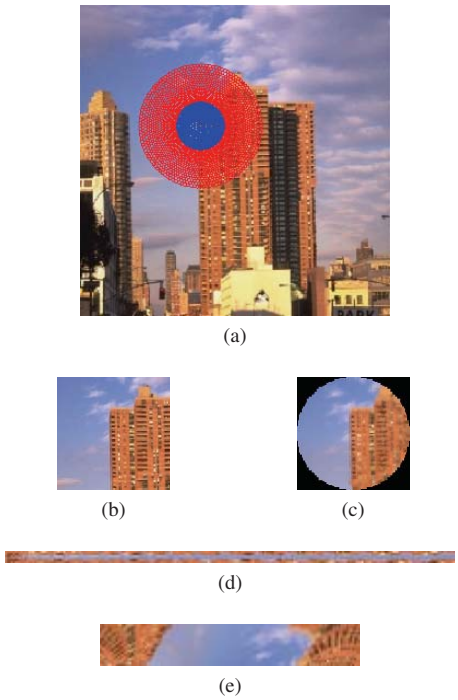


Figure 8. Topological properties: (a) retinal model superimposed to an image; (b) original image patch; (c) retina-like resampled image patch reconstruction; (d) cortical fovea; (e) cortical periphery.

4. Conclusions

We have presented a model for retina-like resampling of cartesian images composed of a high-resolution fovea and a space-variant log-polar periphery. Important features of the model are the gap-free transition between fovea and periphery, and the use of simple mathematical equations to define the mapping between cartesian and retina-like domains. The model provides a coherent cortical map representation for both foveal and peripheral region, which can use an efficient implementation based on the integral image representation.

Experiments using incremental PCA demonstrate that our model, despite massively reducing the amount of image data, still preserves most of the information contents of the ORL Database of Faces. This is also observed through visual comparison of original cartesian images and the images resampled using the model.

Future work includes investigation of specific characteristics of the model, in special its topological properties and behaviour in the frequency domain. It should be noted that our model has the advantage of presenting periodicity in one of the dimensions of both cortical images – foveal and peripheral – which presents potential advantages for computing its Fourier Transform.

We also intend to investigate the well-know properties of log-polar representations to achieve rotation- and scale-invariance for feature extraction in content-based image retrieval and robotic vision applications. The model seems to be particularly interesting to be used in conjunction with a visual attention mechanism [7], where the salient visual locations should be centred on the fovea.

References

[1] L. S. Balasuriya and J. P. Siebert. An artificial retina with a self-organised retinal receptive field tessellation. In *Proceedings of the AISB'03 Symposium on Biologically-inspired Machine Vision, Theory and Application*, pages 34–42, Aberystwyth, UK, 2003.

[2] A. Bernardino, J. Santos-Victor, and G. Sandini. Model-based attention fixation using log-polar images. In V. Cantoni, A. Petrosino, and M. Marinaro, editors, *Visual Attention Mechanisms*. Plenum Press, New York, NY, 2002.

[3] M. Bolduc and M. D. Levine. A review of biologically motivated space-variant data reduction models for robotic vision. *Computer Vision and Image Understanding*, 69(2):170–184, 1998.

[4] P. Cavanagh. Local log polar frequency analysis in the striate cortex as a basis for size and orientation invariance. In D. Rose and V. G. Dobson, editors, *Models of the Visual Cortex*, pages 85–95. Wiley, New York, NY, 1985.

[5] H. M. Gomes and R. B. Fisher. Learning-based versus model-based log-polar feature extraction operators: A comparative study. In *Proceedings of the Sixteenth Brazilian Symposium on Computer Graphics and Image Processing*, pages 299–306, São Carlos, Brazil, 2003.

[6] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Pearson, 3rd edition, 2008.

[7] L. Itti and C. Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40(10-12):1489–1506, 2000.

[8] T. Kurita, K. Hotta, and T. Mishima. Scale and rotation invariant recognition method using higher-order local autocorrelation features of log-polar images. In *Proceedings of the 3rd Asian Conference on Computer Vision*, volume 2, pages 89–96, Hong Kong, China, 1998.

[9] S. E. Palmer. *Vision Science: Photons to Phenomenology*. MIT Press, Cambridge, MA, 1999.

[10] F. Saramaia and A. Harter. Parameterisation of a stochastic model for human face identification. In *Proceedings of the Second IEEE Workshop on Applications of Computer Vision*, pages 138–142, Sarasota, FL, 1994.

[11] H. Vieira Neto and U. Nehmzow. Incremental PCA: An alternative approach for novelty detection. In *Proceedings of TAROS 2005: Towards Autonomous Robotic Systems*, pages 227–233, London, UK, 2005.

[12] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 511–518, Kauai, Hawaii, 2001.