9th IEEE-RAS International Conference on Humanoid Robots
December 7-10, 2009 Paris, France

Smooth Foveal Vision with Gaussian Receptive Fields

Daniela Pamplona and Alexandre Bernardino

Abstract— **Despite the huge amount of information, the human brain is able to perceive and interpret visual signals in real time. One of reasons is that visual information is selectively sampled in the retina providing higher acuity in the center (where usually the most important information is) than in periphery. Humanoids vision can benefit from such space-variant representations of the visual field with utility not only in image data reduction but also in others applications as vergence, active tracking, as demonstrated in the last decades' research. However, classical methods model foveation processes with non-smooth receptive fields with are a weak match to the human physiology. Instead we propose an alternative representation using Gaussian kernels. While increasing redundancy, Gaussian receptive fields provide a smoother representation of Foveal images and model certain properties of data acquisition in human vision. In addition, we propose an algebraic approach for the analysis, synthesis and processing of Foveal images, using simple matrix computations and operator theory. We show how to derive the equivalent Foveal operators to common Cartesian domain linear processing routines such as image geometrical transformations and filtering operations. We present experiments illustrating the performance of the proposed methodology in comparison to classical approaches for space-variant image processing both in image reconstruction and in motion estimation/tracking tasks.**

I. Introduction

Visual perception in biological systems is often characterized by space-variant acquisition and processing mechanisms, which reduce the amount of instantaneous information to process. The retina of mammals have a three layered structure: the first layer is composed by the photoreceptors: rods and cones[1]. Cones are distributed non-uniformly as function of the visual position, according to a log-polar law [21]. They have a high concentration in the *fovea*[2] and a decreasing density toward the periphery of the visual field. The second layer is called inner nuclear area and transmits signals form the first to the last layer. The last layer is composed by ganglion cells that pool together the information of several photoreceptors. The distribution of these cells also follow a log-polar law but with sharper decrease of density than photoreceptors [14], [28], [27].

This complex system is the biological solution for dealing with information from more than 1 million photoreceptors. Foveation, the process of transform a regular sampled image into a log-polar geometry, decreases exponentially the number of computations, keeping the acuity in the center and a large field of view.

All authors are with the Institute for Systems and Robotics, Instituto Superior Técnico, Av.Rovisco Pais, 1, 1049-001 Lisboa, Portugal {dpamplona,alex}@isr.ist.utl.pt

[1]Cones dedicated to daylight color processing, while rods are for low light vision and fast motion

[2]The central part of the retina, covering about 1 deg of visual angle.

Several authors have proposed methods to acquire and process log-polar images and have developed applications in multiple fields (check [6] and [30] for recent reviews). Particularly in humanoid robotics, where robots have moving eyes and must operate in real-time mimicking some aspects of human behaviour, several applications of interest have exploited the properties of log-polar images, namely depth perception and vergence control [5], [7], image motion computation and tracking [2], [24], ego-motion estimation [22], [8], visual attention [17], [1] and integrated binocular head control [4], [3], [19]. However, most of the existing methods do not appropriately model the information acquisition properties of biological computational elements, thus not fully exploiting the analogy the natural vision systems. In this paper we try to revisit existing foveal vision methods and go one step further in matching data acquisition properties of biological systems.

Classical approaches to foveation focus on modeling the distribution of receptive fields but do not exploit properly their shapes. Usually, a uniform resolution image (from now on denoted Cartesian), is subdivided in compact regions of with positions and sizes following a log-polar law, called *superpixels*. Then, the pixels in the original image belonging to each of those regions are "averaged" and the results stored in memory. To address image processing operations in space-variant images, one of the most formal methods proposed to date was introduced in [26], that performs image processing and geometrical transformations using graph based operations.

Our approach, instead, considers ganglion cells' receptive fields as the basic units of image analysis, with a closer resemblance to its biological counterparts. Images are generated by sampling the information on the modeled retina with receptive fields of Gaussian shapes, whose locations and sizes can be determined in an application dependent manner. In this paper the locations are chosen to match the distribution of these cells in the human retina (which is approximately log-polar) and sizes are chosen in a way to avoid image aliasing effects. Fig 1 shows schematically the spatial support of receptive fields in the superpixel approach and in our proposal. Notice that the significant amount of overlap between receptive fields in our approach and their shape smoothness are required to reduce image aliasing.

Other works have proposed overlapping receptive fields models to follow more closely biological data. The models in [29] and [20] have RF's with a log-polar distribution but with circular shapes and a moderate amount of overlap. [29] proposes a tessellation with a linear relation between receptive field size *vs* eccentricity, and a receptive field overlap of

50%. The proposal in [20] also uses circular receptive fields but tries to minimize the amount of overlap between them. However, these receptive fields have a cylindrical shape (thus sharp boundaries) and reduced overlap factors, therefore not providing enough low-pass filtering to avoid aliasing effects. By using Gaussian receptive fields, we provide the largest amount of low-pass filtering for the smallest spatial support (modulated Gaussians have a minimum joint time-frequency localization [10]), thus making Gaussian receptive fields ideal in terms of minimizing computations for the same bandwidth.

To model the analysis, synthesis and image processing operations in the foveal domain we propose the use of matrix operations. In contrast to existing approaches, that use custom or graph based operations, we model cartesian-to-foveal image acquisition with matrix operations, which allow us to represent the reconstruction process (foveal-to-cartesian) using simple pseudo-inverse methods.

We present a procedure to transform linear (cartesian) maps in the foveal domain (foveal-to-foveal). Using operator theory in Hilbert spaces, we derive an algebraic method that models the filtering as matrix-vector product, reducing the number of computations in account with the space-variant geometry. We illustrate the application of such methods to image geometrical transformations and filtering operations, and at a higher level, in the tracking task.

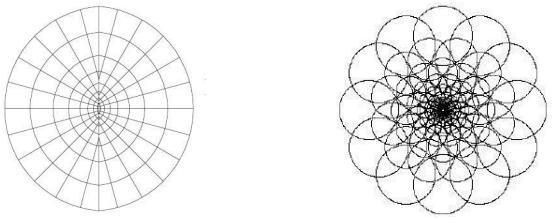


Fig. 1. Human vision models that inspire the foveation process. Models based of *superpixels* (left) and *receptive fields* (right).

The paper is organized as follows. Section II describes our model of foveation using Gaussian Receptive fields. In section III we formulate the analysis and synthesis problems with matrix algebra operations. Then, in section IV, image processing operations in the Foveal domain are defined and also formulated as matrix computations. Using adjoint operator theory in Hilbert spaces we propose methods to represent in the Foveal domain common linear Cartesian processing operations. Results are presented in section V, illustrating the performance of our approach with respect to the classical methods. Finally, section VI presents the conclusion of our work and directions for future research.

II. Receptive Field Foveation

Usual approaches to foveation are inspired on the distribution and shape of the photoreceptors on the retina of mammals. However, most often such approaches do not consider that the cortical image is not simply a reflex of the image sensed on retina, but rather is a representation of the image information in the form of a sample hierarchy [31]. The ganglion cells are the neuron cells responsible to receive visual information from photoreceptors and send it to the visual system of the brain [9]. The term receptive field comes after neurophysiology experiments demonstrating that visual cells are responsive only to stimulus in a confined region of the visual field. Also, not all parts of the RF contribute equally to the cell response, thus leading to the concept of RF *profile*. Profile functions have a limited spatial support (the region where a stimulus elicits cell response) and a well-defined center or location (where a stimulus elicits the maximal RF response). The information computed by each ganglion cell can be modeled as the receptive field response to the Cartesian image:

$$f[k] = \langle \phi_k, c \rangle = \sum_{i=1,j=1}^{M,N} \phi_k[i,j] c[i,j] \tag{1}$$

where c is the Cartesian image, $[i,j]$ are the coordinates and ϕ_k is the profile function.

We model each profile as a Gaussian function ϕ_k where the mean μ_k is the center of the receptive field and the standard deviation σ_k defines its support.

$$\phi_k[i,j] = g_{\mu_k,\sigma_k}[i,j] = \frac{1}{\sigma_k\sqrt{2\pi}} e^{-\frac{||(i,j)-(\mu_{k_i},\mu_{k_j})||^2}{2\sigma_k^2}} \tag{2}$$

Taking into account the distribution and scale of receptive fields in the mammal visual system [14], [15], $\{\mu_k\}$ should have a log-polar distribution and $\{\sigma_k\}$ should increase linearly with the distance to the center (as in Fig. 1). Moreover, due to the discretization of the Gaussians and the image boundary effect, the kernels should be normalized so that $\sum_{i,j} \phi_k[i,j] = 1$. The left hand side of Fig. 3 shows the side view of a single Gaussian receptive field.

As a way to illustrate the advantage of smooth receptive fields, we show on Fig. 2 the reconstructions of images performed with the classical and the proposed approach. We can verify that our approach produces smooth transitions between pixels and performs a better reconstruction in the periphery. In section V, we quantitatively evaluate both approaches on a large image data base.

(a) Cartesian image (b) Super pixel (c) Gaussian receptive fields

Fig. 2. Original image (left) and the reconstructed images using the usual super pixel approach (middle) and the Gaussian receptive fields approach (right).

III. Analysis and Synthesis

In this section we present methods for the foveation (analysis) and reconstruction (synthesis) of the Cartesian

image. The analysis process consists of representing the Cartesian image with a code obtained from sampling it with Gaussian receptive fields. This code can be represented as a simple array or, if sampling points are a 2D transformation of the uniform grid, as an image. The synthesis process consists of reconstructing the Cartesian image from the Foveal code.

We model the problem considering both the Cartesian image and the Foveated image belonging to Hilbert spaces. Foveation is defined as an operator between these spaces. This allows to represent both the analysis and synthesis processes as matrix operations, not only simplifying the calculus, but also allowing the exploitation of the algebraic properties of Hilbert spaces and operator theory to address complex image representation and processing problems. For, instance, in this setting, the analysis process will be defined as a simple matrix multiplication, and the reconstruction process defined as its Moore-Penrose pseudo-inverse.

A. Analysis

For fixed M, N, the space C, of the usual Cartesian images with values in $\mathbb{R}^{M,N}$, is associated with the usual operations and the norm induced by the inner product of two images c_1 and c_2:

$$c_1 \cdot c_2 = \sum_{i=1,j=1}^{M,N} c_1[i,j]c_2[i,j] \tag{3}$$

It is thus considered as a discrete Hilbert space. Analogously, the space F of the Foveated images with fixed size $K < M \times N$ is also a Hilbert space. Moreover, the foveation process is an operator that maps C on F. Let the operator Fov be defined in a matrix form as:

$$\begin{aligned} \text{Fov} : C &\rightarrow F \\ c &\rightarrow \Phi c \end{aligned}$$

where Φ is a matrix of which each row k contains the values of a ganglion cell profile function ϕ_k and c is the Cartesian image (the profile functions and Cartesian images are reshaped in order to become a single vector). This matrix has size $(K, M \times N)$, which is the number of Foveated pixels *vs* the number of Cartesian pixels. Although Φ can be a large number, it is centrosymmetric [18] and very sparse, thus its space in memory can be reduced.

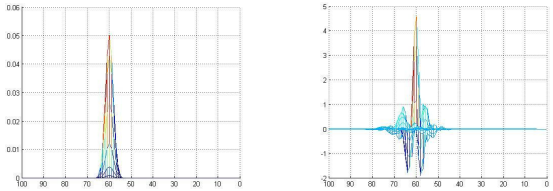
B. Synthesis

Foveation provides an incomplete representation of the image, thus Fov is injective but not surjective (there is not an inverse operator). However, Fov is right invertible, meaning that there is a Fov^{-1} such that $f = \text{Fov}(\text{Fov}^{-1}(f))$.
We define:

$$\begin{aligned} \text{Fov}^{-1} : F &\rightarrow C \\ f &\rightarrow \Phi^{+} f \end{aligned}$$

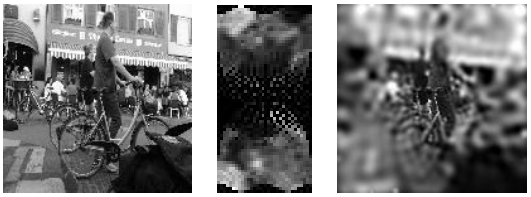
where Φ^{+} is the Moore-Penrose pseudo-inverse [23] of Φ.

If $f = \text{Fov}(c)$, then $\tilde{c} = \text{Fov}^{-1}(f)$ is the best solution in the least squares sense, *i.e.*, $||\tilde{c} - c||^2$ is the minimal solution for the inversion problem. Moreover, since the rows of Φ are linearly independent, we have $\Phi^{+} = \Phi^{T}(\Phi\Phi^{T})^{-1}$. Though we are not concerned in this paper with computational complexity issues, the number of computations in the pseudo-inverse can be reduced using the fact that Φ is centrosymmetric [18]. Moreover, there are iterative methods for obtaining approximations to the pseudo-inverse that trade-off computation time by approximation quality [23].



(a) Gaussian receptive field (b) Reconstruction function

Fig. 3. Shape of the analysis and synthesis functions, corresponding to rows of Φ and Φ^{+}, respectively.



(a) (b) (c)

Fig. 4. Cartesian (a), Foveated (b) and reconstructed (c) images

The rows of Φ^{+} represent the reconstruction functions. In Fig. 3 we can see the shape of the typical analysis and synthesis functions. Note that, despite the shape of the analysis functions is very regular (Gaussians), the shape of the synthesis functions is not. Notwithstanding they have a limited support which can be exploited to implement synthesis methods faster than raw matrix multiplication.

Fig. 4 illustrates the analysis and synthesis processes, showing the Cartesian image, the Foveal code and its reconstruction. The Foveal image is displayed in log-polar coordinates (logarithm of the distance to the center *vs* angle) where, due to boundary effects and oversampling in the center, part of the space is not represented (black pixels). Note that there is a faithful reconstruction of the center due to the high sampling density. In the periphery, where profile functions have larger support, the reconstruction has less detail, but due to the smooth shape of Gaussian receptive fields, the reconstruction does not present strong discretization artifacts. In section V we evaluate the average relative error as a function of the radial distance in a large image set.

Depending on the position and size of receptive fields, the matrix Φ can be ill posed. To avoid these problems, one can simply increase the value of σ_k for each receptive field, or, use Tikhonov regularization [13]. This, however, increases the approximation error and should be used only when Φ has a very large condition number. In our experiments we did not require regularization but retinas with other receptive

fields distribution and shapes may benefit from it.

IV. Operations on the Foveal Domain

One of the major difficulties about working with Foveal images is to apply the usual Cartesian operations in the Foveal domain, *i.e.*, without explicitly reconstructing the Foveal image to the Cartesian domain. This happens because of the peculiar shape of Foveated pixels and the complex pixel neighborhood relationships, that make cumbersome even the computation of a simple image translation.

A classical model for image processing in space variant domains is given by the Connectivity Graph of [26]. In that approach a graph-like structure is used to represent neighborhood relationships in Foveal images. Then, graph transformations and associated pixel computations are defined to implement the desired operations (image translations, edge detection, etc.). However, the definition of such graphs is a hard to program and error prone process. Furthermore, it lacks a theoretical support to analyze certain operations like composition of transformations or inversions.

In this section we present a method to derive the equivalent operations in Foveal images corresponding to common Cartesian image processing operations. We consider linear and bounded operations that map C on itself (endomorphic operators). We will present a procedure to define the equivalent operators on the Foveated space and represent them as matrix multiplications.

If P_C is an operator on the Cartesian domain, we want to define P_F that, when applied to f, simulates $\mathrm{Fov}(P_C(c))$. When P_C is linear and bounded, there exists P_C^*, the adjoint operator of P_C [11] such that, for each Foveated pixel k:

$$\begin{aligned}\mathrm{Fov}(P_C(c))[k] &= \langle \phi_k, P_C(c)\rangle \\ &= \langle P_C^*(\phi_k), c\rangle \\ &\approx \langle P_C^*(\phi_k), \mathrm{Fov}^{-1}(f)\rangle \end{aligned} \tag{4}$$

Since P_F domain is Foveal, we define

$$\begin{aligned} P_F : & F \to F \\ & f \to P_C^*(\Phi)\,\mathrm{Fov}^{-1}(f)\end{aligned}$$

where ϕ_k is the receptive field k and $P_C^*(\Phi)$ is a matrix whose rows are $P_C^*(\phi_k)$. Given that Fov^{-1} is represented by matrix Φ^+, then P_F is a matrix of size $K \times K$, defined by $P_C^*(\Phi)\Phi^+$. Note that, for fixed receptive fields geometry, P_F can be computed offline, meaning that the number of online computations is $O(K^2)$.

The simplicity of this approach allows to easily represent chains of operations by using matrix multiplications. Whereas in classical techniques (e.g. [26]) the definition of a chain of operations requires the construction of a new custom graph, in the proposed approach the combined operation is obtained by a simple matrix multiplication.

In the next section, we give examples of the transposition of Cartesian image processing operations to the Foveal domain: geometrical transformations and image filtering.

A. Geometrical Transformations

Without loss of generality we illustrate the case of translation. Other geometrical transformations are similar. Let the translation $(\Delta i, \Delta j)$ be given by the operator[3]:

$$\begin{aligned} T_C : & C \to C \\ & c[i,j] \to c[i-\Delta i, j-\Delta j]\end{aligned}$$

Expanding the inner product, we have:

$$\begin{aligned}\langle \phi_k, T_C(c)\rangle &= \sum_{i,j} \phi_k[i,j] T_C(c[i,j]) \\ &= \sum_{i,j} \phi_k[i,j] c[i-\Delta i, j-\Delta j] \\ &\text{take } i' = i-\Delta i, j' = j-\Delta j \\ &= \sum_{i',j'} \phi_k[i'+\Delta i, j'+\Delta j] c[i',j'] \\ &= \langle T_C^{-1}(\phi_k), c\rangle \end{aligned} \tag{5}$$

where T_C^{-1} is the operator associated with the translation $(-\Delta i, -\Delta j)$. Therefore, for this operation, the adjoint operator of T_C is the inverse translation T_C^{-1}. In practice, this consists in inverse transforming the receptive field positions and shapes and then apply to the untranslated image.

Finally, we define the translation on the Foveal domain for each pixel k as:

$$T_F(f(k)) = \langle T_C^{-1}(\phi_k), \mathrm{Fov}^{-1} f\rangle \tag{6}$$

meaning that, for the Foveated image, we have that

$$T_F(f) = T_C^{-1}(\Phi)\,\mathrm{Fov}^{-1}(f)$$

where $T_C^{-1}(\Phi)$ is a matrix where each row k is given by $T_C^{-1}(\phi_k)$.

Analogously, for any invertible coordinate transform of the Cartesian plane S_C, the corresponding S_F on the Foveal domain is given by

$$S_F(f) = S_C^{-1}(\Phi)\,\mathrm{Fov}^{-1}(f) \tag{7}$$

Fig. 5 is an example of a vertical translation on the Foveal domain. In this case, the number of computations online depends on 1284 pixels instead of the 10000 corresponding Cartesian.

B. Image Filtering

Image filtering operations can usually be performed with the aid of *masks* (FIR filters). Here we illustrate the derivation of Foveal filters analogous to given Cartesian *masks*. Let us take the case of the Laplacian filter. The Laplacian is a 2-D isotropic measure of the 2nd spatial derivative of an image. Many procedures, such as edge detection, use this filter as an intermediate step. There are various masks to approximate the discrete Laplacian kernel [12]. Let us consider the most usual one:

$$d = \frac{1}{4}\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \tag{8}$$

[3]While dealing with images, we assume that the image was not defined beyond its given limits

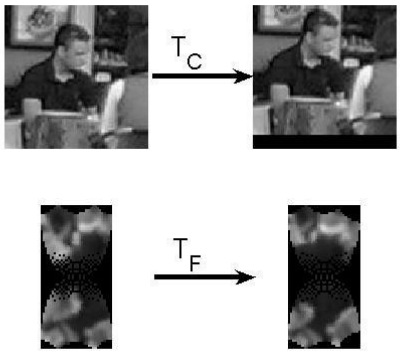


Fig. 5. On the upper row the initial Cartesian image and its vertical translation of 10 pixels, on the lower row the correspondent Foveal images.

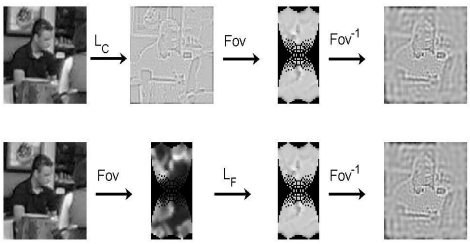


Fig. 6. Illustration of the Laplacian operator applied to Foveal images (bottom) in comparison to *ground truth* (top). On the top row a Cartesian image is filtered with a Laplacian mask and after, converted to the Foveal domain. On the bottom, the Foveal image corresponding to the original Cartesian in processed with the Foveal Laplacian operator. As can be seen in the images of the last columns, both computations give similar results.

The second derivative operator can be computed by the convolution of the image with this kernel.

$$\begin{aligned} L_C : C &\to C \\ c &\to d * c \end{aligned}$$

The symbol $*$ denotes the convolution operation.

Applying the procedure above, the corresponding Foveal operator is defined by:

$$\begin{aligned} L_F : F &\to F \\ f &\to (d \star \Phi)\,\mathrm{Fov}^{-1}(f) \end{aligned}$$

where $(d \star \Phi)$ is a matrix whose row k is the correlation between d and ϕ_k (the proof that correlation is the adjoint operator of convolution follows in an analogous manner to the translation case).

Fig. 6 compares the Laplacian obtained using the proposed approach with a *ground truth* provided by computing the Laplacian in the Cartesian domain and mapping the result to the Foveal space.

Since any linear filtering operation can be implemented by convolution, this formulation can be extended to any linear and bounded operator, like Laplacian of Gaussian, Gabor filters, etc.

Method	Aerials	Misce	Seqs	Texts	Total
Super Pixel	388	725	442	1843	907
Gaussian RF	354	594	325	1774	816

TABLE I

ERROR AVERAGE OF THE IMAGE RECONSTRUCTION USING MSQ

Method	Aerials	Misce	Seqs	Texts	Total
Super Pixel	287	1030	149	1833	1269
Gaussian RF	260	965	137	1835	1260

TABLE II

ERROR STANDARD DEVIATION OF THE IMAGE RECONSTRUCTION USING MSQ

V. RESULTS

In this section we show some quantitative results comparing our approach with the classical super pixel approach [26]. We consider the following problems: image reconstruction of Foveal images and tracking in a video sequence. In all tests the number of log-polar pixels and their distribution are the same. We generate the super pixel tessellation first, and then, with their centroid position ξ and area η, we define the center of the receptive field at ξ and the standard deviation of the Gaussian was set by $\sqrt{\eta}$. These choices were taken to have the same number and distribution of pixels in both approaches and perform a fair enough comparison between the two methods.

A. Reconstruction

If SP is the super pixel method for foveation and SP^{-1} its reconstruction, we compare both methods using the Mean Square Error:

$$\epsilon_{SP} = \frac{1}{MN}\sum_{i,j=1}^{M,N}[c[i,j] - \mathrm{SP}^{-1}(\mathrm{SP}(c[i,j]))]^2 \qquad (9)$$

$$\epsilon_{GF} = \frac{1}{MN}\sum_{i,j=1}^{M,N}[c[i,j] - \mathrm{Fov}^{-1}(\mathrm{Fov}(c[i,j]))]^2 \qquad (10)$$

over the "USC-SIPI Image Database" [25]. On this database, there are 4 types of images: Aerials (38 images), Miscellaneous (44 images), Sequences (70 images in 4 sequences) and Textures (154 images). The Cartesian images were normalized to size 100×100, and the Foveal images have 1284 pixels.

Fig. 2 shows an example of an image on the data base, its foveation and reconstruction using both the super pixel method and the Gaussian Foveation method.

Both the mean square error and standard deviation on reconstruction are consistently smaller in all image data sets

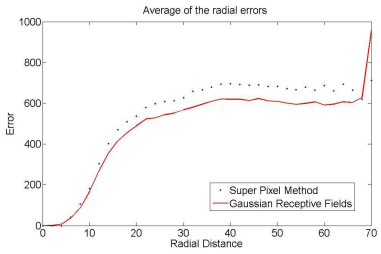


Fig. 7. Average of the radial errors using the MSE

using our approach (see Tables I and II). Moreover, a deeper analysis of the radial squared error, *i.e.*, the sum of the errors for all pixels with radius r, shows that our approach performs better than the usual one at all image eccentricities (see Fig. 7). Once the variance of the foveation with Gaussian receptive fields error is less then the variance of the super pixel approach error we can conclude that out approach is less subject to the particular characteristics of the input images.

We verify that, over the same conditions, our approach performs better that the usual one. As Fig. 2 illustrates, reconstruction using our method provides a better interpolation and smoothness in the transition between receptive fields. This provides not only less error in terms of the defined metrics but also better and more pleasant visualization since the discretization effect almost disappears.

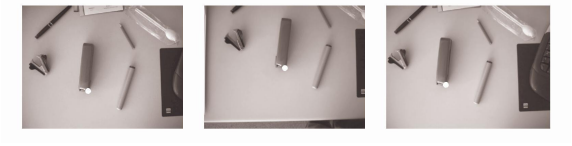
B. Tracking

The tracking problem consists in discovering and compensating the motion of one object in a video during time. There are numerous methods to solve this problem. We have applied template matching. This method is very simple: the main idea is to search for the motion that produces minimal mismatch between the image and a translation of a template in a discrete set of hypothesis.

We consider two tracking approaches for this problem: passive tracking (considering that the camera is fixed) or active tracking (considering that the camera dynamically changes its parameters to track the object). The latter type of search makes more sense in a biological paradigm because the eyes of humans track objects when they move in the environment, trying to center their projections on the highest resolution area (fovea).

Using two video sequences, we compared Gaussian Foveation with the Super Pixel approach, on passive (Fig. 8) and active (Fig. 10) tracking experiments, with several target velocities[4] (Fig. 9). The increase of velocity allows us to test the robustness of the tracking method on increasingly longer displacements. On both approaches, the grid of possible motions had resolution $\Delta h = 2$ and $\Delta v = 2$ pixels (horizontal and vertical, respectively), on a set of 7×7 hypothesis, and the number of Foveated pixels was 1284. The ground truth

[4]on velocity one we analyzed all frames of the video sequence, on velocity two we analyzed every second frame, on velocity three we analyzed all third frame, and so on.



(a) Frame 15 (b) Frame 150 (c) Frame 450

Fig. 8. Passive tracking using Gaussian Foveation on film A. The white mark is the estimated position of the tracking point.

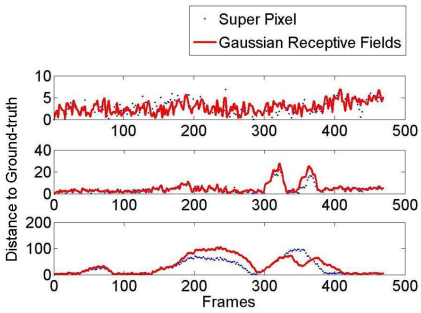


Fig. 9. Distances to ground truth on film A with velocity one, two, and three, using active tracking.

was calculated frame-by-frame by manually selecting a fixed landmark on the object.

In most of the cases, tracking results were very similar, meaning that, for velocity one, both approaches were able to follow the focus point. The approaches failed at velocity 3,(see Fig. 9) both in film A and B. However, there was one case (see Fig. 10) where the Super Pixel approach was not able to follow the tracking point and the Gaussian Foveation was. This does not demonstrate a general behavior but indicates a potential higher robustness of the Gaussian Foveation method.

(a) Frame 15 (b) Frame 250

(c) Frame 260

Fig. 10. Active tracking using Super Pixel on film B. The white mark is the estimated position of the tracking point

VI. Conclusion

Space-variant vision and eye movements are fundamental characteristics of the human visual system. As humanoid robots try to replicate human behavior and operate in real-time on unconstrained environments, they may benefit from artificial implementations of foveal vision. This paper presented a novel formulation for space-variant image synthesis, reconstruction, and processing for artificial vision systems that tries to mimic two important aspects of the human retina: the distribution and the shape of retinal ganglion cells. In the contrary to classical approaches that rely on graph based approaches and non-smooth receptive fields, we adopt an algebraic methodology that deals conveniently with smooth and overlapping receptive fields. Such type of receptive fields model more closely biological data acquisition systems but have large overlap with many neighbors, which we have found to be more elegantly represented using algebraic methods than graph based methods. The foveation process is modeled as a matrix multiplication, allowing the representation of the reconstruction process as a pseudo-inverse problem. Furthermore, using adjoint operators in Hilbert spaces we were able to derive the Foveal equivalent operators to common Cartesian image processing functions: geometrical transformation and filtering. Results show the advantage of the proposed approach with respect to graph based methods in terms of the quality of image reconstruction and robustness in tracking applications. On the negative side, due to the large overlap between receptive fields and extended neighborhoods, the proposed approach is computationally more demanding than usual methods. Notwithstanding, our foveation operations are easily parallelizable and can be implemented in accelerated hardware (GPUs, FPGAs).

In future work we intend to further characterize the properties of the proposed formulation. On the theoretical side we aim at investigating the use of Frame Theory [16] to derive error bounds for the foveation operations with respect to their Cartesian counterparts. On the practical side we aim at evaluating the benefits of the smooth foveal representation for the purposes of pattern categorization.

Acknowledgments

This work was supported by the European Commission, Project IST-004370 RobotCub, and by the Portuguese Government - Fundação para a Ciência e Tecnologia (ISR/IST pluriannual funding) through the POS Conhecimento Program that includes FEDER funds, and through project BIOLOOK, PTDC / EEA-ACR / 71032 / 2006. The authors would like to thank Dr. Carlos Alves and Dr. Cornelius Weber for the valuable suggestions and comments.

References

[1] I. Ahrns and H. Neumann, "Space-variant dynamic neural fields for visual attention," in *CVPR*, Fort Collins, Colorado, June 1999.

[2] A. Bernardino, J. Santos-Victor, and G. Sandini, "Foveated active tracking with redundant 2D motion parameters," *Robotics and Autonomous Systems*, vol. 39, no. 3–4, pp. 205–221, June 2002.

[3] A. Bernardino and J. Santos-Victor, "Visual behaviors for binocular tracking," *Robotics and Autonomous Systems*, vol. 25, pp. 137–146, 1998.

[4] ——, "Binocular visual tracking: Integration of perception and control," *IEEE Trans. on Robotics and Automation*, vol. 15, no. 6, pp. 137–146, December 1999.

[5] ——, "A binocular stereo algorithm for log–polar foveated systems," in *Proc. of the 2nd Workshop on Biologically Motivated Computer Vision*, November 2002, pp. 127–136.

[6] F. Berton, G. Sandini, and G. Metta, "Anthropomorphic visual sensors," in *Encyclopedia of Sensors*, M. P. C.A. Grimes, E.C. Dickey, Ed. American Scientific Publishers, 2005, vol. X, pp. 1–16.

[7] C. Capurro, F. Panerai, and G. Sandini, "Dynamic vergence using log-polar images," *IJCV*, vol. 24, no. 1, pp. 79–94, August 1997.

[8] K. Daniilidis, "Attentive visual motion processing: computations in the log-polar plane," *Computing*, vol. 11, pp. 1–20, 1995.

[9] S. Edelman, "Receptive fields for vision: from hyperacuity to object recognition," Weizmann Institute CS-TR 95-29, Tech. Rep., 1995.

[10] D. Gabor, "Theory of communication," *J. IEE*, vol. 93, pp. 429–459, 1946.

[11] T. Gohberg and S. Golberg, *Basic Operator Theory*. Birkhï¿½user, 1981.

[12] R. Gonzales and R. Woods, *Digital Image Processing*. Prentice Hall, 2002.

[13] P. C. Hansen, "The use of the l-curve in the regularization of discrete ill-posed problems," *SIAM Journal on Scientific Computing*, vol. Volume 14, no. Issue 6, pp. 1487–1503, November 1993.

[14] H. Kolb, "How the retina works," *American Scientist*, vol. 91, no. 1, January- February 1993.

[15] T. Linderberg and L. Florack, "Foveal scale space and the linear increase of receptive field size as a function of eccentricity," Royal Institute of Technology, Sweden, Tech. Rep. ISRN KTH NA/P 94/27 SE, 1994.

[16] S. Mallat, *A Wavelet Tour of Signal Processing, 2nd Ed.* Academic Press, 1999.

[17] G. Metta, "An attentional system for a humanoid robot exploiting space variant vision," in *Intl. Conf. on Humanoid Robots*, November 2001, pp. 22–24.

[18] W. Pye, T.Boullion, and T. Atchison, "The pseudoinverse of a centrosymmetric matrix," *Linear Algebra and Its Applications*, vol. 6, 1973.

[19] E. Rivlin and H. Rotstein, "Control of a camera for active vision: Foveal vision, smooth tracking and saccade," *Intl. Journal of Computer Vision*, vol. 39, no. 2, pp. 81–96, 2000.

[20] G. Sandini and V. Tagliasco, "An anthropomorphic retina-like structure for scene analysis," *Computer Vision, Graphics and Image Processing*, vol. 14, no. 3, pp. 365–372, 1980.

[21] E. Schwartz, "Spatial mapping in the primate sensory projection: Analytic structure and relevance to perception," *Biological Cybernetics*, vol. 25, pp. 181–194, 1977.

[22] C. Silva and J. Santos-Victor, "Egomotion estimation using log-polar images," in *Intl. Conf. on Computer Vision*, January 1998.

[23] T. Stromer, "Irregular sampling, frames and pseudoinverse," Master's thesis, University of Viena, Austria, 1991.

[24] V. J. Traver and F. Pla, "Similarity motion estimation and active tracking through spatial-domain projections on log-polar images," *Computer Vision and Image Understanding (CVIU)*, vol. 97, no. 2, pp. 209–241, February 2005.

[25] S. University of Southern California and I. P. Institute, "The usc-sipi image database," http://sipi.usc.edu/services/database.

[26] R. Wallace, P.-W. Ong, and E. Schwartz, "Space variant image processing," *International Journal of Computer Vision*, vol. 13, no. 1, September 1994.

[27] B. A. Wandell, *Foundations of Vision*. Sinauer Associates, Inc., 1995.

[28] C. Weber and J. Triesh, "Implementations and implications of foveated vision," *Recent Patents on Computer Science*, vol. 2, no. 1, pp. 75–85, 2009.

[29] S. Wilson, "On the retino-cortical mapping," *International Journal on Man-Machine studies*, vol. 18, pp. 361–389, 1983.

[30] M. Yeasin and R. Sharma, "Foveated vision sensor and image processing - a review," in *Machine Learning and Robot Perception*, B. S. S. in Computational Intelligence, Ed. Springer Berlin, Heidelberg, 2005, vol. 7, pp. 57–98.

[31] S. Zucker and R. Hummel, "Receptive fields and the reconstruction of visual information," *Human Neurobiology*, vol. 5, 1986.