A NEW FOVEAL CARTESIAN GEOMETRY APPROACH USED FOR OBJECT TRACKING

José Martínez, Leopoldo Altamirano
National Institute of Astrophysics, Optics and Electronics
Computer Science Department
Luis Enrique Erro No 1. Sta Ma. Tonantzintla
72840, Puebla, México,
{josemcr,altamirano}@inaoep.mx

ABSTRACT
Foveal vision has been used as a way for sampling and reducing the amount of data in Cartesian images for vision systems. For this sampling, there are different approaches as the Log Polar Transform, the Exponential Cartesian Geometry and the Foveal Wavelet Transform between others. In this paper a new approach to obtain the foveal sampling and its application to single object tracking is presented. The approach uses the log polar formulation for the sampling but preserves the Cartesian properties of the information in the original image. In this way, it is possible to overcome the problems of non-linearity of the Log Polar Geometry. Furthermore, it allows an easier object location as the hierarchical processing used in the Exponential Cartesian Geometries and it is easy to implement because it does not imply complex operations for the sampling of the images and the recovery of the original image through the foveated image, contrary to the Foveal Wavelet Transform. The proposed geometry has been tested in diverse images sequences where a single object is tracked successfully by appearance based methods which demonstrate the effectiveness of the proposal.

KEY WORDS

1. Introduction
Object tracking is a very important problem for many applications such as surveillance systems, mobile robots, target aiming and others. Most of cameras used for this task are space-invariant, which means they have a uniform resolution for the whole image. Techniques for tracking and fixation for space-invariant cameras are well established. However, the extensive computation required in tracking algorithms using uniform resolution means that only a small field of view can be analyzed at sustainable frames rates. For example, the processing for object identification and location is performed on restricted areas in the image called search windows, instead of searching in the full image. Some strategies work with a subsampled image, for example the Wavelet Transform is applied in [1] to perform multi-resolution analysis or pyramidal imaging analysis to sample the original image and thus, produce scaled images where the tracking will be performed in a hierarchical way. Other very important approach that has been proposed is the foveal sampling also called foveal vision. This technique proposes the sampling of certain zones in the image with high resolution while the rest is sampled in a decreasing form in relation with the center of the fovea. Foveal vision has been studied with a lot of interest because this has an excellent performance in biological visual systems and diverse strategies have been developed to produce foveated images from Cartesian images. Among these strategies the most representatives are: (1) the Log Polar Transform which makes use of the mathematical formulation based on the research of Schwartz [2], who found evidence that the density of receptor cells on the primate's retina and their retinotopic mapping onto the first area of the striate cortex can be modeled by a logarithmic polar geometry. (2) The so called Exponential Cartesian geometry based on multiresolution schemes. This uses a hierarchical pyramid of images to represent the fovea with higher resolution in the top of the pyramid and with lower resolution in the base by using a foveal polygon [3]. (3) The Foveate Wavelet Transform [4,5] whose idea is similar to the Exponential Cartesian Geometry but with the application of the Wavelet Transform to produce a reconstruction of the original image with higher resolution in a region and decreasing resolution around the periphery of such region.

The use of foveal images, in the single object tracking task implies that the vision system maintains the gaze over the object of interest by maintaining the foveal region or the eye of the fovea over the object. By this way, only important information is kept in the high resolution area. However, in active vision systems, the control mechanisms can be complicated because some times the mechanisms do not have the sufficient speed to keep the camera over the object. It means that any of the foveation strategies must be able to move the center of the fovea over the low resolution area too; consequently instead of moving the head of the camera, the foveation can be performed along the sampled image such as the gazing is done in biological systems without moving the head. This is called the moving foveal property. Both properties, foveation and a moving fovea should be
contemplated if a new strategy to produce foveal images is designed.

Figure 1. Distribution of the sampling over the Cartesian plane and the foveal plane.

Figure 2. A Cartesian image, its foveal image with same number of rings and sector, and the recovered image.

Although the previous strategies to foveation present both properties, some problems limit their use in certain applications and that will be discussed in this paper. For example, in the case of single object tracking, linear algorithms cannot be applied in log polar images. For the case of exponential Cartesian geometries, the identification of the object along the different levels can be difficult because the abrupt changes in resolutions between levels [5]. Finally, the Foveate Wavelet Transform itself implies complex operations which make difficult its implementation in hardware algorithms that can be underlying for real time vision systems.

In this paper a new foveation strategy, inspired by the idea of being able to apply linear algorithm for the single tracking object task, is presented. This approach makes use of the mathematical formulation of the Log Polar Transform to define the way in which the sampling is performed over the Cartesian image. Similar to the exponential Cartesian geometries, the sampling is done such that the foveated image preserves the spatial information of the object in the foveal region with some gradual deformations along the periphery but these does not split the object. Other property is that the implementation is easy and not requires complex operations.

The rest of the paper is the following: in section 2 some of most representative foveation strategies are explained, in section 3 the proposal of this work is described. Section 4 shows some experiments concerning to the object tracking by using the foveated images from video sequences. Section 5 discusses some aspects of the proposal. Finally, conclusion and future work are presented in section 6.

2. Foveation strategies

2.1 The Log Polar Transform

Studies on primates have shown that their visual system present a not uniform distribution of photoreceptors in the retina. They are more densely situated in the central region called fovea, while they are sparser in the periphery. Consequently, high resolution is presented in the central region and it decreases moving away from the fovea to periphery. An important fact is that such distribution can be modeled by using a logarithmic-polar distribution. In according with the model of Kruger [6], the logarithmic factor of the distribution can be formulated taking into account the number of rings $N_r$ and the number of photoreceptor $N_a$ in each ring and such distribution can be formulated as follows:

$$ a = e^{\frac{\log(N_a)}{N_r}} \quad (1) $$

$$ \rho = \sqrt{x^2 + y^2} \quad (2) $$

$$ \eta = \arctan\left(\frac{y}{x}\right) \quad (3) $$

$$ \xi = \log_{p_0}\left(\frac{\rho}{\rho_0}\right) \quad (4) $$

$$ \gamma = \frac{N_a}{2\pi} \eta \quad (5) $$

Equations (2-5) shows the way to map a Cartesian coordinate $(x,y)$ to the called Cortical or foveal plane that concludes with the generation of the foveal coordinates $(\xi,\gamma)$, see Figure 1 and 2. In the model of Kruger, the factor $a$ is used as the logarithmic base for the distribution. The constant $p_0$ is used as the minimum radius or size of the fovea and suggested as different of zero to avoid a singularity in the origin and to allow a homogenously sampled region. The factor $p_{max}$ corresponds to the half-size of the retinal image (Cartesian image) [6]. Finally, the inverse mapping is formulated as it is indicated in equations (6-8). Such definition is useful to construct a lookup table that saves which Cartesian coordinate correspond to each foveal coordinate in a relation of one to one by varying $\xi$ and $\gamma$ with discrete values such that it satisfies $0 \leq \xi \leq N_r$ and $0 \leq \gamma \leq N_a$.

$$ \eta = \frac{2\pi}{N_a} \gamma \quad (6) $$

$$ x = p_0 a^\xi \cos(\eta) \cdot (7) $$

$$ y = p_0 a^\xi \sin(\eta) \cdot (8) $$

In general, the previous mapping is known as the Log Polar Transform and this has been used as a way to obtain foveated or log polar images from Cartesian images. In addition, several works have proposed the use of log polar
images on vision systems applications. For example: estimation of time to impact by using optical flow over the foveated images [7], detection of stationary objects in an indoor scene [8], face detection and tracking [9] among others. In [10] log polar images have been proposed for tracking single objects as well as multiple objects by using differences between images and grey pixel information of the object. In active vision, an extensive study of vergence control and object tracking by using log polar images in a stereo camera configuration is done in [11], [12] and [13]. Other works have proposed the use of a binocular tracking system based on log polar images such as in [14], where the system proposed gazes holding of a moving object in a clutter scene, similar to the work proposed by [15] and [16]. In the last six works, the object identification and location is made through a disparity estimation method applied on foveal images in combination with some gradients methods.

We can observe that all the works mentioned above for object tracking use features based methods. Appearance based methods are not exploited because many of them are difficult to apply directly due to the non-linearity of the transform, producing abruptly deformations, rotations or divisions over the object of interest. For this reason, methods such as correlation are hard to be implemented in foveal images. In fact, when the factor \( p_0 \) is different of zero, some area of the original image is not taken into the sampling so, if the object of interest moves away from the center, new information will be added in the next frame and other will disappear, see third image in Figure 2. By other side, the well known property of scaling and shift invariance in the cortical plane is only valid for centric ones, other functions non-symmetric with the origin will be mapped to very complicated ones [5].

2.2 Exponential Cartesian Geometries

Although for certain applications, object tracking has been successful by using Log Polar images, there are others applications where such methods become inefficient specially where algorithms have to process cluttering scenes, occlusions or gradual changes in the shape of the object. In consequence, it is necessary to research in novel methods or strategies that allow the integration of appearance based methods to object tracking in foveal images. However, appearance based methods are only applicable if spatial relation of pixels in a Cartesian image is preserved in the foveal image.

Exponential Cartesian Geometries were proposed as a way of foveation although not as a direct foveal transform. This approach rests on the idea of the multi-resolution representation and consists in the construction of a pyramid of images from the original by using a strategy known as foveal polygon [3]. The original image is divided as shown in figure 4. This number of areas is defined by the number of rings \( m \) and the subdivision factor \( d \). The first image of the pyramid corresponds to the fovea and is taken directly from the original image. The second image is a combination of the average of cells determined by the factor of subdivision corresponding to the ring and the average of four cells or pixels of the first image. The third image is generated by the same way and so on until the last image corresponds to the same field of vision of the original image.

The advantages of this proposal are that the spatial relation of pixels is preserved in each image of the pyramid. It implies that appearance methods for object tracking can be applied. Also, the formulation is simple and allows the moving fovea property. The scaling in images is not difficult because the reduction is perform by averaging \( 2^{n-1} \) pixels where \( n-1 \) represents the previous level. All this makes the implementations of the direct and the inverse transform easy in software [17] or hardware [18].

However, despite the strength of this approach, some disadvantages are presents when this geometry is used for single object tracking. The principal disadvantage is derived from the different resolutions between the images of the pyramid and emerges when the tracking is performs hierarchically over the images. For example, when a matching algorithm is used, if the object of interest is complete located in one of the images of the pyramid, then the template will find a high peak in the correlation function that will situate correctly the object. However, if the object gets out of the foveal area, then its shape will remain divided in two consecutive or more levels of the pyramid and consequently in different scale. It can introduce noise in the correlation function and can become ambiguous to find the exact position of the object.

2.3 Foveate Wavelet Transform

This is a wavelet-based foveation strategy and provides various merits such as its linearity preservation, orientation selectivity, and high flexibility while exhibiting desirable visual effects reflecting the variable resolution result. It has been used for applications in video compression [5] and in active vision for object tracking [19] and among others.

This approach proposes the generation of four child images generated by using wavelets. These child images are a division by two of the original image such as the pyramidal approach and can be used to recreate the original image fully and thus, convey low-frequency and high-frequency content of the image in both the horizontal and vertical directions. Depending on the distance from the foveation center, high-frequencies can be masked and
compression can be achieved producing the foveation effect, see figure 4.

The advantage of the wavelet approach over the pyramidal decomposition is that no redundant information is added to build the wavelet pyramid. However, the hierarchical processing is implied too besides that the method is computationally more complex which makes it difficult to implement in hardware algorithms, very useful to increase speed in real time applications.

2.4 The Pseudo Log Polar Grid

A variation of the Log Polar Plane is used in the Pseudo-Log Polar Fast Fourier Transform (PPFFT) where a FFT performs the calculation of frequencies in an oversampled set of non-angularly equispaced points which are called Pseudo Polar Grid (PPG) [20]. Figure 5 shows this Pseudo Polar sampling based on the Radon Transform which, in combinations with the Fourier Transform properties in the complex field, allows a fast application of the Fourier Transform by sampling data from the original image as it is shown in Figure 5. Although the PPG is not proposed for the object tracking task, because is only used for sampling, not to produce foveal images, the idea of a non-angularly sampling results interesting and it inspired the proposal of this work.

3. The Proposed Foveal Cartesian Geometry

Taking into account the disadvantages of the traditional foveal strategies and also properties of foveation and moving fovea, it was designed a foveal geometry that overcome such difficulties, sampling with high resolution on the center and decreasing over the periphery, with the formulation necessary to allow moving fovea and in addition, preserving the spatial relation of Cartesian images.

3.1 Implementation

The central idea of this approach rests on the pseudo log polar grid presented in section 2.4. The aim was how to construct a grid whose sampling were high in the center and sparser in the periphery but using square rings and that such rings could be mapped onto the same plane to produce one image and not multi images as exponential Cartesian geometries or the Foveate Wavelet Transform.

The idea emerges in a very simple form by observing the formulation of the Log Polar Transform and the foveation property. First of all, the foveal area was defined by the factor \( p_0 \) which represents the square of minimum radio. By doing this, \( p_0 \) sets the size of the area in the original image that will be copied identically and that will be called foveal square. Later, around this first square area, the rings are located along the image in a logarithmical factor such as it is defined in equation (1) but with squares of radio defined by:

\[
\rho_\xi = p_0 a^\xi \quad \xi = 1, \ldots, N_r \quad (9)
\]

Once the distance between each square is defined, the number of sampled points on each square is set by an equispaced number of points that can be placed around the foveal square such as it is shown in Figure 6. This model is easy to implement and the size of the foveal image is determined by \( 2(p_0 + N) \). By this way, square rings placed around the fovea are denominated as the periphery. From Figure 6 it is easy to see that size of square rings of the periphery has a direct relation with square rings for the sampling in the Cartesian image. Thus, the factor of division for the sampling in each ring is determined by the relation:

\[
\Delta \rho_\xi = \frac{\rho_\xi}{\rho_0 + \xi} \quad \xi = 1, \ldots, N_r \quad (10)
\]

With the previous relations, the mapping can be performed easily and thus a foveal image with the characteristics mentioned at the beginning of this section can be generated. Table 2 shows an algorithm for the direct transform to produce foveal images from Cartesian images as also for the inverse mapping. Of course, depending now only from the number of desirable rings, the size of the fovea is set by \( p_0 \) and the maximum size of the square ring is set by \( p_{max} \). The value of the factor \( a \), calculated in equation (1), will be a real number which must be used in equation (9) without truncation. The value to be truncated or rounded must be the result of equation (9). For this reason, for a few initial values of \( \xi \), the radio of their respective rings will be the same and,
because squares rings with different radio are only of interest, then those with same value wont be taken into account as rings for the foveal image. However, values of the equation (9) are important for the mapping and the inverse mapping and should be remembered. Therefore previous to the mapping, a preparation procedure is performed and must be perform each time the fovea size $p_0$ changes its value, see first algorithm of Table 2.

![Cartesian image](image)

**Figure 7.** A Cartesian image, their sampling by using the mapping proposed by the Table 2 but with a one to one sampling. The last image is the resulting foveal image, enlarged for viewing.

### 4. Experiments

The proposed foveation strategy was tested over several sequences of images of 512x512 pixels where the task consisted in the tracking of a single object selected in the scene by the user. The configuration of parameters is: $p_0 = 30$ and $N=60$, producing that the number of squares rings with different radio remains equal. The size of the foveal image is then of 180x180, a considerable reduction but still with enough information. Figure 8 and 9 presents some sequences where the tracking was performed. For each sequence, tree kinds of images are shown: (1) the original Cartesian image, where the object is selected by the user as well as the position of the object is pointed out. (2) The foveated images where the tracking algorithm is performed. (3) The one to one inverse mapping that shows only how the gazing moves a long the original image as a consequence of the moving fovea always positioned over the object of interest. Of course, the inverse mapping is used to recover the coordinates of the object’s position in the original image.

For the tracking, a simple correlation based on the SAD (Sum of Absolutes Differences) is used. With this strategy a number of 35 frames per second are analyzed in a conventional PC with Pentium 4 processor and Windows XP operative system. This rate can be augmented if we apply strategies used on Cartesian Images to save computational time as trajectories estimators, auto-adjust of the template window and the search window, among others. The precision in the identification and location of the object can be improved by using sophisticated correlation function as for example the correlation with subtracted mean. The aim here consist into present that object tracking with the proposed geometry is feasible and that such mapping preserves the principles of the foveal vision.

Table 1 shows a comparative of the proposal in relation with the single object tracking performed over Cartesian images. The configuration of the foveated strategy remains equal ($rings=60$, $p_0=30$). The SAD correlation is used as tracking methods. The second column shows the number of pixels to process in the worst case for the correlation operation by using the complete image CS. In the fourth column is shown the number of pixels processed in the worst case when a search window strategy WS is used. In this case, the size of the windows search encloses a field vision of 150x150 pixels. In the case of the foveal Cartesian image, a search window with the same field of vision corresponds to the size of 100x100 pixels, in according to the mapping.

<table>
<thead>
<tr>
<th>Image</th>
<th>Size</th>
<th>CS</th>
<th>WS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cartesian</td>
<td>512x512</td>
<td>6,969 720</td>
<td>432 000</td>
</tr>
<tr>
<td>Foveal</td>
<td>180x180</td>
<td>675 000</td>
<td>147 000</td>
</tr>
</tbody>
</table>

CS = Complete Search. WS = Windows Search Strategy

### 5. Deformation of the Object

Figure 7 shows the sampling performed with the proposed strategy over a Cartesian image and the foveal image. It can be observed that linearity is preserved on the foveal region and the first rings rounding this one. The image begins to suffer deformations that do not split the object and the deformations are product of a kind of scaling effect derived of the sampling along the periphery. If the object of interest gets out of the foveal region its shape starts to deform but this is gradual and not abrupt and can be faced with a robust location algorithm. Of course, in the tracking of a single object, the gazing or center of the fovea is located in the next position where the object is located. It makes possible the using of appearance based methods as much as feature based methods because the proposed foveation strategy generates foveal images with Cartesian properties where linear algorithms can be applied such as if the foveal image were a Cartesian image itself.

### 6. Conclusions

In this paper we have presented a new Foveal Cartesian Geometry based on a foveal vision system. This geometry allows extracting a sampling of digital images taken with space-invariant sensors. This sampling is done with high resolution over the center of the image and decreasing over the periphery. The mapping as well as the inverse mapping is simple and easy to implement. The Foveal Cartesian Images generated are compact and thus computational time is saved by using the foveal images generated with the proposed strategy in the single object tracking task. Furthermore, the object of interest mapped to the Foveal Cartesian Image and enclosed in the foveal region is not deformed abruptly such as in traditional foveal images and although the object suffers deformation, it is gradually in proportion to its position.
over the periphery. Experiments show that the matching algorithm was able to identify and locate the object without problems. The proposed geometry makes feasible to use appearance tracking methods with Foveal Cartesian Images. Finally, the simple definition of the mapping and the inverse algorithm makes feasible the possibility of implement the proposed strategy in hardware.

Future work includes the constructions of a more complete object tracking application including diverse strategies used in Cartesian images for tracking single and multiple objects.

References


Table 2. Algorithm for the mapping from Cartesian coordinates to Foveal Cartesian coordinates and its inverse.

<table>
<thead>
<tr>
<th>Preparation(Naρδφm) \ ln(ρδφ/ρ₀)</th>
<th>Mapping(x,y)</th>
<th>InverseMapping(x',y')</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. \ a = e^N</td>
<td>1. \ x' = x' - fcx</td>
<td>1. \ x = x - x₀</td>
</tr>
<tr>
<td>2. \ r = 0</td>
<td>2. \ y' = y' - fcy</td>
<td>2. \ y = y₀ - y₀</td>
</tr>
<tr>
<td>3. \ Γ(0) = ρ₀</td>
<td>3. \ r = max[(x₀),(y₀)]</td>
<td>3. \ ρ = max[(x),(y)]</td>
</tr>
<tr>
<td>for \ i = 1 \ to \ Nr \ do</td>
<td>4. if \ ρ ≤ ρ₀ \ then</td>
<td>4. if \ ρ ≤ ρ₀ \ then</td>
</tr>
<tr>
<td>5. \ ρ(i) = [ρ₀ aᵢ]</td>
<td>5. \ ρ' = ρ₀</td>
<td>5. \ ρ' = ρ₀</td>
</tr>
<tr>
<td>6. \ if \ ρ(i) ≤ Γ(0) \ then</td>
<td>6. \ Γ = [Γ⁺[ρ₀]]</td>
<td>6. \ x = x₀ + t₀</td>
</tr>
<tr>
<td>7. \ r = r + 1</td>
<td>7. \ ρ̇ = \frac{Γ(0)}{ρ₀}</td>
<td>7. \ y = y₀ + t₀</td>
</tr>
<tr>
<td>8. \ Γ(i) = ρ(i)</td>
<td>8. \ Δρ₀ = \frac{Γ(0) - ρ₀}{ρ₀}</td>
<td>8. \ Δρ₀ = \frac{Γ(0) - ρ₀}{ρ₀}</td>
</tr>
<tr>
<td>9. \ end \ if</td>
<td>9. \ x = x₀ + Δρ₀ + x₀</td>
<td>9. \ y = y₀ + Δρ₀ + y₀</td>
</tr>
<tr>
<td>10. \ end \ for</td>
<td>10. \ x' = x' + Δρ₀ + fcx</td>
<td>10. \ x = x₀ + Δρ₀ + x₀</td>
</tr>
<tr>
<td>11. \ y' = y' + Δρ₀ + fcx</td>
<td>11. \ \ return (x₀,y₀)</td>
<td>11. \ \ return (x₀,y₀)</td>
</tr>
<tr>
<td>12. \ end \ if</td>
<td>13. \ \ return (x',y')</td>
<td>13. \ \ return (x',y')</td>
</tr>
<tr>
<td>14. \ \ return</td>
<td>14.</td>
<td></td>
</tr>
</tbody>
</table>
Figure 8. Single object tracking by using Foveal Cartesian Images. Original images in the first line have size of 512x512 pixels; the foveal Cartesian images in the second line have a size of 180x180 pixels. The third line of images corresponds to the recovery of the original image with the size of 512x512 pixels.

Figure 9. Single object tracking by using Foveal Cartesian Images with the same size in each kind of image just like in images of figure 7.

Acknowledgments

This work is in part supported by a grant from CONACYT under No. 189903.