Space Variant Vision for an Active Camera Mount

F. Panerai , C. Capurro, G. Sandini

LIRA-Lab - DIST University of Genova
Via Opera Pia 13 - 16145 Genova, Italy

Technical Report TR 1/95

February 1995

Abstract

Robot systems that rely on vision as their main sensory capability need to be able to cope with changes in the visual environment and to manage a wide field of view. Moreover, in order not to loose real-time response capabilities, selective visual sensing is indeed highly desirable. The "built-in" selection in space and time provided by space variant sensors acts as a filter on the visual field having considerable implications for robotic applications. This paper focuses the attention on log-polar vision in the context of active control of visual sensors. The geometric distribution of sensing elements in the log-polar mapping provides visual task simplification and computational advantages. Correlation measurement techniques in the log-polar framework are formalized and two different uses are proposed. By performing global measurements on convergent log-polar images, a binocular mount can drive its cameras towards correct vergence configuration. It is also shown how image shifts can be detected by using one-dimensional correlation measurements in the log-polar domain, and a possible use of this technique aimed at stabilizing gaze or tracking moving objects is presented. Both the reduced algorithm complexity, due to space variant topology, and the computational advantages, due to the limited number of pixels, make log-polar mapping a good candidate image geometry to obtain real-time responses in the context of reactive vision systems.

The research has been supported by by the ASI, the Italian Space Agency and by the Esprit project VAP-II.

Contents

1 Introduction: Active Visual Perception

Robots need enough perception to accomplish their tasks; active sensing strategies[1; 2] represent the way they can achieve it only by selecting necessary information. In particular, robot vision has taken advantage from the paradigm of active vision[3; 4; 5] that has been defined as: *"The control of sensors and the control of processing to make observation of the external world" (ECUS/NSF Workshop on Vision[6], Ruzenagaard, July 92)* It has been also pointed out that *real-time* and *continuous operation* are the most important characteristics of active vision systems[7]: the system is always running and it returns results within a fixed delay. On the basis of previous statements the following steps and motivations should be kept in mind when experimenting with active vision:

1. **Define the task.** Reactive behaviors to environmental stimuli or more in general integration of basic visual behaviors, obtained through eye control, can simplify vision and support behavior itself.

2. **Identify information necessary for behavioral requirements.** Once the task has been defined the vision system must know what information is to be retrieved from images (i.e. position, velocity of a target).

3. **Select the attentional area in space and time.** Fixed delay responses require limiting the incoming visual data. In this regard it is very important to know where the information is located, and concentrate the processing on it, thus avoiding useless computation.

4. **Choose processing strategies.** Algorithms must retrieve all visual information needed to feed the control loop in at least near real-time.

5. **Study control issues.** Latency due to image processing and noise in the extracted information are the main problems to deal with.

From these considerations it is clear that the choice of image representation is a crucial issue, especially when examining the last three points. The log-polar representation of the image is characterized by a variable resolution on the visual field (an high resolution in the foveal region and a decreasing one going towards the periphery); together with a controlled motion of the sensors it acts as a filter on the observed scene, providing "built-in" selection both in space and time. Other peculiarities of this space variant representation (e.g. symmetries, polar topology, reduced number of pixels) can be exploited in order to fasten and simplify the extraction of the information useful in active vision tasks[8]. These are some of the reasons that have proven space variant vision as being appropriate in different robotic applications [9; 10; 11].

According to the salient points previously mentioned, the paper is organized as follows: section 2 briefly introduces space variant topology with some details useful for algorithm implementations. In section 3 the main behavioral tasks of our active platform and the visual information useful to achieve them are outlined; section 4 is devoted to the choice and the description of visual processing; section 5 presents the active platform and section 6 the experimental results on vergence control and image stabilization techniques.

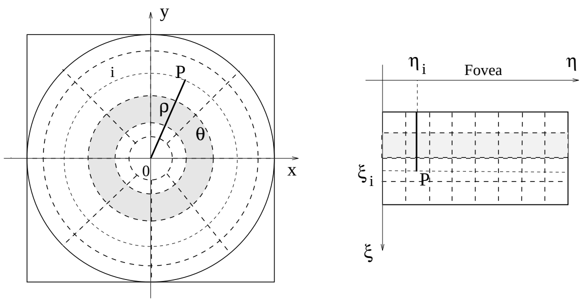


Figure 1: Parameters of retino-cortical mapping. (Left) Any position P in the retinal plane can be expressed in terms of (ρ, θ) or (x, y) coordinates. (Right) In the log-polar plane the same position is identified by (η_i, ξ_i).

2 Log-polar Mapping

The first studies carried out by Daniel and Whitteridge[12] (1961) on visual systems point out that the retinal topology can be optimally described in term of ρ (radius) and θ (orientation). Subsequent studies mainly by Schwartz[13], but also by Hubel and Wiesel[14] yield to the well known analytical formulation of the mapping that occurs between the retina[1] (ρ, θ) (retinal plane) and the visual cortex (η, ξ) (log-polar or cortical plane). The main parameters of the derived logarithmic-polar law (see equations (1) and (2) for the continuous and discrete formulas) that straightforwardly takes into account the linear increment of the photoreceptor size going from the foveal region towards the periphery, are shown in Fig. 1:

$$\begin{cases} \xi = \ln_a \frac{\rho}{\rho_0} \\ \eta = q\ \theta \end{cases} \quad (1) \qquad \text{and the discrete version} \qquad \begin{cases} \xi = \ln_a \frac{\rho_i}{\rho_0} & i \in [1, .., N_{circ}] \\ \eta = q\ \theta_j & j \in [1, .., N_{ang}] \end{cases} \quad (2)$$

where ρ_0, that corresponds to the radius of the innermost circle of the log-polar layout, is obtained on the basis of the dimensions[2] $(width, height)$ of the smallest receptive field and the desired number of cells N_{ang} for each circle. N_{circ} is the number of cell for each radius. The parameter a comes out by imposing that $\Delta\xi$ between two consecutive circles is equal to 1 pixel in the log-polar domain; $1/q$, that corresponds to the minimum angular resolution of the log-polar, is $2\ \pi/N_{ang}$:

$$\rho_0 = \frac{N_{ang}\ width}{2\ \pi} \quad (3) \qquad \ln_a \frac{\rho_{i+1}}{\rho_i} = 1 \quad (4)$$

Substituting in (4) $\rho_i = \rho_0$ and $\rho_{i+1} = height + \rho_0$ we obtain

$$a = \frac{\rho_0 + height}{\rho_0} \quad (5)$$

[1] In this paper we will refer to the retinal plane also through the cartesian coordinates (x,y) with in mind that $\rho = \sqrt{x^2 + y^2}$ and $\theta = \arctan \frac{y}{x}$.

[2] i.e. expressed in micron or retinal pixels

At present the log-polar transformation is obtained at frame rate by using remapping software routines. Hardware remappers[15] and prototypes [16; 17] of space variant CCD's have already been designed and manufactured in the last years and soon a compact camera[18] using C-MOS technology will be available from of an ongoing research project.

3 Basic Visual Tasks and Selection of Visual Information

On the basis of studies [19; 20; 21; 22] on visual systems of some primates and vertebrates it comes out that they show five different types of ocular movements; each of these has a particular function and seems to be activated by precise input stimuli. These behavioral movements are vergence, smooth pursuit, saccadic, miniature and vestibular movements. Though not explicitly considered in the previous classification, cyclotorsion[23] is another important ocular movement related to stereopsis, which is evident in some vertebrates like cats and owls and present in humans too. Apart from vestibular movements, that are related to inertial stimuli (not visual), and miniature movements (the role of which is not yet completely understood) the other four movements are chosen as behavioral tasks for our robotic head. Having a binocular system we need to make a further distinction between visual information (the robot visual stimuli) that relates to monocular and binocular view.

- *Monocular stimuli*: all the information related to motion[3] in the visual field such as linear shift, rotation, expansion or contraction can be perceived monocularly and used according the following scheme:
  - *linear shift*: to control smooth pursuit and saccadic movements.
  - *rotation*: to control cyclotorsion.
  - *expansion* or *contraction*: in vergence control.
- *Binocular stimuli*: visual information related to stereo perception can be expressed as an index of similarity of a stereo pair or on the basis of precise displacement measurement between correspondent points in the left and right images. In both cases the extracted information can be used e.g. to control vergence and cyclotorsion or to segment objects in the horopter[4] to perform binocular tracking.

[3]Our tasks are related to ocular movements only, therefore recognition and navigation issues are not considered in in this paper.

[4]In the human system, any variable point on a circle which also contains the fixation point and the two retinas on the same circle will have a constant angular disparity and it is called *horopter*[24].

4 Correlation Techniques for Space Variant Processing

Given the task and having chosen the most suitable representation of the image (in terms of number of pixels, properties and layout) for focusing the visual attention on an object, it is important to define appropriate processing to measure parameters necessary to feed the control loop. It is worth noting that some of the visual cues we are interested in, such as motion estimation recoverable through a 2D processing in the retinal domain, can be obtained by processing the image along one dimension in the chosen log-polar plane. It is well known that in the retinal plane the components of expansion and rotation are both functions of x and y, while in the log-polar domain they can be expressed as functions of only one coordinate, respectively ξ and η. Perhaps the fact that also the translational component of motion can be estimated with a good approximation solely along the η axis in the log-polar domain is less evident. In this way we are able to discriminate among a number of directions in the retinal plane that correspond to the number of angular positions represented on this axis. We can, however, obtain a precise estimate of the displacement by considering also the ξ direction.
Considering the active and continuous control of camera parameters strictly coupled with the processing, two basic correlation techniques are investigated in order to compute the above mentioned visual information and to obtain a stable behavior.

4.1 Computation of global indices of correlation

Global indices of correlation can be used to obtain a measurement of the similarity between two images. There are many different types of indices, but trying to keep a good balance between their accuracy and their robustness, two of them especially stand out. The first takes into consideration the quadratic difference of the two images:

$$C_1 = \frac{\sum_{\xi,\eta}[I_1(\xi,\eta,t_1) - I_2(\xi+\Delta\xi,\eta+\Delta\eta,t_2)]^2}{N^2}$$

where I_1 and I_2 are the grey levels and N is the number of the pixels of the image. In case of correlation between a log-polar stereo pair ($t_2 = t_1 = t$, $I_1 = I_{left}$, $I_2 = I_{right}$ and $\Delta\xi = \Delta\eta = 0$) index C_1 performs reasonably well for objects that do not show a considerable difference between their projection on the stereo images, but it is very sensitive to changes of brightness and average luminosity of the images. The latter problem becomes almost completely negligible when index C_1 is used to indicate the degree of correlation between subsequent frames acquired from the same camera ($t_1 = t$ and $t_2 = t-1$). The second index is a global correlation index weighted by the variance of the pixel intensity[11]:

$$C_2 = 1 - \frac{\sum_{\xi,\eta}(I_1(\xi,\eta,t_1) - \mu_1(t_1)) \cdot (I_2(\xi+\Delta\xi,\eta+\Delta\eta,t_2) - \mu_2(t_2))}{\sqrt{\sum_{\xi,\eta}(I_1(\xi,\eta,t_1) - \mu_1(t_1))^2 \cdot \sum_{\xi,\eta}(I_2(\xi+\Delta\xi,\eta+\Delta\eta,t_2) - \mu_2(t_2))^2}}$$

Again I_1 and I_2 are grey levels of the images and μ_1, μ_2 represent their mean values. Index C_2 is almost invariant to brightness, much more stable to environmental changes, but

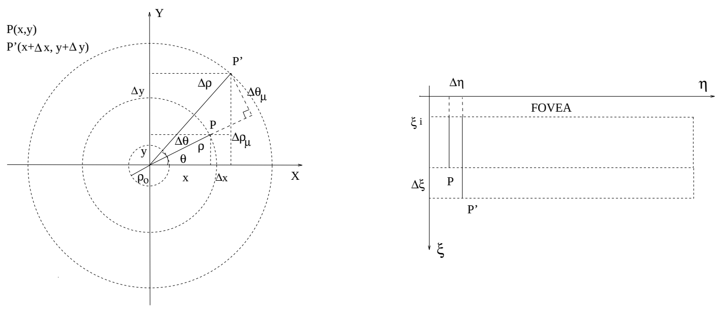


Figure 2: Coordinate systems: retinal plane (left), log-polar plane (right).

computationally more expensive. It has proved particularly useful in the control of vergence with $t_2 = t_1 = t$, $I_1 = I_{left}$, $I_2 = I_{right}$ and $\Delta\xi = \Delta\eta = 0$; it is normalized between 0 and 1 and has a minimum in the correspondence of the correct vergence angle. In this case the space variant mapping deserves a special consideration because it is only by using it that these indices gain most of their effectiveness. Without the mapping there would be no good index maximum except in the case where the target covers more than half of the image.
The way parameters $\Delta\xi$ and $\Delta\eta$ are varied allows to select the kind of information we want to retrieve when processing two subsequent log-polar frames. If we compute, for example,

$$\min_{\Delta\xi,\Delta\eta} C_1 \quad with \quad \Delta\xi \in [1,..,N_{circ}] \ , \ \Delta\eta \in [1,..,N_{ang}]$$

we are able to estimate the amount of rotation $\Delta\eta$ and zoom $\Delta\xi$ of one image compared to the previous one. Otherwise, if the goal is to measure the *virtual*[5] displacement between two log-polar images (may be also a stereo pair) by using the same kind of computation, $\Delta\xi$ and $\Delta\eta$ need to be varied differently as it will be explained in the following section.

4.2 Log-polar correlation

Correlation in log-polar plane, in order to recover linear displacements occurred in the retinal domain, must be considered in a different way. In fact the effective measure of distance (amplitude and direction) between two pixels in log-polar images is a function of the position of the pixels, while in the retinal domain distance is invariant with pixel coordinates and depends only on the reciprocal position of the pixels. The log-polar correlation can be in general defined as

$$Corr(I_1(\xi, \eta, t_1), I_2(\xi + \Delta\xi, \eta + \Delta\eta, t_2)) \tag{6}$$

where $Corr$ is the function chosen to express the correlation between the two images I_1 and I_2. In order to map retinal linear shifts in the log-polar domain, we can compute the $\Delta\xi$, $\Delta\eta$ increments as a function of the correspondent Δx, Δy increments and of the position (ξ, η) in the log-polar plane. Consider the coordinate system in Fig. 2. By differentiating

[5] the transformation of a retinal shift $(\Delta x, \Delta y)$ into the log-polar domain.

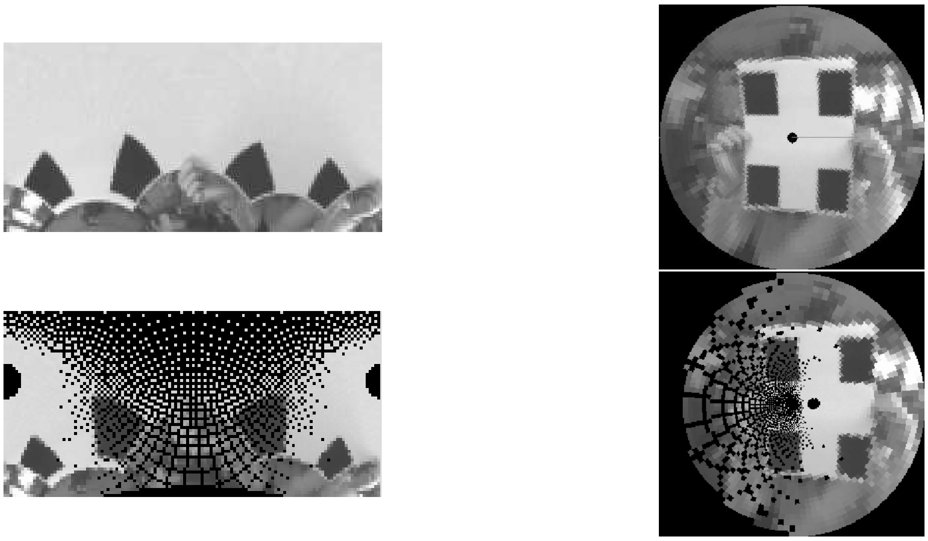


Figure 3: Example of a *virtual* retinal right shift of 20 pixels performed in the log-polar domain (even the hole that corresponds to the fovea results 20 pixels right shifted). Log-polar image dimension is are 128×64.

the system of equations (1), that provides the definition of the log-polar mapping, and by integrating the obtained formulas from initial positions i, j to final positions fi, fj, it is possible to evaluate $\Delta\xi$, $\Delta\eta$ increments in the discrete domain[6]:

$$\begin{cases} d\xi = \frac{d\rho}{\ln a \;\; \rho} \\ d\eta = q\ d\theta \end{cases} \qquad (7) \qquad\qquad \begin{cases} \Delta\xi = \xi_{fi} - \xi_i = \int_i^{fi} d\xi = [\log_a \frac{\rho_i + \Delta\rho}{\rho_i}] \\ \Delta\eta = \eta_{fj} - \eta_j = \int_j^{fj} d\eta = [q\ \Delta\theta] \end{cases} \qquad (8)$$

where, considering that $\rho_i = \rho_0 a^i$ from the log-polar mapping definition (4),

$$\rho_i + \Delta\rho = \frac{\Delta\theta_\mu}{\sin\Delta\theta} \qquad (9)$$

$$\Delta\theta = \arctan\frac{\Delta\theta_\mu}{\rho_i + \Delta\rho_\mu} \qquad (10)$$

and

$$\Delta\theta_\mu = \Delta y \cos\theta_j - \Delta x \sin\theta_j \qquad (11)$$

$$\Delta\rho_\mu = \Delta y \sin\theta_j + \Delta x \cos\theta_j \qquad (12)$$

Substituting (9), (10) into (8) we obtain the final expression for $\Delta\xi$, $\Delta\eta$:

$$\begin{cases} \Delta\xi = [\ \frac{1}{\ln a} \ln \frac{\Delta\theta_\mu}{\sin(\frac{\Delta\eta}{q})\rho_0 a^i}\ ] \\ \Delta\eta = [\ q \arctan \frac{\Delta\theta_\mu}{\rho_0 a^i + \Delta\rho}\ ] \end{cases} \qquad (13)$$

[6]square brackets are used to indicate the integer part.

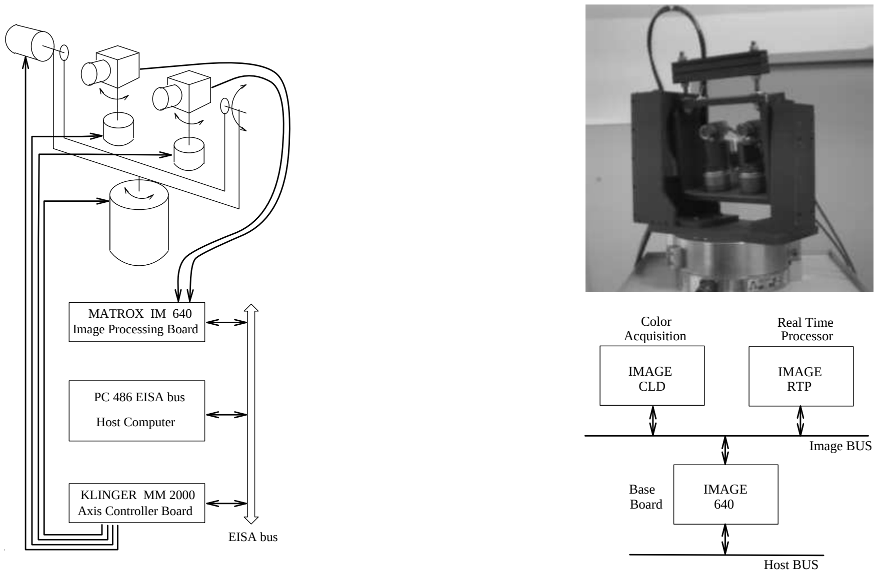


Figure 4: (left) The set-up used in the experimental phase; (right) a picture of the head mount and a detail of the image processing hardware.

Fig. 3 shows an example of the effect introduced in the log polar domain by a linear retinal right shift. The images, from the top left to the bottom right, are respectively the original log-polar image, the same image remapped into the retinal plane, the log-polar image of the original *virtually* right shifted and the remapped one.

5 The Controlled Platform

The active vision hardware consists of a robotic platform, a motor control board, an image processing device set and a host computer. Figure 4 shows a front view of the mount and some details of the image processing hardware. The platform is quite bulky if compared to more recent active vision systems[25] and it is not optimized in relation to the mechanical structure. The original aim of this design was to experiment with a limited number of d.o.f. in order to define the minimum requirements necessary to achieve autonomous behaviors. The range for vergence rotations extends approximately from -45 to 45 degrees. Common tilt and pan rotations have a larger range (-160 to 160 degrees), enabling the system to point almost anywhere in space. Stepper motors have been used for independent vergence movements, while DC direct drive motors control common tilt and pan rotations. Stepper motors provide an excellent performance, reaching velocities up to 1400 deg/sec with accelerations up to 80000 deg/sec^2. DC motors provide a better torque but a lower performance in velocity and acceleration (190 deg./sec., 1400 deg./sec^2). The four motors are controlled by a single board (Klinger MM2000) plugged into a 486 PC EISA bus. The image processing hardware (Matrox IM-640 board set), which is in charge of processing the visual information, is plugged on the same bus. We would like to stress here that the reduced number of pixels typical of space variant sensors allows an efficient use of a limited computational power. This approach may be very relevant and appropriate for applications where a reduction of system complexity, size and weight are demanded.

6 Experimental Results

6.1 Space variant vergence

Robot vision systems that are using foveated visual architectures must improve in the future their ability to control the non-uniform spatial resolution photoreceptor arrays in order to recover the necessary amount of information from the observed scene. Vergence movements have been studied in the past by many researchers[19; 21; 20]. In recent years control of vergence has received attention from active vision researchers and successful implementation results have been achieved within the context of space invariant frameworks. The space invariant approach relies on computationally expensive techniques like binocular segmentation, blur detection and disparity estimation[26; 27]. On the other hand, the reduced computational cost and the reduced algorithmic complexity offered by space variant sensing, together with other interesting features[28], can be exploited in order to efficiently control vergence. Within the log-polar framework, a vergence configuration can be considered correct when the registration between the stereo pair is as high as possible. Control of vergence seems advantageous in the context of log-polar image representation because it enables to:

- extract information from the whole log-polar image, avoiding the segmentation issues,
- emphasize binocular fusion rather than geometrical intersection of optical axes,
- exploit a global measurement index between the stereo pair,
- reduce computational cost and algorithmic complexity.

In section 4.1 we touched the idea of using correlation techniques to measure the similarity between the log-polar stereo pair in order to control camera vergence. Correlation indices C_1 and C_2 can be specified for stereo processing by imposing $t_2 = t_1 = t$, $I_1 = I_{left}$, $I_2 = I_{right}$ and $\Delta\xi = \Delta\eta = 0$. The advantage of using correlation techniques within a space variant framework can be even more emphasized by considering the following comparison among four different geometric mappings. The leftmost side of Fig. 5 show the standard uniform sampling tessellation; on the right, there are three different types of space variant mappings: log-polar, log-cartesian, horizontal log-cartesian. The plot below correspond to the behavior of the correlation index C_2 for the four different tessellations with respect to

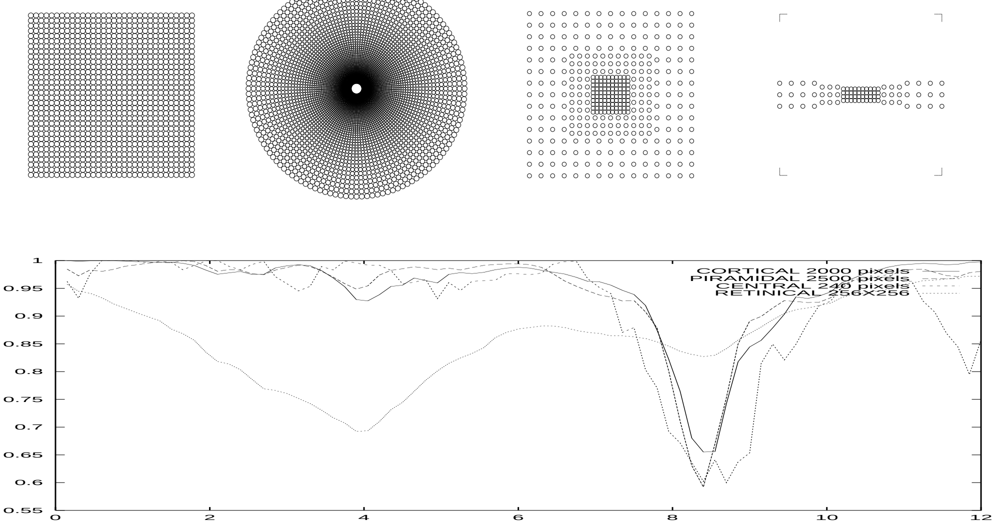


Figure 5: Plottings correspond to different space variant mappings. From left to right: uniform (retinical), log-polar (cortical), log-cartesian (piramidal), horizontal log-cartesian (central) tessellations.

a complete span over the vergence range. It can be seen that space variant mappings are unaffected by false minima deriving by incorrect matching in the periphery of the visual field, while it occurs for the cartesian geometry. In addition, it can be seen that as the binocular camera system approaches the correct vergence configuration the degree of correlation between the stereo pair becomes even higher in space variant mappings. This fact can be better understood by considering that in log-polar coordinates the horopter cover a higher percentage of the image plane. To support this Weiman and Fisher [29] showed how magnitude of intrinsic disparity for a convergent log-polar view is pushed further toward the periphery when compared to the retinal case. The presented comparison also put into evidence the fact that correlation index C_2 behaves almost similarly for the log-polar and for the log-cartesian mappings. On this regard we argue that by choosing a log-polar image representation other advantages can be gained that log-cartesian image representation do not offer. This will be further motivated in the section 6.2.

Consider now the vergence scenario in Fig. 6(left) adopted in the experimental phase. The two indices C_1 (center) and C_2 (right) are plotted with respect to the angle θ (read out from the motors) for different speed values. Deep valley shaped profiles are centered around the correct vergence configuration. Spatial filtering provided by log-polar sensing and computationally inexpensive global measurements on the stereo pair allow to identify the correct vergence configuration. The problem could be finding vergence direction, because the correlation measurement on itself does not provide any information about direction. We have found that as long as the vision system keeps the cameras moving, a kind of gradient descending technique can be used to control the angular velocity of the cameras. The graphs of Fig.6 show that the derivative of the correlation is low when cameras are far from the correct vergence configuration while it is high when cameras approach correct

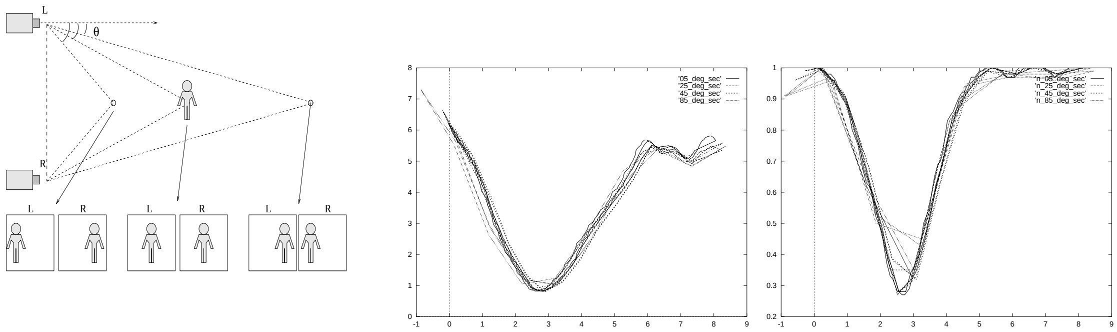


Figure 6: Minimization over time of *global indices* is obtained through the control of $\dot{\theta}$. Graphs correspond respectively to the correlation index C_1 and C_2 plotted for different velocities (continuous lines correspond to the slowest motion) of a verging-diverging movement with θ varying between 0 and 8 degrees.

vergence configuration. Using the inverse derivative of correlation and a normalization factor equal to the correlation index C_2 to smooth the noise coming out from the derivative of correlation index, a velocity signal of the following form can be used to control the motion of the cameras towards the correct vergence configuration:

$$\dot{\theta} = -\frac{C(\theta, t)}{\frac{dC(\theta,t)}{d\theta}} \tag{14}$$

Two considerations are important at this point: the first is that a similar result can be recovered analytically by considering a different expression of index C_1, for example, where the dependence on *time* and *angular position* of the cameras are put into evidence:

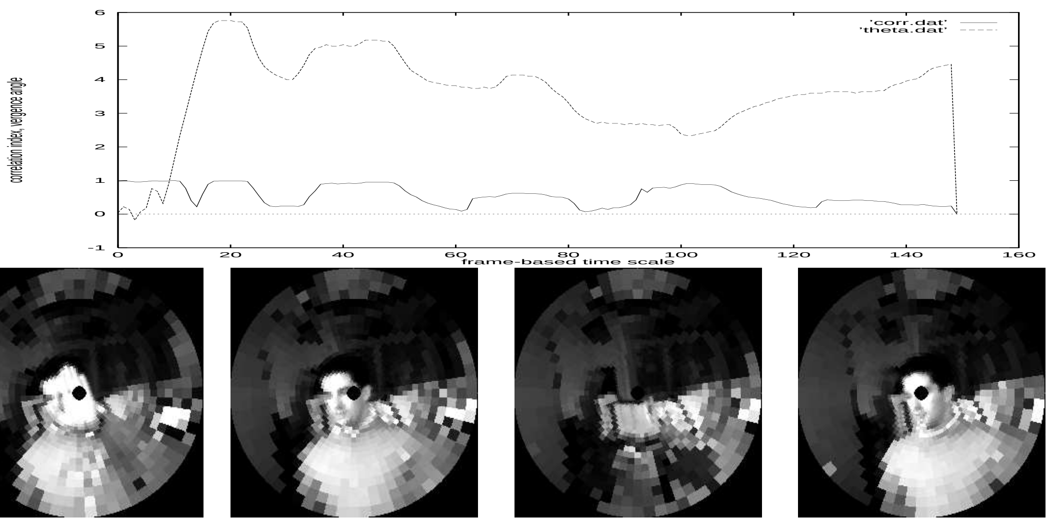
$$C_1(\theta, t) \;=\; \frac{\sum_{\xi,\eta}[I_{l\_\xi\eta}(\theta, t) - I_{r\_\xi\eta}(\theta, t)]^2}{N^2}$$

If we differentiate it, it comes out that the velocity that minimize this index[7] is proportional to the inverse of $\frac{dC(\theta,t)}{d\theta}$, as evidenced in (14), and it is:

$$\dot{\theta} = -\frac{\sum_{\xi,\eta}[(I_{lt} - I_{rt})\ (I_l - I_r)]}{N^2\ \frac{dC(\theta,t)}{d\theta}} \tag{15}$$

where I_{lt} and I_{rt} are the temporal derivatives of images I_l and I_r. Even if this expression is interesting from a theoretical point of view, because $\dot{\theta}$ becomes zero when the vergence is correct ($I_l = I_r$) or when the target is moving in the horopter ($I_{lt} = I_{rt}$), the computation of the expression in the numerator of (15) can be very noisy and this is the reason why we chose (14) to control vergence. The second consideration is about the reading of angular position directly from the motors. To avoid coarse errors when computing the derivatives of correlation indices with respect to camera position it is necessary to synchronize acquisition

[7] the same holds for index C_2



frame=30 frame=62 frame=85 frame=124

Figure 7: Vergence angle and correlation index behavior are plotted in an evolving scenario. The four remapped images (64×30 format) below correspond to correct vergence configurations. Frame references enable to match them onto plateaus present in dotted profile of vergence trajectory.

and data reading. A delay introduced between the acquisition and the reading from the motors corresponds to an additional error in angular position and it increases with motor speed[8]. Fig. 7 shows the behavior of the system responding to changes in environment by adjusting vergence in accordance with visual stimuli. A hand, a face and the background at different depths are sequentially presented to the vision system; the correlation index is plotted against time (expressed in a frame based scale) and correct vergence configuration can be identified by low values of the correlation index and almost constant angular positions.

6.2 Gaze stabilization and motion tracking

Correlation techniques are used in this section to recover image shift estimates in terms of direction and shift amount. Using a log-polar representation of the image we can measure retinal image shifts due to rotation of the camera about its nodal point, or due to translational movements. Rotation about and motion along the optical axis are not considered; however they are not among the degrees of freedom of our robotic platform. When mainly rotation is involved we can measure a constant retinal shift $\Delta\rho$ (see Fig. 8 in the center)

[8]If it is not possible to synchronize acquisition and data reading, two pairs of frames can be acquired before computing a new velocity for the motors, so that derivatives are computed in correspondence of a constant speed.

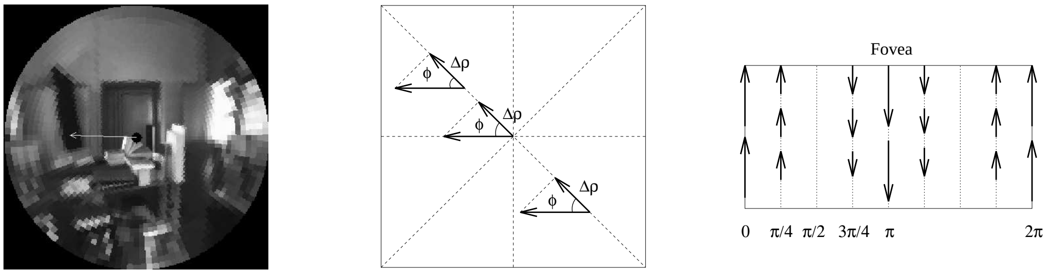


Figure 8: (left) Image from a pan sequence, the arrow indicates the direction of motion; (center) retinal shift; (right) log-polar shift.

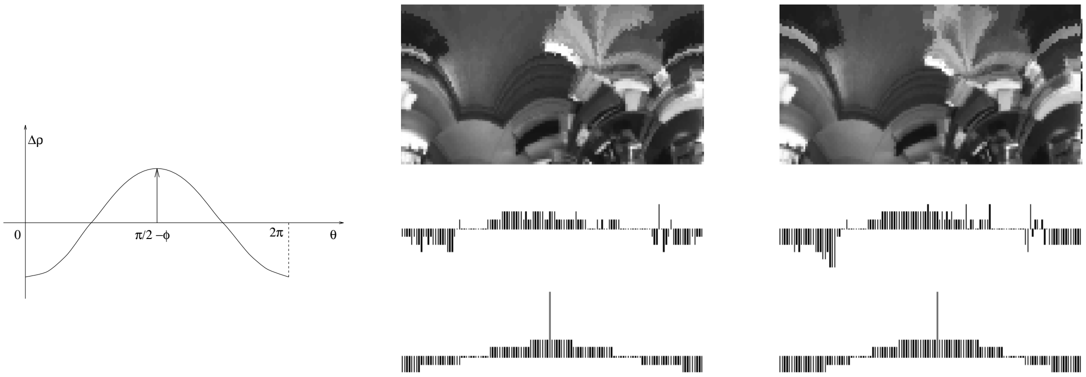


Figure 9: (left) Remapped image and two subsequent log-polar frames from the same pan sequence; histograms of computed $\Delta\rho$ vs. θ on the upper part and of the corresponding DFT first harmonic below. The range of θ extends from 0 to 360 degrees.

along radial lines corresponding to ξ direction (columns) in the log-polar domain(see Fig. 8 in the right). When both translation and rotation are present (and this is a common situation in which stabilization is required) a constant phase ϕ of the shift with respect to the radial direction is still preserved. In this case only the direction of the shift can be recovered (translational egomotion gives rise to shift amplitudes on the image plane that depend on the distance to objects in the scene) in order to actively compensate for motion. The same kind of processing can be applied to evaluate motion of foveated objects by assuming that it can be locally approximated by shift in the retinal image plane. In general this approach can be used in all the situations in which conventional[9] correlation techniques prove themselves successful. We would like to emphasize here that the above mentioned classes of motion fields, traditionally sharing a 2-D retinal representation, can be recovered with a good approximation by a 1-D analysis on a reduced number of pixels (64×30 and 128×64 log-polar images have been tested). Consider as an example the remapped image in Fig. 8(left) acquired during a pan movement of the camera. Correlation equations presented in section 4.2 can be explicitly tailored to detect, for example, image shifts in the mentioned

[9] referred to the retinal plane

case of rotation:

$$Corr(I_t(\xi,\eta), I_{t-1}(\xi+\Delta\xi(\Delta\rho),\eta)) = \min_{\Delta\rho} C_1$$

where

$$\Delta\xi = [\ \log_a \frac{\rho_i + \Delta\rho}{\rho_i}\ ] \qquad \text{and} \qquad \Delta\rho = \Delta\rho_\mu = \frac{\Delta x}{\cos\theta} = \frac{\Delta y}{\sin\theta}$$

Plotting $\Delta\rho$, that is constant along each radial direction, against the direction itself (θ) we come to a discrete sinusoidal function (see figure 9-left) whose values can be analyzed in terms of amplitude and phase in order to recover the vector of motion. Figure 9 illustrates histogram behavior of $\Delta\rho$ versus θ corresponding to a pan of the camera: two subsequent log-polar frames of the same pan sequence of figure 8 are shown above the raw data histograms and the derived DFT first harmonic interpolation. Positive peaks on DFT histograms represent the estimated motion direction the amplitude of which is given by first DFT harmonic amplitude. The described processing is extremely simple and can run in real time also on a 486-PC based architecture.

7 Conclusions

This paper has presented some experimental results on the use of log-polar image representation in the context of active control of visual sensors. Use of correlation techniques in the log-polar domain has proven to be efficient to extract visual cues for controlling vergence behavior and providing robust motion estimation of the visual field. In the context where an active and continuous control of camera's parameters is required in order to reach a stable system behavior, sometimes a less accurate real time processing is more suitable than a very precise processing requiring a long time. On the other hand, computational advantages due to reduced number of pixels allow an efficient use of limited computational power. This may be relevant and appropriate for applications that require reduction of system complexity and consequently a reduction of physical size and weight.

8 Acknowledgments

The authors would like to thank T. Uhlin, F. Recio, J. Nielsen and P. Questa for helpful discussions and the valuable contributions to software development.

References

[1] R. Bajcsy and P. Allen. Sensing strategies. In *U.S. - France Robotics Workshop*, Philadelphia, PA, 1984.

[2] R. K. Bajcsy. Active perception vs passive perception. In *Proc. Third IEEE Computer Society Workshop on Computer Vision: Representation and Control*, pages 13–16, Bellaire, MI, 1985.

[3] D.H. Ballard, R.C. Nelson, and B. Yamauchi. Animate vision. *Optics News*, 15(5):17–25, 1989.

[4] J. Aloimonos, I. Weiss, and A. Bandyopadhyay. Active vision. *International Journal of Computer Vision*, 1(4):333–356, 1988.

[5] J. Aloimonos. Purposive and qualitative active vision. In *Proc. of Int. Workshop on Active Control in Visual Perception*, Antibes, France, April 1990.

[6] J. Crowley and R. Bajcsy. Ecus/nfs workshop on vision. In *Workshop on Active Vision Hardware*, Ruzenagaard, Sjællands Odde, Denmark, July 1992.

[7] J. Crowley. What is active vision? In *Workshop on Active Vision Hardware (Slide-book)*, Le Sappey, France, February,1995.

[8] C. Capurro, F. Panerai, E. Grosso, and G. Sandini. A binocular active vision system using space variant sensors: Exploiting autonomous behaviors for space application. In *Int. Conf. on Digital Signal Processing*, Nicosia, Cyprus, July 1993.

[9] C. F. R. Weiman and R. D. Juday. Tracking algorithms using log-polar mapped image coordinates. In *SPIE Int. Conf. on Intelligent Robots and Computer Vision VIII: Algorithms and Techniques*, volume 1192, pages 843–853, Philadelphia (PA), 6-10 November 1990.

[10] T. Deitrich. Target tracking with a retina-like spatially variant sensor. Technical report, University of Pennsylvania, May 1991.

[11] J. Nielsen and G. Sandini. Learning mobile robot navigation : A behavior-based approach. In *IEEE Int. Conf. on Systems, Man and Cybernetics*, San Antonio, Texas, October 1994.

[12] D. Whitteridge P. M. Daniel. The representation of the visual field on the cerebral cortex in monkeys. *J. Physiol. London*, 159:203–221, 1961.

[13] E. L. Schwartz. Spatial mapping in the primate sensory projection: Analytic structure and relevance to perception. *Biol. Cybernetics*, 25:181–194, 1977.

[14] D.H. Hubel and T.N. Wiesel. Functional architecture of macaque monkey cortex. *Proc. R. Soc. Lond.*, 198:1–59, 1977.

[15] T.E. Fisher and R.D. Juday. A programmable video image remapper. In *Proc. SPIE*, volume 938, pages 122–128, 1988.

[16] I. Debusschere, E. Bronckaers, C. Claeys, G. Kreider, J. Van der Spiegel, P. Bellutti, G. Soncini, P. Dario, F. Fantini, and G. Sandini. A 2d retinal ccd sensor for fast 2d shape recognition and tracking. In *Proc. 5th Int. Solid-State Sensor and Transducers*, Montreux, June 25-30 1989.

[17] A.S. Rojer and E.L. Schwartz. Design considerations for a space variant visual sensor with complex logarithmic geometry. In *Proc. Int. Conf. on Pattern Recognition*, Philadelphia, PA, 1990.

[18] F. Ferrari, G. Sandini, L. Hermans, C. Guerin, A. Manganas, P. Dario, and H. Frowein. Tide project 1038 ibidem, tecnical annex. Technical report, Ibidem Consortium, June 1994.

[19] R.H.S. Carpenter. *Movements of the Eyes.* Pion Ed., 1988.

[20] J. Pola and H.J. Wyatt. Smooth pursuit: Response characteristics, stimuli and mechanisms. In R.H.S. Carpenter, editor, *Eye Movements.* The Macmillan Press, 1991.

[21] D.A. Robinson. The systems approach to the oculomotor system. *Vision Research*, 26(1):91–99, 1986.

[22] H.J. Wyatt and J. Pola. Smooth eye movements with step-ramp stimuli: the influence of attention and stimulus extent. *Vision Research*, 27(9):1565–1580, 1987.

[23] J.D. Pettigrew M. Cooper. A neurophysiological determination of the vertical horopter in the cat and owl. *The Journal of Comparative Neurology*, 184:1–26, 1979.

[24] C.F.R. Weiman N.C. Griswold, J.S. Lee. Binocular fusion revisited utilizing a log-polar tesselation. *Computer Vision and Image Processing*, pages 421–457, 1992.

[25] P.M. Sharkey, D.W. Murray, S. Vandevelde, I.D. Reid, and P.F. McLauchlan. A modular head/eye platform for real-time reactive vision. *Mechatronics*, 3(4), 1993.

[26] D.J. Coombs, T.J. Olson, and C.M. Brown. Gaze control and segmentation. In *Proc. of the AAAI-90 Workshop on Qualitative Vision*, Boston, MA, July 1990.

[27] K. Pahlavan, T. Uhlin, and J.O. Eklund. Integrating primary ocular processes. In *Proc. Second European Conference on Computer Vision*, pages 526–541, Santa Margherita, Italy, 1992. Springer-Verlag.

[28] G. Sandini and M. Tistarelli. Vision and space-variant sensing. In H. Wechsler, editor, *Neural Networks for Perception.* Academic Press, 1991.

[29] C.F.R. Weiman and T.E. Fisher. Log-polar binocular vision system. Technical Report NAS 9-18637, Nasa Final Report, Danbury, CT, December 1994.