Sensor geometry for dynamic vergence: characterization and performance analysis *

Alexandre Bernardino José Santos-Victor

Instituto de Sistemas e Robótica - Instituto Superior Técnico,
Av. Rovisco Pais, 1096 Lisboa Codex, PORTUGAL
e-mail: {ajmb,jasv}@isr.ist.utl.pt

Abstract

Real-time control architectures for active vision impose strong constraints on the computational complexity of image processing algorithms. In many cases, fast and reactive visual routines are achieved by drastically reducing the image size, at the cost of degrading the accuracy of the algorithms.

The non-uniform sampling of the images, such as the log-polar mapping, is an efficient way of reducing the amount of data to process [7, 17], while maintaining a good resolution in the center of the visual field. Additionally, log-polar images have several advantages over cartesian images, namely for the control of vergence in stereo active vision devices.

In the work described here, two different vergence control strategies are presented. Both are based on the correlation of low-resolution log-polar images and the control law is a function of direct image measurements, leading to fast visuo-motor processes.

In this paper we focus on the problem of characterizing the performance of the vergence system relative to various log-polar sensor layouts and control strategies. The quality of a given sensor layout is assessed by analyzing the shape of the correlation function for various values of the vergence angle.

Therefore, we introduce a set of quality criteria, depending on the sensor layout, that ultimately determine the performance of the control algorithm. The influence of the log-polar sensor parameters on the shape of the correlation function is analyzed, and criteria to choose adequate layouts are established. Experiments on real image data, in closed loop control, are shown.

1 Introduction

Research on active vision has increased significantly in the past few years [1, 2]. In many laboratories, Stereo Heads have been designed and built, thus offering an experimental platform for the study and test of many active vision problems (see [4] for an overview).

Very often, such systems have been inspired after biological vision systems [3, 11] and visuo-motor processes like saccades, vergence, and smooth pursuit that can been found in the human oculomotor system, have been implemented and tested in various active robot heads. In many cases, the mechanical design of such systems resembles closely the degrees of freedom and structure of the human oculomotor system and, although not having the whole motion capabilities, they can still perform the basic movements required to carry out most of the visual tasks.

In nature, however, the main need to perform eye movements is the space variant resolution of the retina, as opposed to the regularly sampled cartesian images that are widely used in most active vision systems. The human retina, for instance, has a central area - the fovea - with a high density of photoreceptors, while in the periphery this density is low. The space variant resolution of the image retina can be modeled as a logarithmic-polar law, as the space between the photoreceptors increases linearly with the distance to the center [13].

Accordingly, eye movements in biological vision systems are required to direct the high resolution fovea towards the scene regions where attention is focused. The advantage of this arrangement is the concentration of the

*Alexandre Bernardino is currently supported by a grant from PRAXIS-XXI research program. This work was partially funded by projects PRAXIS/3/3.1/TPR/23/94 and JNICT-PBIC/TPR/2550/95.

computational resources in the image regions of interest, while the entire visual field can still be considered and, for example, peripheral events can be detected and further processed. Therefore, we can obtain the visual information needed for many tasks, without considering all the data contained in uniformly sampled images, thus avoiding the waste of processing effort.

Even if, apart from few exceptions [8], log-polar images have seldom been used in active vision systems, they offer important advantages over cartesian images for a number of visual tasks. For instance, they have interesting invariant properties relative to shape scaling and rotation [16] and, they are particularly suited for binocular fusion [17] and vergence control.

The work described here has two main contributions. The first aspect is related to the performance characterization of log-polar sensors when used for vergence control. In fact, the design of a log-polar sensor depends on the specification of a number of layout parameters whose influence in the overall vision algorithm is hardly ever addressed. Here, we study the influence of each of these parameters in the overall performance of the vergence system.

The input to the vergence control system is a correlation index between two log-polar images. Hence, the behavior of the correlation index, as a function of the layout parameters, will ultimately determine the global algorithm performance. We introduce criteria that allow an evaluation of the various design parameters and, therefore, establish a methodology to choose an appropriate design. This evaluation is done both through **analytical comparisons** between log-polar and cartesian images and by **extensive testing with real images**.

A second relevant aspect in this work is the vergence control system itself. Extensive work has been done in disparity estimation that could be used for vergence control. Most of these techniques cannot be directly used with the log-polar image geometry, since a translating target will be imaged as a space-variant template that deforms when crossing the visual field.

The two control systems proposed, result from the approach of relying on a continuous stream of visual input that is used to achieve the correct vergence. The *iterative vergence* control is based on the temporal optimization of the correlation index between the two images, without any explicit estimation of disparity. The *pre-programmed vergence* control is based on a set of pre-defined disparity channels whose output is used to determine the best disparity estimate.

In spite of these differences, the algorithms are based on the same type of image clues and the performance analysis is very similar. In the paper, we extensively describe and compare both control approaches.

In our opinion, this paper addresses a number of important issues in vision and active vision systems, particularly when considered together. First, there is the concern to characterize the performance of the vision system as a whole. This analysis has consequences at two levels. It provides guidelines for the log-polar sensor design for a specific task. Secondly, it is related to the particular class of control systems to be used in a given problem. In our opinion, the holistic analysis of global performance characteristics, sensor design and control system specification may have a significant impact in various applications of vision algorithms and robotics.

1.1 Structure of the paper

In Section II, we review the main aspects of the log-polar mapping, particularly those that are relevant to the problem addressed : vergence control through the use of purposive cues extracted from log-polar image stereo pairs. In particular, we describe the sensor geometry parameters that play important role in the performance of the vergence process.

In Section III, the vergence behavior is introduced, referring its importance in biological and artificial systems. A review of usually applied techniques for vergence control is made, with greater emphasis on correlation techniques, where the advantages of log-polar over cartesian geometry are illustrated and discussed.

The two proposed control algorithms, *iterative vergence control* and *preprogrammed vergence control*, are described in Section IV.

Section V is devoted to the characterization of the performance of the vergence system, as a function of the sensor layout, based on the shape of the correlation function. This procedure determines qualitative criteria to assess the influence of using different log-polar sensor layouts, on the response of the vergence algorithm.

In Section VI, we present a real-time implementation and various experiments that illustrate the influence of sensor geometry on the performance of the vergence system, according to the characterization described in Section V. Closed loop experiments are described and discussed. Finally, in Section VII some conclusions are drawn, and the main directions of future research established.

2 Log-Polar Mapping

Even though the research activity based on the active vision paradigm has grown in the past years, far less attention has been given to the problem of imaging geometry and sampling. In fact, the main reason for ocular movements in many biological systems consists on the existence of a space variant resolution in the retina. In the human visual system, the receptors at the retina are distributed in space with increasing density toward the center of the visual field (the fovea), and decreasing density from the fovea towards the periphery. Therefore, eye movements are used by the human oculomotor system to point the central high resolution fovea in the direction of the scene areas to observe more carefully.

However, apart from few exceptions [8], most active vision systems use cartesian image representations. Instead, the vergence system described in this paper will be based on a log-polar space variant image sampling.

We start by briefly reviewing some of the existing work on the log-polar image mapping [13, 16, 17]. We will describe the notation used throughout the paper and focus on the design parameters that will be studied in the experiments.

The log-polar transformation is a conformal mapping from the points on the cartesian plane (x, y) to points in the log-polar plane (ξ, η). The mapping is described by :

$$\begin{cases} \xi & = \log_k(\rho) \\ \eta & = \theta \end{cases} \tag{1}$$

where k represents a decaying resolution factor and ρ and θ denote the commonly used polar coordinates :

$$\begin{cases} \rho & = \sqrt{x^2 + y^2} \\ \theta & = \arctan y/x \end{cases} \tag{2}$$

The log-polar mapping exhibits a singularity in the origin of the coordinate plane, as ρ tends to zero. This means that the mapping will not be defined in a neighborhood of the origin :

$$\xi \in [\log_k(\rho_{min}), \log_k(\rho_{max})],$$

where ρ_{max} depends on the cartesian plane dimension. As defined initially, the logarithmic-polar mapping yields continuous values in the domain $[\log_k(\rho_{min}), \log_k(\rho_{max})[\times[0, 2\pi[$. These values must be discretized into $E \times A$ cells (pixels), where E is the number of eccentricities and A is the number of cells per eccentricity. Figure 1 shows the discretized cartesian and log-polar planes.

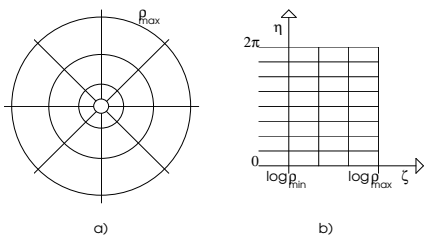


Figure 1: Space variant mapping from the cartesian plane, (a), to the log-polar plane, (b).

In the ξ direction, the full range $(\log_k \rho_{max} - \log_k \rho_{min})$ is mapped in a range of E pixels, which imposes a constraint on the value of the logarithmic parameter, k :

$$E = \log_k \rho_{max} - \log_k \rho_{min} \Rightarrow k = (\frac{\rho_{max}}{\rho_{min}})^{\frac{1}{E}} \tag{3}$$

According to this, and defining r and s as the pixel coordinates in the discretized log-polar plane, standing for "ring" and "spoke", respectively, the mapping to be performed in order to set $r \in [0, E[$ and $s \in [0, A[$ is given by :

$$\begin{cases} r &= \log_k \rho - \log_k \rho_{min} \\ s &= \frac{A\theta}{2\pi} \end{cases} \tag{4}$$

In this paper special attention is given to the main parameters that define the geometry of the log-polar sensor, which are: the number of eccentricities, E, the number of cells per eccentricity, A, the logarithmic factor, k, and the innermost radius, ρ_{min}. The performance of the vergence algorithms will be analyzed as a function of these parameters. While only the parameter A affects discretization in the η direction, all the other parameters affect the ξ direction. Moreover, the following expressions, derived from equation (3), show that they are not independent:

$$\begin{aligned} k &= (\tfrac{\rho_{max}}{\rho_{min}})^{\frac{1}{E}} \\ E &= \log_k(\tfrac{\rho_{max}}{\rho_{min}}) \\ \rho_{min} &= \tfrac{\rho_{max}}{k^E} \end{aligned} \tag{5}$$

As referred in [10], adequate "default" values can be determined for ρ_{min} and k parameters :

$$\begin{aligned} \rho_{min} &= \tfrac{1}{2\tan\frac{\pi}{A}} \\ k &= 1 + \tfrac{2\pi}{A} \end{aligned} \tag{6}$$

These values are obtained by using a circular approximation for the shape of the log-polar cells, and assuming that the resolution in the fovea of the log-polar image is similar to the resolution of the original cartesian image.

One procedure to define a log-polar sensor geometry consists in setting the number of cells per eccentricity, based on the image size, and then defining at most two of the parameters ρ_{min}, E and k. Alternatively, one can also use the default values described in equation 6.

As an illustration of the log-polar mapping, Figure 2 shows the mapping of a cartesian image with constant ρ_{min} and different values for the resolution factor, k. This figure shows the main advantages of the log-polar mapping. On one hand there is an important data reduction while the information content of the image is still sufficient to perceive its structure. On the other hand, we have high resolution at the fovea, whereas the low resolution peripheral vision still allows us to cover a large field of view.

3 Vergence using correlation

When following a target, the motion of the human eyes can be decomposed into saccadic, smooth pursuit and vergence movements [11]. Saccadic and smooth pursuit movements use image position and velocity errors, in order to compensate lateral motion of the target. Vergence movements, instead, are based on disparity cues to compensate motion in depth.

Vergence is one of the basic eye movements to focus visual attention on a given scene point. When observing a target, moving away along the same gaze direction, our eyes rotate by equal amounts in opposite directions. The purpose of the vergence movements is to control the angle between the optic axes, in such a way that a fixation point along some specified gaze direction is kept at the center of the visual field. When this is the case, we say that the target is "foveated". Figure 3 illustrates the vergence angle ν, in a binocular system.

The vergence process enables foveated animals to register objects of interest on the fovea of each eye. As a consequence, it allows the extraction of the greatest possible amount of information. Moreover, when verging on a target, it is possible to find its retinotopic location through disparity filtering, thus allowing to isolate the target from the background [5].

As the vergence angle is directly related to the target distance, depth cues can be used to help the vergence system in compensating depth changes. Depth cues such as binocular disparity and accommodation are important inputs to the vergence control system, and humans rely also on other cues such as motion and shading [3].

Figure 2: The original cartesian image, shown on top, is mapped into log-polar representations (in the center row) with different resolutions, E=64, E=32 and E=16, from left to right. In the bottom row are shown the images remapped back to the cartesian plane, where it is possible to notice the effect of the different resolutions.

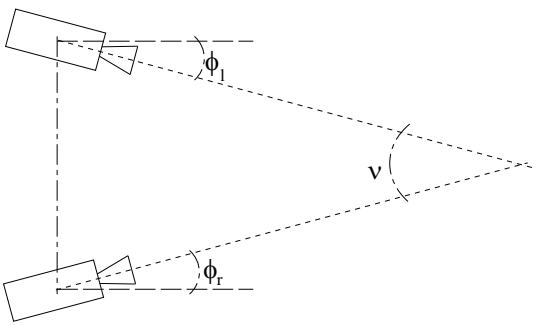


Figure 3: Stereo cameras verging at a point. The vergence angle, $\nu = \phi_l - \phi_r$, is shown.

Traditional algorithms to calculate explicit disparity measures include cepstral filtering and phase correlation, but these techniques are very time consuming, and therefore, not very adequate for real-time applications. By relying on more active visual processes, it is possible to avoid the explicit computation of disparity and still detect correct vergence configurations.

The basic strategies to control vergence eye movements are still under discussion in the biological literature. Hypotheses include continuous feedback control [11], sequential preprogrammed motor responses [14], and channel-like processing of disparity [9].

In the approach used here, we test and compare two algorithms for vergence control. Both of them rely on closed loop control and on simple correlation measures of log-polar images. The first and simpler algorithm consists basically in estimating, over time, a similarity measure between two stereo images and controlling the vergence angle in order to increase their overlap, in a fashion similar to [8]. The second algorithm consists in a set of channels tuned to different disparities. A different channel is activated depending on the instantaneous level of disparity and, as disparity changes, control switches to other channel, resulting in a sequence of pre-programmed responses. This approach follows, in some sense, the ideas of [14] and [9].

As both algorithms use correlation measures, we start by analyzing the differences between the correlation of images of a stereo pair in the vicinity of adequate vergence situations, on the cartesian and log-polar domains.

3.1 Cartesian images versus log-polar images

Suppose that a stereo pair is composed by one image $I(x, y)$ and a translated version $I(x + \delta_x, y)$. For a small horizontal disparity, δ_x, we can write :

$$I\left(x + \delta_x, y\right) = I\left(x, y\right) + \frac{\partial I}{\partial x} \cdot \delta_x \tag{7}$$

As a measure of similarity we use, in this analysis, the integral of the squared difference between the two images. Considering cartesian coordinates, we have:

$$c_c = \iint_{x,y} \left[I\left(x + \delta_x, y\right) - I\left(x, y\right)\right]^2 dxdy \tag{8}$$

Replacing (7) in (8), it yields :

$$c_c = \delta_x^2 \cdot \iint_{x,y} \left(\frac{\partial I}{\partial x}\right)^2 dxdy \tag{9}$$

To compare the correlation of cartesian images with the correlation of log-polar images, we need to change the coordinates of equation (9) from the cartesian to the log-polar domain. We have:

$$\frac{\partial I}{\partial x} = \frac{\partial I}{\partial \xi} \cdot \frac{\partial \xi}{\partial x} + \frac{\partial I}{\partial \eta} \cdot \frac{\partial \eta}{\partial x} \tag{10}$$

where the partial derivatives of the log-polar mapping are:

$$\begin{cases} \frac{\partial \xi}{\partial x} = \frac{\cos\theta}{\rho \cdot \log k} \\ \\ \frac{\partial \eta}{\partial x} = -\frac{\sin\theta}{\rho} \end{cases} \tag{11}$$

The determinant of J, the jacobian of the log-polar to cartesian mapping, is given by :

$$|J| = \log k \cdot \rho^2 \tag{12}$$

and therefore equation (9) can be rewritten in log-polar coordinates as:

$$c_c = \delta_x^2 \log k \iint_{\xi,\eta} \left(\frac{\partial I}{\partial \xi} \frac{\cos\theta}{\log k} - \frac{\partial I}{\partial \eta} \sin\theta\right)^2 d\xi d\eta \tag{13}$$

Similar calculations can be done for the correlation in the log-polar domain. For small deviations from vergence we have:

$$I\left(\xi+\delta_\xi,\eta+\delta_\eta\right)=I\left(\xi,\eta\right)+\frac{\partial I}{\partial\xi}\cdot\delta_\xi+\frac{\partial I}{\partial\eta}\cdot\delta_\eta$$

and the correlation measure can be expressed as:

$$c_l=\iint_{\xi,\eta}\left(\frac{\partial I}{\partial\xi}\cdot\delta_\xi+\frac{\partial I}{\partial\eta}\cdot\delta_\eta\right)^2 d\xi d\eta \tag{14}$$

Using (11), the increments δ_ξ and δ_η can be written as a function of the horizontal disparity δ_x, as :

$$\begin{cases}\delta_\xi=\frac{\cos\theta}{\rho\cdot\log k}\cdot\delta_x\\ \delta_\eta=-\frac{\sin\theta}{\rho}\cdot\delta_x\end{cases} \tag{15}$$

Hence the correlation function of the log-polar images is given by:

$$c_l=\delta_x^2\iint_{\xi,\eta}\frac{1}{\rho^2}\left(\frac{\partial I}{\partial\xi}\frac{\cos\theta}{\log k}-\frac{\partial I}{\partial\eta}\sin\theta\right)^2 d\xi d\eta \tag{16}$$

Observing carefully the expressions for the correlation measures in the cartesian and log-polar domains (equations (13) and (16)), we can see that the main difference is the additional factor of $1/\rho^2$ in the integrand of c_l. Thus, in the log-polar images, the importance given to the disparity of peripheral image patches is smaller when compared to more central parts of the image.

For vergence control, this is an important advantage. Suppose that the cameras are verging on an object, in the center of the visual field. Then, the area under vergence contains low disparities, while the area in the periphery belongs to the background and exhibits large disparities. In this case we would like to have low values for the correlation function. In cartesian coordinates, however, as all disparities are equally considered, the disparities of the background may significantly increase the values of the correlation, as opposed to the log-polar case, where the contribution of the background disparities is reduced.

For log-polar images, the central area which represents the focus of attention can be reduced by increasing the design parameter k, thus allowing to keep effective vergence on smaller targets.

The discrete version of the correlation function we used in the previous analysis is known as the Sum of Squared Differences (SSD):

$$SSD=\sum_{x,y}[\,I_l(x,y)-I_r(x,y)\,]^2,$$

where $I_l(x,y)$, $I_r(x,y)$ stand for the left and right images.

However, the SSD norm is very sensitive to illumination changes between left and right images, which often occur if the two cameras have different apertures. On the contrary, the zero-mean normalized cross-correlation ($ZNCC$) behaves robustly [6] relative to changes in the illumination conditions :

$$ZNCC=\frac{\sum(I_l-\mu_l)(I_r-\mu_r)}{\sqrt{\sum(I_l-\mu_l)^2\sum(I_r-\mu_r)^2}}$$

where μ_l and μ_r are the average gray levels for the left and right images, respectively. For the remaining of this paper, the following correlation function used is:

$$ZNCC^-=1-ZNCC$$

and highly correlated images will present low values of this function.

We have performed experiments to illustrate the advantages of the log-polar geometry over the cartesian images, for the vergence behavior. An object was placed in front of the camera rig and the cameras moved to

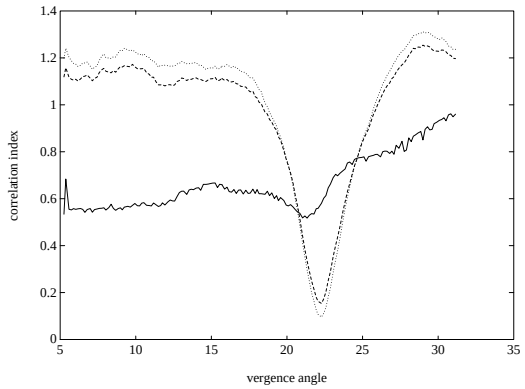


Figure 4: $\mathcal{C}(\nu)$: correlation index as a function of the vergence angle. Solid line: $\mathcal{C}(\nu)$ for a cartesian pair of images. Dashed and dotted lines: $\mathcal{C}(\nu)$ for log-polar images with different layouts. The minimum values represent the correct vergence angle.

sweep the entire vergence space. The cartesian images size is 128x128 pixels whereas the log-polar images size is 64x32 pixels, thus reducing the amount of data to be processed by a factor of 8. Figure 4 shows a plot of $\mathcal{C}(\nu)$, the $ZNCC^-$ index, as a function of the vergence angle, for both the cartesian and log-polar images.

This experiment shows that for the cartesian images, the correlation index exhibits various local minima and a nonuniform behavior. On the other hand, for the log-polar images, 8 times smaller, the curve is much smoother and presents a sharp minimum in the neighborhood of the correct vergence angle.

In the experiments made with $ZNCC^-$ of log-polar images, all correct vergence configurations were associated to local minima in $\mathcal{C}(\nu)$, except for some very cluttered environments. With cartesian images errors occurred more often.

4 Control algorithm

In this section we propose and compare two strategies for vergence control. Both of them use correlation measures from log-polar images to drive the cameras motion, using a standard feedback control architecture, as represented in Figure 5.

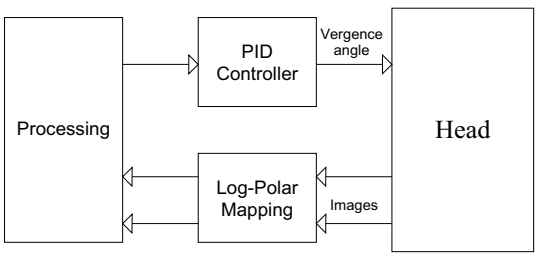


Figure 5: Vergence loop control diagram.

Both vergence control mechanisms are based only on correlation clues acquired continuously over time. As in [12] vision is used as a purposive sensor, providing feedback signals to close control loops and no calibration information is needed in order to accomplish the desired tasks. The difference among the two strategies remains on the processing step, as will be described in the next sub-sections.

4.1 Iterative vergence control

The first algorithm is based on the minimization of the correlation of the log-polar stereo pair, calculated as the $ZNCC^-$ between the two images. As we have seen previously, the correlation function for log-polar images exhibits a distinctive minimum, when an object is appropriately verged.

The $ZNCC^-$ calculation provides a non-negative correlation index, $\mathcal{C}_i$. However, this signal cannot be directly used by a PID controller, since is does not convey any information about whether to verge or diverge the cameras (error sign) and, even when an object is under correct vergence, the error may be non-zero and give origin to a non-stationary situation.

The control goal is to minimize the correlation index, hence indicating the most likely vergence configuration. Usual gradient descent techniques require the calculation of the gradient of the function $\mathcal{C}(\nu)$ where ν is the vergence angle. The incremental update ($\delta\nu_i$) would be given by:

$$\delta\nu_i = -\alpha\nabla\mathcal{C}(\nu_i) \tag{17}$$

where a discrete approximation of the gradient can be obtained by :

$$\nabla\mathcal{C}(\nu_i) \approx \frac{\Delta\mathcal{C}_i}{\Delta\nu_i}.$$

In practice, the determination of the gradient is very sensitive to noise. In the present case, both motor control and image acquisition processes introduce noise that degrades significantly the estimates of $\mathcal{C}(\nu_i)$ and $\Delta\nu_i$.

In [8], another algorithm was proposed to calculate the control reference, by using the inverse gradient of the correlation, and the correlation value itself, as a normalization factor to smooth the noise coming out from the gradient :

$$\delta\nu_i = -\frac{\mathcal{C}_i}{\nabla\mathcal{C}(\nu_i)}$$

Similarly to the gradient descent, we have found this algorithm very unstable and potentially presents a singularity in the minimum ($\nabla\mathcal{C}(\nu)$ becomes zero), which poses problems to the control.

Most of the instability of these algorithms is due to the noise in the calculation of the gradient of the correlation index. Our approach avoids relying on values of any differential measure to determine an appropriate signal to control. For the experiments made, the designed algorithm relies only on the value of the correlation index, the sign of its first difference which indicates the increase or decrease of $\mathcal{C}(\nu)$ and the sign of the previous reference value:

$$s_i = \mathcal{C}_i \cdot [-sgn(\Delta\mathcal{C}_i)] \cdot sgn(s_{i-1}) \tag{18}$$

As $\mathcal{C}_i$ is always non negative, if we are moving in the right direction, then $\Delta\mathcal{C}_i$ is a negative value and the direction of vergence is maintained. In the other way, if we are moving in the wrong direction, the control reference inverts the direction of motion. The factor $\mathcal{C}_i$ respects the intuitive fact that when far from the minimum, the reference value should be high and when $\mathcal{C}_i$ approaches the minimum, the reference should be low.

This control strategy is less sensitive to additive noise than the gradient descent. Figure 6 refers to a simulation of both methods, supposing that the correlation function is a parabola with uniform additive noise. It shows the correlation function and the evolution in time of the vergence angle for the control laws described by equations 17 and 18.

4.2 Preprogrammed vergence control

Several theories exist for vergence eye movement control. In [14] the theory of "dual mode vergence" eye movement is sustained. In this theory, the initial portion of the movement is composed by fast relocations of the eyes, followed by slower movements to attain accurate bifixation. Their experiments with human individuals have shown that the first mode consists in a step-like behavior, as the product of sequential preprogrammed responses which are triggered when retinas disparity becomes sufficiently large. Additionally [9] proposed a number of discrete "disparity" channels, which would be activated depending on the disparity level. Each channel, once activated, would trigger the control process to produce an additional preprogrammed movement.

Like in [9], our second control strategy is based on a set of channels corresponding to a number of (horizontal) disparity values. The horizontal disparity field in the cartesian images gives origin to complex disparity patterns in log-polar images. These patterns are calculated off-line and stored in look-up tables.

Therefore, each channel corresponds to an hypothesized disparity between the two log-polar images. In run time, the output of each channel is given by the $ZNCC^-$ between the images, where the image patches are locally shifted by the disparity considered.

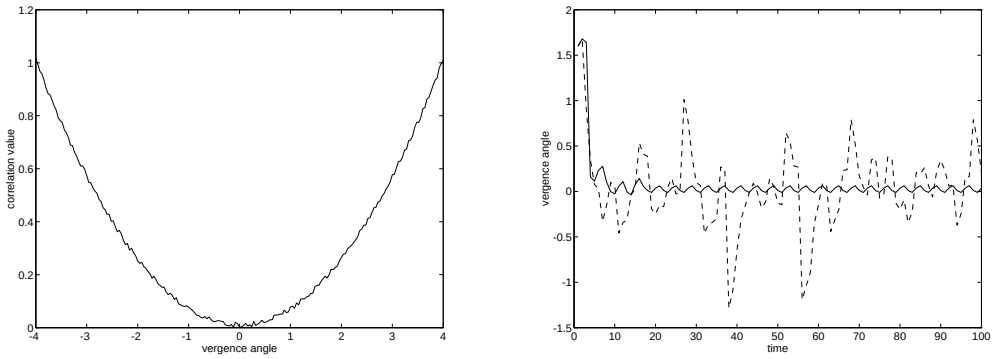


Figure 6: Performance of different control laws. Left: Parabolic correlation function with additive noise. Right: System response with proposed control law (solid line) and gradient descent (dashed line).

In run time, the $ZNCC^-$ between the pixels in both images whose correspondence is given by the channel look-up tables. A global minimization for the several disparity channels is done, and the one with the minimum value provides an estimate for the disparity of the current images.

To compare the results of applying this approach both to cartesian and log-polar images, we have acquired a sequence of stereo images, with a central object with various disparities. Since the object disparity is known, we can compare the ground truth value with the output of the different disparity channels.

Figure 7 shows the correlation surface obtained by computing the $ZNCC^-$ for the various disparity channels, and the various target disparities. In this experiment we have considered channels with disparities of $\{ 0, \pm1, \pm2, \pm3, \pm5, \pm8, \pm12, \pm17, \pm23 \}$ pixels.

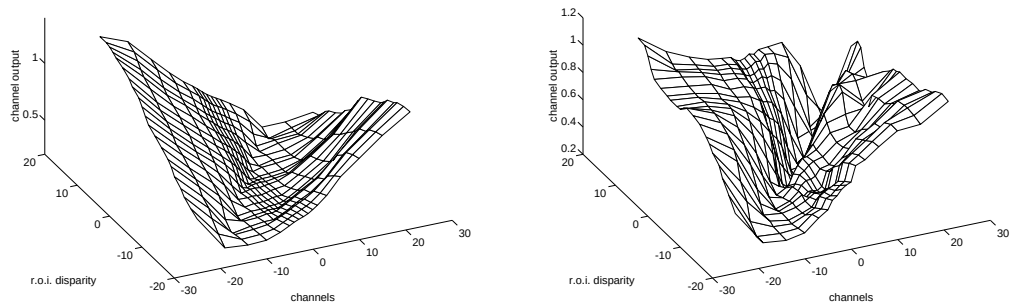


Figure 7: Calculation of the $ZNCC^-$ in several disparity channels for cartesian (left) and log-polar (right) images

Figure 7 shows that in general, the selected channel corresponds to the target disparity, both for cartesian and log-polar images. However, for the log-polar images, the correlation surface is sharper in the neighborhood of the true disparity, whereas it is almost flat for the cartesian images. Therefore the selection of the optimal channel is easier in the log-polar images, when compared to cartesian images, where erroneous disparity estimates are more likely.

In [9] it is argued that, for control purposes, each channel has a low pass filter characteristic with different gains and time constants. Our approach, instead, follows a simpler strategy: the value of the disparity of the activated channel is fed into the PID controller so that each channel inherits the dynamic characteristics of the controller.

5 Performance characterization of the correlation function

In this paper we are interested in evaluating the influence of the main design parameters of the sensor layout on the performance of the vergence behavior. The performance of this behavior obviously depends on:

- Characteristics of the visual measurements obtained from the environment.
- How are these measurements used for control purposes.

In the previous section we have discussed the vergence control problem and introduced two control strategies that use the correlation of log-polar images as the input. In this section we focus on the characterization of the ideal shape of the observation (correlation) function, for the proposed control algorithms. We discuss qualitatively how changes in this function, influence the vergence control performance. Then, in Section VI, the overall system performance will be evaluated according to the influence of the sensor geometry on the shape of the correlation function.

As stated before, the correlation of log-polar images as a function of the vergence angle, $\mathcal{C}(\nu)$, is a locally convex function with a minimum for the correct vergence configuration. Its width, steepness and minimum value are characteristics that influence the behavior of the vergence control algorithm, in the following way :

Width - The width of the region around the minimum, without other local minima, establishes the maximum displacement of the object for which the control algorithm can still detect the local minimum. It should be as large as possible to allow verging on targets with significant disparities and to maintain vergence in objects with important depth displacements between sampling cycles.

Steepness - The steepness near the minimum determines the sensitivity of the algorithm to noise. If the correlation function is very steep in a neighborhood of the minimum, then the amplitude of noise or disturbances needed to break monotonicity must be very large. Hence, it influences the detectability of the target disparity.

Minimum - The minimum value qualitatively defines the confidence on the disparity estimate, since it measures the similarity of the images for the best-detected vergence angle. Therefore, it is expected to be as small as possible, once the cameras are appropriately verging on an object.

The particular shape of $\mathcal{C}(\nu)$ around the minimum, depends on the log-polar sensor geometry. The performance analysis of the various design parameters is extensively addressed in the experimental tests, and will be discussed in detail.

6 Results

The vergence system described in the previous sections was implemented on the **Medusa** Stereo Head, shown in Figure 8. The head has four mechanical degrees of freedom, including independent vergence for both cameras,

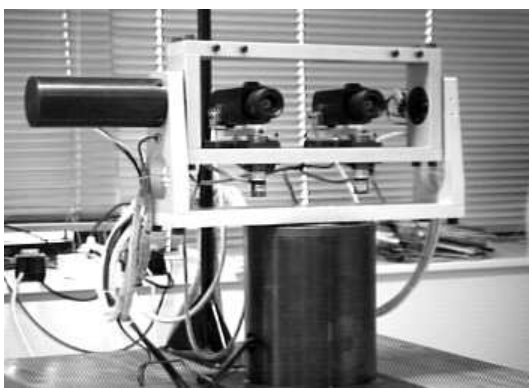


Figure 8: Medusa stereo head used for the experimental tests.

and a common tilt and pan axes.

All the computations, including image mapping and processing, are done on a PENTIUM PC computer running at 133MHz. The mapping from a pair of 128x128 cartesian images to 64x32 log-polar images takes about 10 milliseconds while the $ZNCC^-$ calculation takes only 0.6 milliseconds. The sampling rate currently achieved is about 6 Hz, which takes into account the data transfer time from the image frame grabber to host computer memory. Using fast transfer through PCI buses it will be possible to run at video rates.

Several tests were made in order to evaluate the two main aspects addressed in this paper. The first one is related to the geometry of the log-polar sampling mechanism. We have performed various experimental tests in order to assess the influence of the main design parameters of the log-polar sensor. The second set of experiments illustrates the performance of the closed loop vergence control strategies.

6.1 Influence of the log-polar parameters

Various authors have used log-polar image representation and have pointed out many of its advantages [15]. However, it is seldom clear what parameters of the log-polar representation to use, and what is their influence on the expected performance of the algorithms.

Here, we will discuss how parameter changes in the log-polar sensor affect the correlation index, according to the desired qualitative properties of $\mathcal{C}(\nu)$ defined before.

The original cartesian images are obtained with a 512×512 resolution but the designed log-polar sensors are allowed to cover different parts of this space, simulating different sizes for the field of view. Experiments were made in a diversity of situations and sensor layouts. For a given scene, the experiments consisted in sweeping the vergence angle space and computing the correlation index, for the various log-polar geometries. Through the analysis of the shape of the correlation index curve, one can evaluate the effects of changing a given design parameter.

The results shown were obtained in two static scenes. In the first situation the target is placed far away from the cameras (about 4m), covering a small area of the visual field. In the second situation, the target is near the cameras (about 1,5m) where a larger range of disparities is present. Figure 9 shows the two situations, as seen by one of the cameras of the stereo head.

Figure 9: Experiments were made with the target either far from the cameras (left) or near the cameras (right).

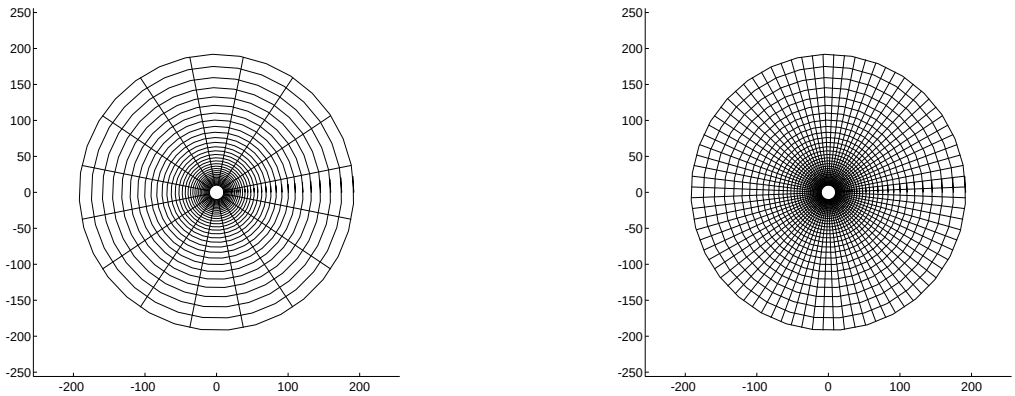
In a first set of experiments, a) and b), we will focus on sensor layouts with different resolutions and singularity regions, which produce log-polar images of different size but that cover the same field of view. In the second set of experiments, c) , the sensor discretization steps are kept constant but different fields of view are covered.

For each experiment, we will show two layouts illustrating the effect of changing in the tested parameter in the sensor geometry. A plot of the correlation value as a function of the vergence angle is shown for several sensors and considering targets at two different distances from the cameras. A table is also presented with the main design parameters of the sensor for each curve in the plots. The table contains the number of cells per eccentricity (A), the number of eccentricities (E), the value of the minimum radius (ρ_{min}), the value of the logarithmic factor *(k)* and the diameter of the sensor (ϕ), measured in pixels.

a) Change in the sensor resolution

The first set of experiments consists in changing the sensor resolution by subdividing the log-polar space with different step sizes. This is done both for the angular and radial directions and does not lead to changes in the size of the sensor.

Figure 10 shows the corresponding sensor layouts and the results obtained for changes of resolution in the angular direction. Figure 11 shows the same information when changes in the radial resolution are considered.



	A	E	ρ_{min}	k	ϕ
solid line	32	64	10.0	1.0473	384
dashed line	64	64	10.0	1.0473	384
dotted line	96	64	10.0	1.0473	384

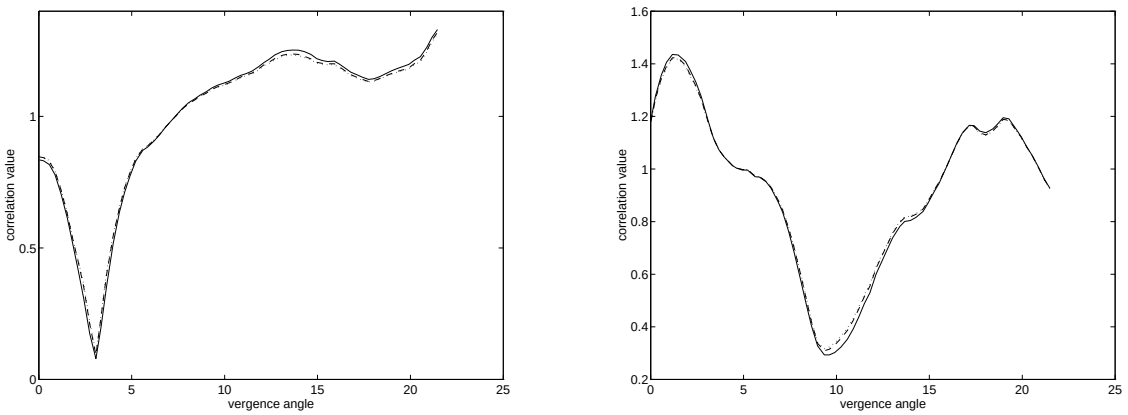
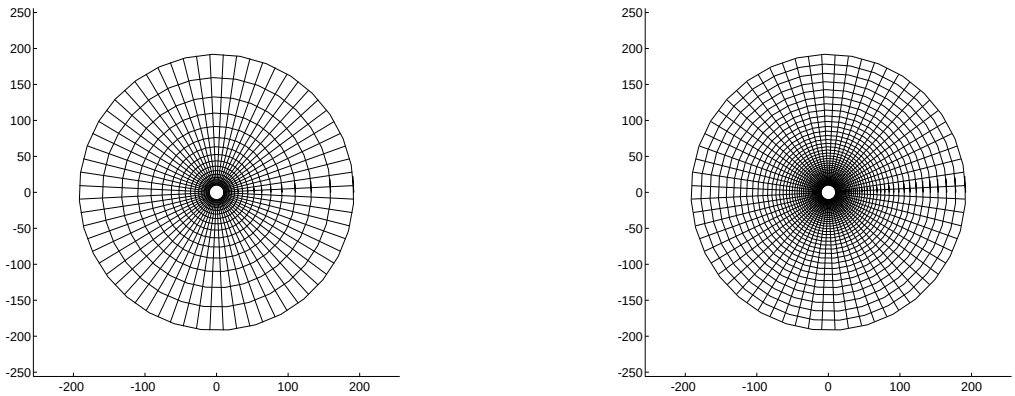


Figure 10: Sensor geometry (top) and plots of $\mathcal{C}(\nu)$ for different angular resolutions, when the target is far from the cameras (bottom left) or near the cameras (bottom right).



	A	E	ρ_{min}	k	ϕ
solid line	128	32	10.0	1.0967	384
dashed line	128	48	10.0	1.0635	384
dotted line	128	80	10.0	1.0376	384

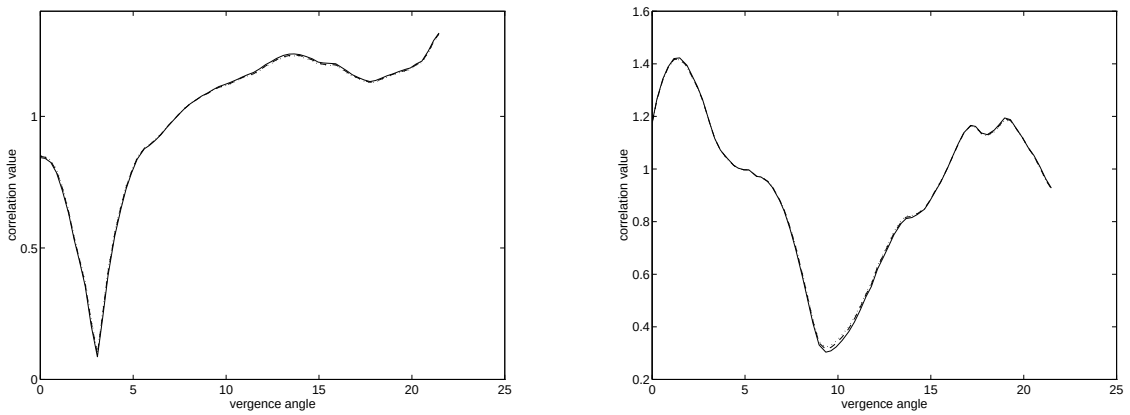


Figure 11: Sensor geometry (top) and plots of $\mathcal{C}(\nu)$ for different radial resolutions when the target is far from the cameras (bottom left) or near the cameras (bottom right).

No major differences were detected in the correlation function, for changes either in the angular or radial resolutions. In fact, the correlation value showed a very low sensitivity to changes in the resolution of the log-polar images and only for extremely low resolution images (both angular and radial), the correlation function exhibited important differences. As can be observed in Figure 12, for the closer target, an incorrect vergence angle is detected for the lowest resolution sensor.

	A	E	ρ_{min}	k	ϕ
solid line	128	64	10.0	1.0473	384
dashed line	8	8	10.0	1.4468	384

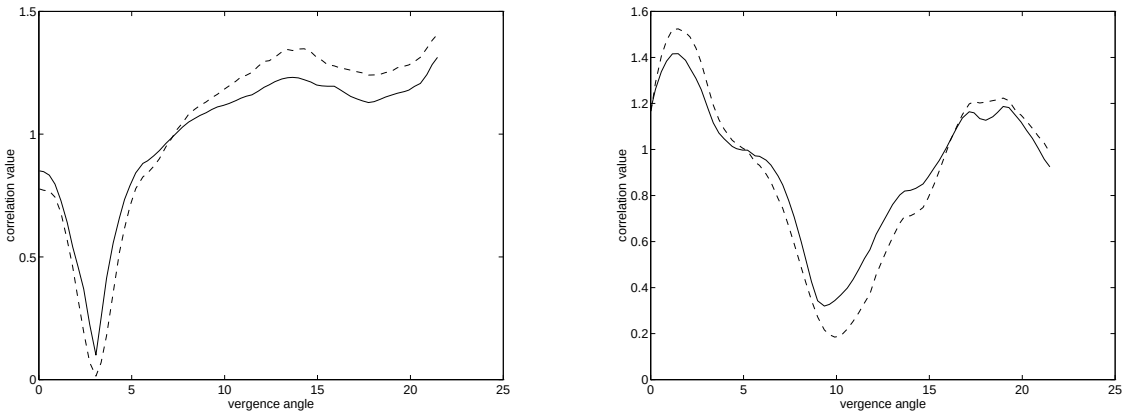


Figure 12: Effect of very reduced overall resolution, when the target is far from the cameras (left) or near the cameras (right).

This insensitivity of the correlation function relative to the image resolution can be explained by the particular image measurements we use, and the nature of the scene. The computation of the correlation measure is based on the accumulation of differences in image patches. For images with small local variations, the correlation value between two patches is similar to the difference of the average gray levels of the patches. Therefore changes in resolution are not expected to produce significant changes in the correlation value. However, for very textured and highly contrasted images, changes in resolution may have an important impact on the performance of the algorithms.

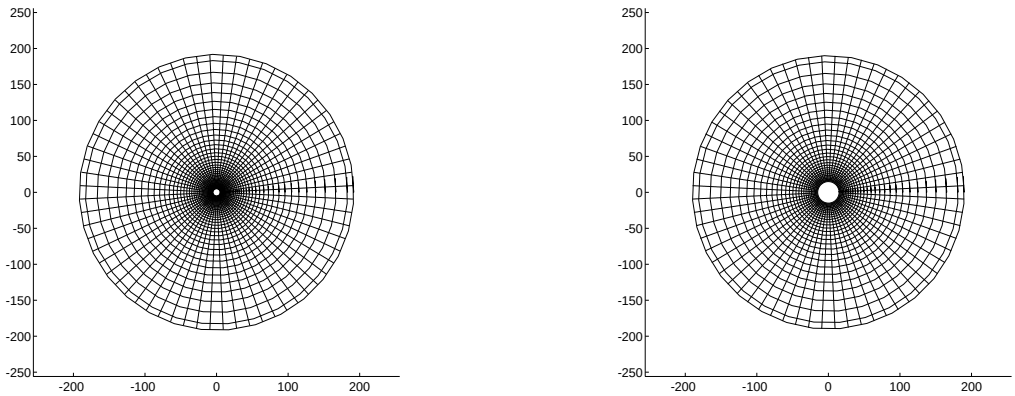
b) Change in the minimum radius

In this experiment, sensors with different values of ρ_{min} are used while the remaining parameters are kept constant. The effect of this change in the sensor geometry, can be interpreted as inserting or removing eccentricities in the inner part of the sensor, thus changing the size of the log-polar images. The addition of eccentricities can increase the number of redundant pixels in the fovea. For one of the sensors used in this experiment, with a minimum radius of 15 pixels, the number of redundant pixels represents 1.08% of the total. For the remaining sensors, this value is zero. Results are shown in Figure 13.

The results reveal that large values of the minimum radius produce a worse profile in the correlation function, according to our quality criteria, and therefore, ρ_{min} should be low. However, very small values of ρ_{min} can increase significantly the amount of redundant pixels in the image.

c) Scaling of the visual field

This set of experiments evaluate changes in the design parameters that do not modify the size of the log-polar images but originate sensors covering different visual fields.



	A	E	ρ_{min}	k	ϕ
solid line	128	55	15.0	1.0473	384
dashed line	128	44	35.0	1.0473	384
dotted line	128	31	45.0	1.0473	384

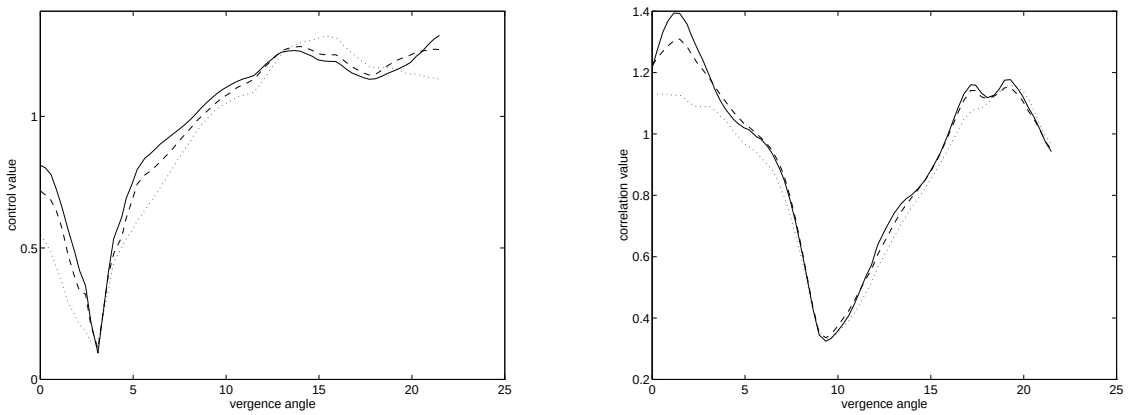
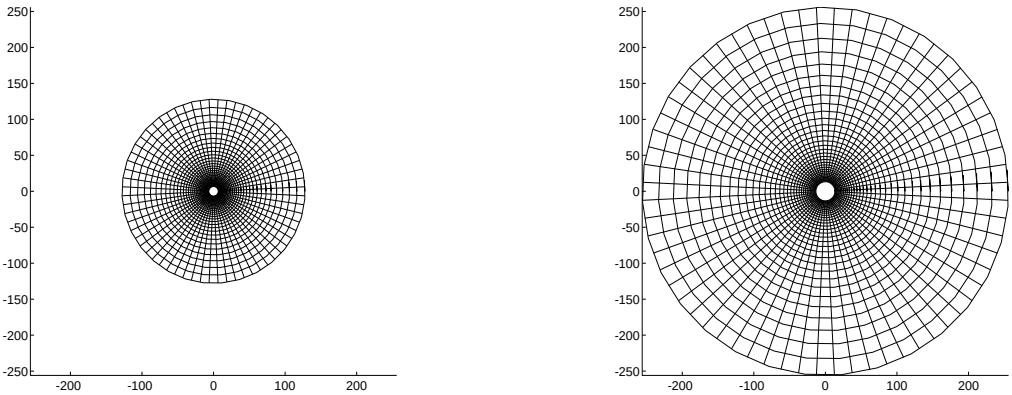


Figure 13: Sensor geometry (top) and plots of $\mathcal{C}(\nu)$ for different values of the minimum radius, when the target is far from the cameras (bottom left) or near the cameras (bottom right).

In this first case, an uniform scaling of sensor cells is produced, which means that the radial size of the cells, in different sensors, is scaled by a constant factor. The size of the minimum radius and the diameter of the sensor are also scaled by the same amount. Results for this case are presented in Figure 14.



	A	E	ρ_{min}	k	ϕ
solid line	128	64	6.667	1.0473	256
dashed line	128	64	10.0	1.0473	384
dotted line	128	64	13.333	1.0473	512

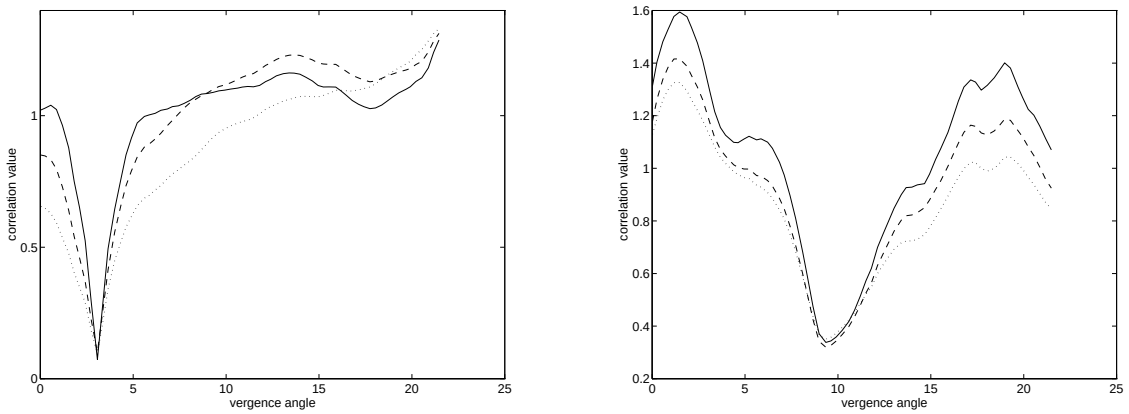
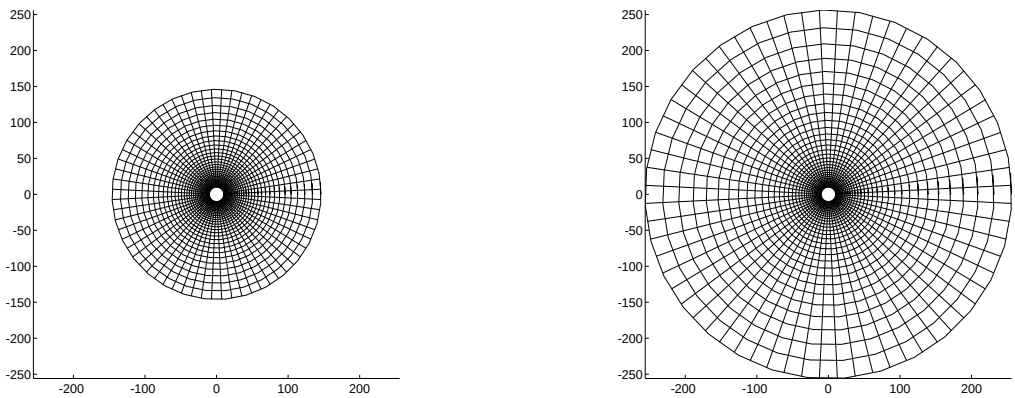


Figure 14: Sensor geometry (top) and plots of $\mathcal{C}(\nu)$ for uniform scaling of the visual field, when the target is far from the cameras (bottom left) or near the cameras (bottom right).

In the second case, another kind of scaling for the visual field is performed. While in the first situation, the rate of growth of the receptive cells was always constant (k was constant), now we will consider the effect of changing the rate of growth. The diameter of the sensors increases with k^E while the minimum radius remains constant. The sensor layouts and plots of the correlation function are shown in Figure 15.

Taking into account both target positions (near or far), we can observe that for smaller visual fields, the correlation function has deeper profiles. As the object of interest is located in the center of the images, the concentration of the information in this area enables the vergence system to perform more reliably. However, reduced visual fields do not allow the detection of objects of interest with large displacements. The plots for large sensors exhibit wider correlation functions, specially when the target is far from the cameras.



	A	E	ρ_{min}	k	ϕ
solid line	128	64	10.0	1.0428	292
dashed line	128	64	10.0	1.0473	384
dotted line	128	64	10.0	1.0520	512

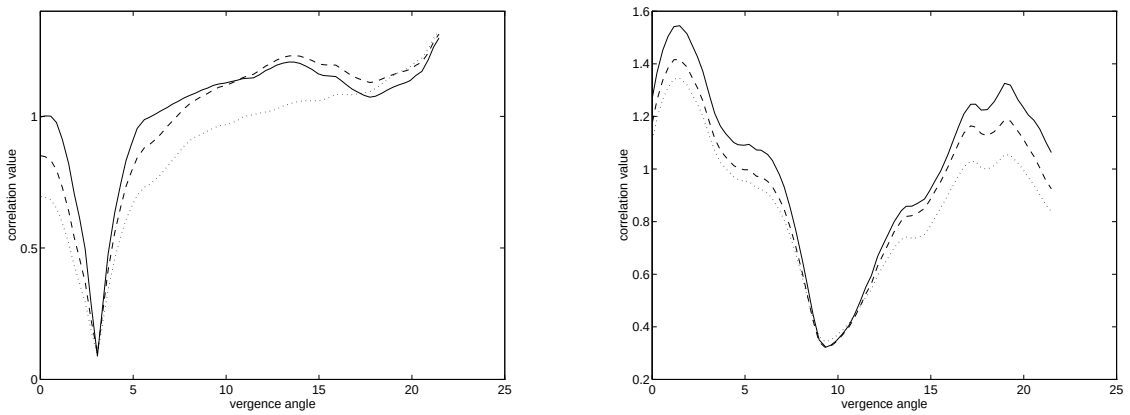


Figure 15: Sensor geometry (top) and plots of $\mathcal{C}(\nu)$ for non-uniform scaling of the visual field when the target is far from the cameras (bottom left) or near the cameras (bottom right).

Discussion

From the experiments performed in this subsection we can extract three important points:

- First we should notice the low sensitivity of the correlation function to changes in the resolution of the log-polar images, allowing the use of extremely small images, hence reducing the image processing time.
- The second point to notice regards the influence of the value of the inner radius. Overall loss of quality in the correlation function was noticed with the increase of this parameter. Of course, as we want to verge on the object roughly located in the center of the retinas, a large blind area in the center of the field of view is very inconvenient. In order to avoid this problem low-values for the minimum radius should be chosen[1] or we should fill the fovea with a pattern other than log-polar[2][17].
- Finally, experiments **c)** and **d)** show that, small fields of view are more sensitive to variations in the vergence angle, thus allowing to have faster convergence rates and more precise vergence. However, large fields of view can produce wider profiles in the correlation function which allow to keep vergence under disturbances.

6.2 Controlled Vergence

We have performed various experiments to test the proposed closed loop vergence control systems. A variety of objects, located at different distances from the stereo head, have been used. In the majority of the situations both approaches were able to verge on the targets located in the center of the images, once the initial disparity was kept within appropriate limits.

In order to measure the system response to a vergence input, we have used a sequence of step stimuli. To start, we have placed an object at a given distance from the cameras. The object was then suddenly moved to different positions after small time periods. This experiment corresponds to a sequence of steps functions applied to the input of the vergence system. Finally, the object was removed from the scene and the system verged on the background.

The closed loop controller used for the two proposed algorithms was a simple proportional controller. Figure 16 shows the time evolution of the vergence angle. The time scale units are sampling periods (about 160ms).

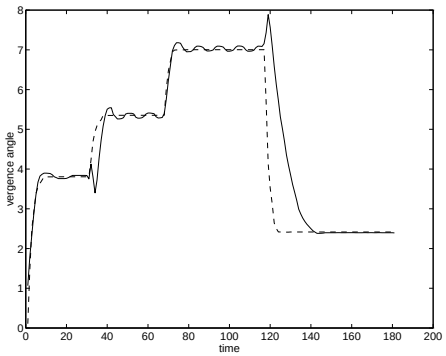


Figure 16: Closed loop response of the vergence control system to step-like stimuli. Solid plot: iterative vergence control. Dashed plot: preprogrammed vergence control.

The main point to notice is the oscillation exhibited by the iterative vergence control when verging in the target. The reason is that this algorithm relies on the sign of the temporal difference of the correlation index. To obtain this signal, the cameras must move continuously, even when the desired vergence angle is achieved. This fact constrains the value of the controller gain, thus limiting the raising time of the response. By increasing the controller gains, the amplitude of the oscillations would be larger.

[1] Very low values for the minimum radius increase redundant information in the foveal area. The amount of redundant information should be kept low, for efficiency purposes.

[2] We could use, for example, the original cartesian pixels.

The preprogrammed vergence control shows a better performance. The correct vergence angles are reached after a short transient period, and in a fashion similar to a first order dynamic system. Furthermore, we noticed that the range of initial disparities for which the system converges is larger for the preprogrammed vergence control. This depends, obviously, in the range of the disparity channels used which is easy to define, while in the iterative vergence control the dependency arises on the size of the region of convergence of the correlation function, which can be as small as 5 degrees.

Figure 17 shows the evolution in time of the control signals.

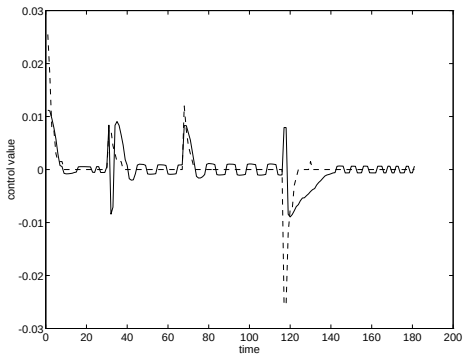


Figure 17: Control value for iterative (solid line) and preprogrammed (dashed line) control modes.

Again, oscillations in the iterative control value can be observed. The control signal for this mode (equation 18), depends on the correlation index, which is a function of the overall mismatch of the images. Therefore, when the system is verging on the background, oscillations have smaller amplitudes due to the lower correlation values (see Figure 18).

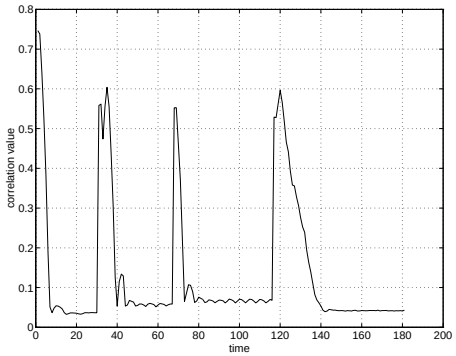


Figure 18: Correlation value for iterative control mode.

Because the controller used for the preprogrammed vergence response is simply proportional, the control signal for this case is equal to the disparity channel selected at each instant, apart from a scale factor (compare with Figure 19). The disparity estimate is used as a direct input to the controller with no further processing, which is not the case of the correlation index in the iterative control mode.

7 Conclusions and further work

In this paper we have shown that vergence control can be achieved using simple correlation measures over log-polar images in a closed loop control architecture.

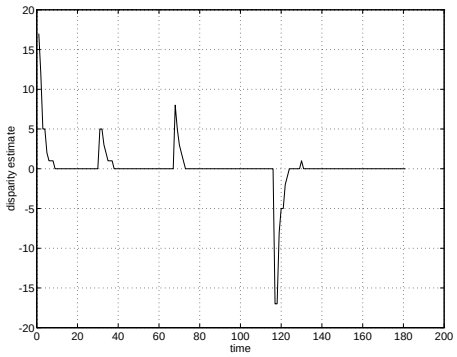


Figure 19: Disparity estimate for preprogrammed control mode.

In this work biological analogies can be established in two main aspects : First, a non-uniform sampling mechanism such as the log-polar mapping is based on the topology of the human retina. Second, the vergence control strategies proposed in the paper have strong similarities with the human vergence system.

Log-polar images have high resolution at the fovea, and decreasing resolution towards the periphery. Besides the important data reduction achieved, the log-polar images are more suited for vergence control when compared to the cartesian images.

Two different control strategies based on the correlation of two log-polar images are presented. The *iterative* approach is based on a hill climbing procedure applied to the correlation index space. Differential measurements are avoided because of noise sensitivity. The *preprogrammed* method uses a set of disparity channels tuned to a number of disparity values. This second method reveals a better performance despite a higher computational cost per iteration.

The system was implemented, in real-time, on the stereo head Medusa. A large number of experiments are described in the paper, using realistic scenes. In a first set of tests, we show how the log-polar sensor parameters affect the performance of the vergence system. These results establish a criterion to choose the log-polar sensor parameters, in order to improve performance. Finally, results with controlled vergence are described, and the control algorithms exhibited good performances. Furthermore, in dynamic situations the system shows fast reaction and settling times.

A relevant aspect of this work is the characterization of the influence of the image sensor design in the performance of the vergence system. Particularly, we establish criteria to assess the quality of the log-polar sensor layout for correct vergence to be achieved, and therefore introduce a procedure to select an adequate sensor design. Additionally, we compare the performance of the two control algorithms proposed.

As a general methodology, the work described here combines, in our view, various important facets for the application of vision algorithms in a large number of active vision problems. First, we consider the choice of a sensor geometry particularly suited to a specific visual task. Then, we characterize and analyze the influence of the sensor design parameters on the overall performance of the system. Finally, the control methodology relies on a continuous stream of visual input to control the cameras motion, thus enjoying of improved reactivity.

Future work should address the problem of integrating the vergence system with the other ocular movements, following a similar approach. A particular effort will be concentrated in the performance characterization of other vision algorithms.

References

[1] Y. Aloimonos, I. Weiss, and A. Banddophaday. Active vision. *Int. Journal of Computer Vision*, 1(4):333–356, January 1988.

[2] R. Bajcsy. Active perception. *Proc. of the IEEE*, 76(8):996–1005, August 1988.

[3] R. Carpenter. *Movements of the eyes.* Pion, London, 1988.

[4] H. Christensen, K. Bowyer, and H. Bunke, editors. *Active Robot Vision : camera heads, model based navigation and reactive control*, volume 6. World Scientific, 1994.

[5] D. Coombs and C. Brown. Real-time binocular smooth pursuit. *Int. Journal of Computer Vision*, 11(2):147–164, October 1993.

[6] O. Faugeras, P. Fua, B. Hotz, R. Ma, L. Robert, M. Thonnat, and Z. Zhang. Quantitative and qualitative comparison of some area and feature based stereo algorithms. In W. Förstner and St. Ruwiedel, editors, *Robust Computer Vision*. Wichmann, 1992.

[7] F. Ferrari, J. Nielsen, P.Questa, and G. Sandini. Space variant sensing for personal communication and remote monitoring. In *Proc. of the EU-HCM Smart Workshop*, Lisbon, Portugal, April 1995.

[8] F. Panerai, C. Capurro, and G. Sandini. Space variant vision for an active camera mount. In *Proc. SPIE AeroSense95*, Florida, USA, April 1995.

[9] M. Pobuda and C. Erkelens. The relation between absolute disparity and ocular vergence. *Biological Cybernetics*, 68:221–228, 1993.

[10] P. Questa. A software simulation environment of a non-linear sampling retina-like ccd sensor. Technical Report 2/92, DIST-University of Genova, 1992.

[11] D. Robinson. The oculomotor control system: A review. *Proc. of the IEEE*, 56(6), June 1968.

[12] J. Santos-Victor, G. Sandini, F. Curotto, and S. Garibaldi. Divergent stereo in autonomous navigation : From bees to robots. *Int. Journal of Computer Vision*, 14(2):159–177, 1995.

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

[14] J. Semmlow, G. Hung, J. Horng, and K. Ciuffreda. Disparity vergence eye movements exhibit preprogrammed motor control. *Vision Reserach*, 34(10):1335–1343, 1994.

[15] M. Tistarelli and G. Sandini. On the advantages of polar and log-polar mapping for direct estimation of the time-to-impact from optical flow. *IEEE Trans. on PAMI*, 15(8):401–411, April 1993.

[16] R. Wallace, P. Ong, B. Bederson, and E. Schwartz. Space variant image processing. *Int. Journal of Computer Vision*, 13(1):71–90, September 1995.

[17] C. Weiman. Log-polar vision for mobile robot navigation. In *Proc. of Electronic Imaging Conference*, pages 382–385, Boston, USA, November 1990.