# Active Vergence Behavior

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**

This paper describes a real-time vergence control mechanism developed for the stereo head MEDUSA [13], in the context of behavior based active control [3]. Real-time control architectures for active vision impose strong constraints on the computational complexity of image processing algorithms.

In many applications, the image size is often drastically reduced in order to achieve fast and reactive routines, at the cost of precision in the output of the algorithms. Non-uniform sampling of the images, through the log-polar mapping, is an efficient way of reducing the amount of data to process [8, 17], while maintaining a good resolution in the center of the field of view. By using active vision devices, such as stereo heads, we are able to inspect a wide field of view with high resolution. Besides the image size reduction feature, the correlation index of log-polar images, as a function of the vergence angle, has major advantages over cartesian images, in the detection of correct vergence configurations [9].

Moreover, in this paper we show that the correlation index can be used for robust vergence control, with very low resolution log-polar images, leading to a very fast visuo-motor process. There is no explicit disparity estimation and the control law is a function of direct image measurements. A performance analysis of the of this method for several log-polar sensors is done, establishing criteria to choose adequate layouts. A real time implementation is described and extensive experiments have been performed.

**Categories: Active and Real-Time Vision, Applications.**

*Alexandre Bernardino is currently supported by a PRAXIS-XXI research program

# Active Vergence Behavior


## Abstract

This paper describes a real-time vergence control mechanism developed for the stereo head MEDUSA [13], in the context of behavior based active control [3]. Real-time control architectures for active vision impose strong constraints on the computational complexity of image processing algorithms.

In many applications, the image size is often drastically reduced in order to achieve fast and reactive routines, at the cost of precision in the output of the algorithms. Non-uniform sampling of the images, through the log-polar mapping, is an efficient way of reducing the amount of data to process [8, 17], while maintaining a good resolution in the center of the field of view. By using active vision devices, such as stereo heads, we are able to inspect a wide field of view with high resolution. Besides the image size reduction feature, the correlation index of log-polar images, as a function of the vergence angle, has major advantages over cartesian images, in the detection of correct vergence configurations [9].

Moreover, in this paper we show that the correlation index can be used for robust vergence control, with very low resolution log-polar images, leading to a very fast visuo-motor process. There is no explicit disparity estimation and the control law is a function of direct image measurements. A performance analysis of the of this method for several log-polar sensors is done, establishing criteria to choose adequate layouts. A real time implementation is described and extensive experiments have been performed.

# Summary

This paper describes a vergence system for stereo heads, using log-polar sampled images. Instead of estimating the disparity, the control law is based on direct information from correlation index evaluating the similarity for two images. The influence of the various parameters in the log-polar sensor have been determined, establishing criteria for the choice of a suitable sensor layout. A real time implementation on a binocular stereo head is described, and experimental results shown.

**Original Contributions** - The main original contributions of the paper are : use of direct image measurements for vergence control without disparity estimation; the control system used to optimize the vergence criterion; extensive experimental analysis of the influence of the various log-polar sensor parameters in the overall system performance. Globally, it emphasizes the usefulness of vision when used in a control loop to perform a visual task.

**Importance of the contribution** - The importance of this work stems from the use of simple and robust algorithms for closed loop vergence control, which does not require a large computational power. Keeping vergence on an object has multiple applications in various domains, as for estimating the distance to target, background and target segmentation, etc.

**Closely related work** - This work is related to other author's work in various aspects : for log-polar image representation [16, 17] have presented the main mathematical aspects which are shown here in a slightly simpler manner; for log-polar images in vergence control [9], our work differs on the control strategy that was used, and presents a thorough experimental analysis of the log-polar sensor parameters and their influence on system performance; for vergence control [6] has used cepstral filtering and most authors use some kind of disparity estimation process, while our control system does not use any explicit disparity estimate.

**Use of the results of this work** - Other researchers can make use of the results of this work in various ways. On one hand we have carefully given all the details of the implementation, thus allowing other researchers to easily implement our method. On the other hand the software can be made available, on request.

**Other submissions** - This work has not been submitted to any conference or journal before.

# 1 Introduction

The paradigm of Active Vision has been a major research issue in the last few years [1, 2]. As a consequence, Stereo Heads have become an important research tool and have been designed and built in many research laboratories (see [5] for an overview). Since the beginning, these systems have been designed with a strong inspiration in their biological analogues [4, 11]. The visuo-motor processes, have been modeled as saccadic, vergence and pursuit behaviors similar to the basic human visual processes that run in parallel and cooperate to accomplish perceptual tasks. Also, the mechanical design of most stereo heads resemble the human oculomotor system, and though not having the whole flexibility of motion, can still perform the basic movements needed to carry out most of the visual tasks.

Not surprisingly, vision sensors are also becoming similar to human eyes. There is a central area with high density of photo-receptors while in the periphery the density is low. This non-uniform resolution of the human retina has been found to follow a logarithmic-polar law, taking into account the linear increment of photo-receptors size with distance to the center [14]. Combining this feature with the ability to move the eyes over the visual field, is possible to obtain the information needed to perform the desired tasks, without considering all the data contained in uniformly sampled images, and avoid the waste of processing effort.

In practice, log-polar mapping has other major advantages over the usual cartesian mapping. Among them are the shape invariance to scaling and rotation [16], improved binocular fusion [17] and, the one we will focus most, adequate correlation properties to vergence control [9].

In spite of these advantages, there are, obviously, disadvantages when compared to the

cartesian representation. One of the strongest drawbacks is the unsuitability of the log-polar representation to perform template matching. Due to the variant nature of the template according to its position, it becomes difficult to estimate, for example, disparity measures in a stereo pair. Although there has been some effort in overcoming this problem [9, 17], we find the log-polar mapping not very appropriate, by construction, for this kind of measures.

Notwithstanding, following active vision paradigms and taking an active part in the sensing procedure, it is possible, as we will show later, to achieve correct vergence configurations avoiding explicit disparity estimation which, for vergence control purposes, is an unessential step. Furthermore, performing an active behavior, the agent becomes more robust to environmental changes, increasing its ability to cope with dynamic scenes. As in [12] vision is used as a purposive sensor, providing feedback signals to close control loops.

The advantages of using continuous control strategies, relying on fast purposive visual cues, is a large advantage to achieve robust real-time vision systems. The system we propose, controls the vergence angle as a function of a correlation index between two log-polar images. We propose a control strategy that leads to robust global performance and requires no calibration. Another relevant aspect in this work is the extensive study of the influence of the log-polar sensor design parameters on the overall system performance. A criterion is established to define a suitable sensor geometry. A real time implementation is shown and experiments described.

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, 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. Performance criteria are defined in order to analyze the behavior of the vergence control algorithm for various log-polar sensor layouts.

The control algorithm is described in Section IV, with special attention given to the particular kind of information extracted from the images : a correlation index function, to be optimized by an adequate hill climbing algorithm.

In Section V, we present a real-time implementation and various experiments that illustrate the influence of sensor geometry on the shape of the correlation function, and on the performance of the control algorithm in realistic situations. Finally, in Section VI 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 toward 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 [9], 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 the existing work on the log-polar image mapping [14, 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 $(\xi, \eta)$. 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 $\rho$ and $\theta$ denote the commonly used polar coordinates :

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

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

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

where $\rho_{max}$ depends on the cartesian plane dimension.

As defined initially, the logarithmic-polar mapping yields continuous values in the domain $[log(\rho_{min}), log(\rho_{max})[\times[0, 2\pi[$. This 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.

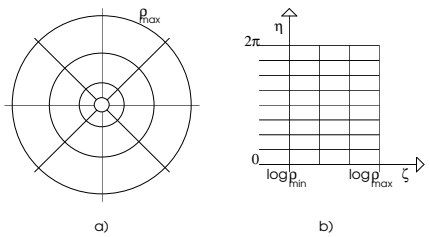


Figure 1: Space variant mapping from the cartesian plane, (a), to the log-polar plane, (b).

In the $\xi$ 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 $\rho_{min}$. The performance of the vergence algorithms will be analyzed as a function of these parameters. While only the parameter $A$ affects discretization in the $\eta$ direction, all the other parameters affect the $\xi$ 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 $\rho_{min}$ and $k$ parameters :

$$\rho_{min} = \frac{1}{2\tan\frac{\pi}{A}}$$

$$k = 1 + \frac{2\pi}{A}$$

This 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 $\rho_{min}$, $E$ and $k$. Alternatively, one can also use the default values.

As an illustration of the log-polar mapping, Figure 2 shows the mapping of a cartesian image with constant $\rho_{min}$ and different values for the resolution factor, $k$. This figure shows

Figure 2: The original cartesian image, shown on top, is mapped with different logarithmic factors, k=1.02, k=1.08 and k=1.15, from left to right. The images in the central row are cartesian representations of the bottom log-polar images.

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 Correlation for vergence

When following a moving target, the human eyes motion 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 target is "foveated". Figure 3 illustrates the vergence angle, $\nu$, in a binocular system.

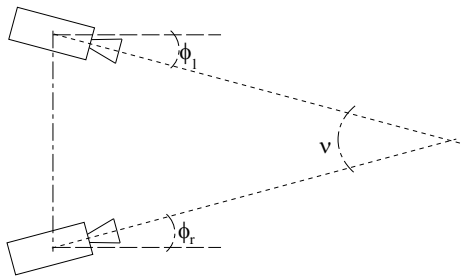


Figure 3: Stereo cameras verging at a point. The vergence angle, $\nu = \phi_l - \phi_r$, is shown.

The vergence process enables foveated animals to register objects of interest on the fovea of each eye. As a consequence, it allows the extraction 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 [6].

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 [4].

Traditional algorithms to extract disparity include cepstral filtering and phase correlation, which are very time consuming and therefore, not very adequate for real-time applications. Furthermore, in these algorithms, once the disparity is calculated, vergence can be achieved in a open loop manner. On the other hand, there is no evidence of the human vergence system being a sampled data system [11]. By relying on more active visual processes, it is possible to avoid the explicit computation of disparity and still detect correct vergence configurations.

In the approach used here, control plays a very important role, as vergence could not be achieved without a continuous control of the vergence angle. The resulting algorithm is very simple and consists basically in estimating, over time, a similarity measure between two stereo images and controlling the vergence angle in order to increase their similarity. An adequate vergence performance depends on two main aspects : the choice of a proper similarity measure and the definition of a good image sampling method (sensor geometry).

The simplest and most commonly used cost function to estimate the similarity between two images is 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. At the correct vergence angle, both images should be very similar and the SSD value close to zero.

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. An alternative solution is the normalized cross-correlation which behaves robustly [7] relative to changes in the illumination conditions :

$$NCC = 1 - \frac{\sum (I_l - \mu_l)(I_r - \mu_r)}{\sqrt{\sum (I_l - \mu_l)^2 \sum (I_r - \mu_r)^2}}$$

where $\mu_l$ and $\mu_r$ are the average gray levels for the left and right images, respectively.

Once a suitable similarity function or correlation index has been chosen, we can wonder about the importance of the image sampling geometry. We have performed an experiment 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 sweep the full vergence space. Figure 4 shows a plot of $\mathcal{C}(\nu)$, the NCC correlation index as a function of the vergence angle, both for 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, the curve is much smoother and presents a sharp minimum in the vicinity of the correct vergence angle. As discussed also in [9], there are large advantages in using the log-polar geometry to correlate image patches and, therefore, for a vergence system.

Finally, to evaluate the influence of the log-polar design parameters in the overall vergence system, we have to define a desirable behavior or shape for $\mathcal{C}(\nu)$, the correlation index as a function of the vergence angle, $\nu$.

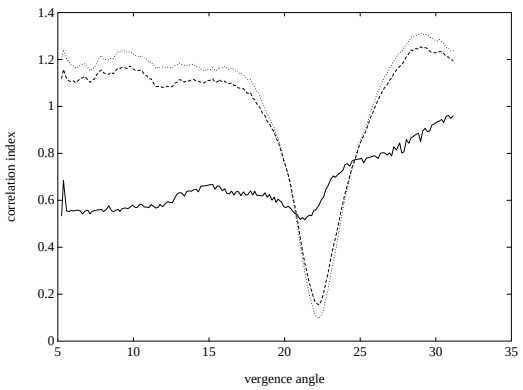


Figure 4: The correlation index as a function of the vergence angle, $\mathcal{C}(\nu)$. The solid line represents the correlation index calculated over a cartesian pair of images. The dashed and dotted lines show the same index but for log-polar images. The minimum values for these two indexes represent the situation where the cameras are correctly verging on a point in space.

## 3.1 What are good properties for a correlation index?

Presumably, by changing the design parameters of the log-polar sensor, the shape of $\mathcal{C}(\nu)$ should also change. Therefore, we need to establish a criterion to evaluate the "quality" of $\mathcal{C}(\nu)$ for a given sensor configuration.

Being always positive, $\mathcal{C}(\nu)$ cannot be used directly as an input to a closed loop controller. Instead, $\mathcal{C}(\nu)$ is a "similarity" index between the stereo images, that provides a quality measure for a given vergence angle, $\nu$. The vergence control system has then to optimize $\mathcal{C}(\nu)$, by applying some sort of hill climbing algorithms. According to this reasoning, some desirable properties of $\mathcal{C}(\nu)$ can be defined :

- The similarity criterion should be robust to offsets in the average value of images. As in [9], our experiments confirmed that the Normalized Cross Correlation (NCC) is very robust to brightness changes, thus fulfilling this requirement.

- $\mathcal{C}(\nu)$, should have a sharp relative minimum for the correct vergence configurations,

and increase monotonically when the vergence angle deviates from the correct value. The existence of a sharp minimum ensures that an hill climbing algorithm can converge to the correct vergence and convergence can still be guaranteed even in the presence of noise or disturbances.

- The "width" of the valleys of $\mathcal{C}(\nu)$ should be large enough for the hill climbing algorithm to converge to the local minimum, starting from a large set of camera positions.

- The value of $\mathcal{C}(\nu)$ should be as low as possible, for the correct vergence. To use the value of the correlation index in the input to the vergence loop controller, this is a very important feature. The motion of the cameras should be slow when approaching the ideal vergence angle.

Except for very cluttered environments, in the experiments made with NCC of log-polar images, all the correct vergence configurations were associated to local minima in $\mathcal{C}(\nu)$. 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.

# 4 Control algorithm

The control system for the vergence system uses direct image measurements from the log-polar images to drive the cameras motion. The control architecture for the vergence loop is represented in Figure 5.

This a standard feedback control architecture. At any instant of time $(i)$, the captured

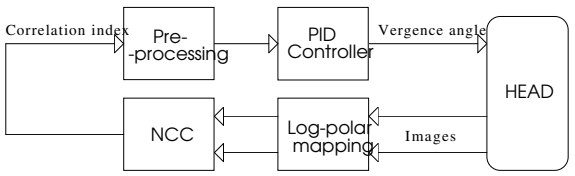


Figure 5: Vergence loop control diagram.

stereo images are mapped to a log-polar pair and the NCC is computed. This procedure provides a non-negative correlation index, $\mathcal{C}_i$, which exhibits local minima in correct vergence configurations. This signal is not suitable as direct input for a PID controller and must be processed somehow to provide an adequate signal to control.

Usual gradient descent techniques require calculation of the gradient of the function $\mathcal{C}(\nu)$ where $\nu$ is the vergence angle. The control reference for relative vergence angle value ($\delta\nu_i$) would be given by:

$$\delta\nu_i = -\alpha\frac{\Delta\mathcal{C}_i}{\Delta\nu_i}$$

where $\frac{\Delta\mathcal{C}_i}{\Delta\nu_i}$ is a discrete estimate of $\nabla\mathcal{C}(\nu_i)$).

Apart from its simplicity, the determination of the gradient was found to be very sensitive to noise. Furthermore, this algorithm is well suited to parabolic shaped valleys, where the reference signal tends to zero smoothly as we approach the minimum, whereas the slope of the valley for $\mathcal{C}(\nu)$ is very high.

In [9], an algorithm was proposed to calculate the reference value to control, 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 sensitive to noise in certain situations and it potentially presents a singularity in the minimum ($\nabla\mathcal{C}(\nu)$ becomes zero), which poses problems to the control.

Finding that 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 in 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 former reference value:

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

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 reverses direction. 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 algorithm proved to converge robustly to the correct vergence when starting, as initial condition, in one of the sides of a vergence valley. When starting in other conditions, the control value is kept under appropriate limits, never taking very large values as it happens when using algorithms relying on gradient information.

The vergence control is based only on very purposive cues, continuously acquired, and a minimization algorithm. No calibration information is needed in order to accomplish vergence. This independence on the knowledge of the intrinsic and extrinsic parameters of

the system turns out to be one of the major advantages of active visual processes.

# 5 Results

The vergence system described in the previous sections was implemented on the **Medusa** Stereo Head, shown in Figure 6. The head has four mechanical degrees of freedom, including

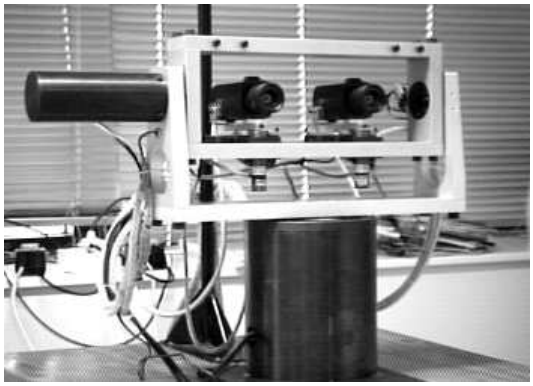

Figure 6: Medusa stereo head used for the experimental tests.

independent vergence for both cameras, and a common tilt and pan axes. All the computations, including image mapping and processing, are done at a 486 PC computer running at 66MHz. The data reduction achieved by using the log-polar mapping makes it possible to have reasonable sampling rates for the visual control (about 3Hz), even when using a modest computational power.

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.

## 5.1 Influence of the log-polar parameters

As we explained, 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 considered parameters in this study are the following :

$A$ - the number of cells per eccentricity

$E$ - the number of eccentricities

$k$ - the logarithmic factor

$\rho_{min}$ - value of the minimum radius.

The dimensions of the original cartesian images are fixed to $128 \times 128$ throughout all the tests and discussions. Experiments were made in a diversity of scenes and situations. For a given scene, the experiment consists in sweeping the vergence angle space and computing the correlation index for several camera positions, and 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 with a static scene with several depth planes. Figure 7 shows the scene, as seen by one of the cameras of the stereo head. The scene consists in a

Figure 7: Experimental setup. There are three different depth planes in this cluttered environment.

flat object roughly located at one meter from the cameras, against a second plane of objects at about 3 meters and the background at 5 meters.

**a) Changing $A$ keeping $E$, $k$ and $\rho_{min}$ constant**

The first experiment consists in changing the value of the number of cells per eccentricity, $A$, while parameters $E$, $k$ and $\rho_{min}$ are kept constant. The corresponding log-polar layouts are shown in Figure 8.

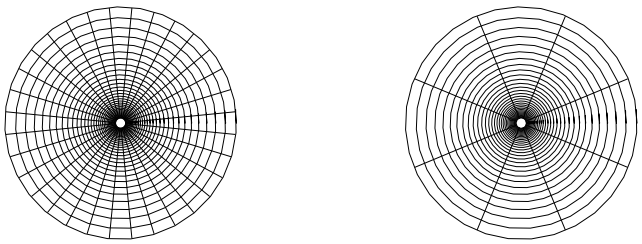

Figure 8: Effect of the change of number of angles per eccentricity while keeping the other parameters constant.

As can be observed in Figure 9, the increase in the number of cells per eccentricity does not lead to substantial improvements to the shape of the correlation index.

In the following experiments we will keep the value of $A$ in 64.

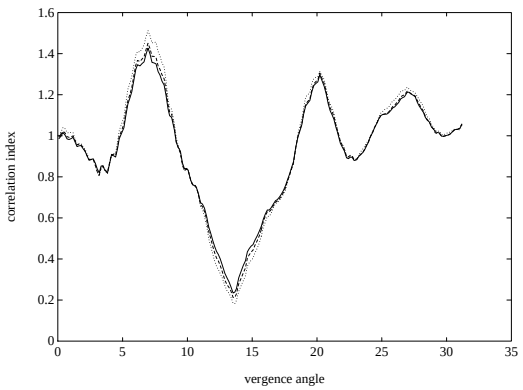


Figure 9: Plots of $\mathcal{C}(\nu)$ for: $E = 64$, $k = 1.05$ , $\rho_{min} = 3.0$ pixels. Solid line: $A = 64$. Dashed line: $A = 32$. Dotted line: $A = 16$.

**b) Changing $E$ for fixed $A$ and $k$**

As stated in equation (5), the parameters $E$, $k$ and $\rho_{min}$ are not independent; the definition of two of them determines the value of the third.

In this experiment the value of $k$ is constant and equal to 1.05. The value of the number of eccentricities, $E$, is changed and the inner radius, $\rho_{min}$, is determined by equations (5). The corresponding sensor layouts are illustrated in Figure 10.

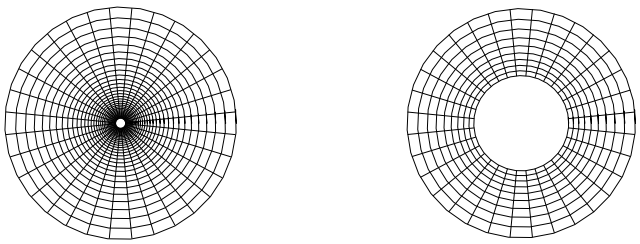

Figure 10: Decreasing the number of eccentricities while keeping the logarithmic factor constant leads to the increase of the minimum radius.

The correlation index, $\mathcal{C}(\nu)$, for the various values of $E$ are illustrated in Figure 11. It is possible to assess the influence of $E$ and $\rho_{min}$ in the shape of the correlation index, as a function of the vergence angle. As can be observed, for higher values of $E$ (lower values of

$\rho_{min}$) the valley around the correct vergence angle is wider and deeper which should lead to better performance.

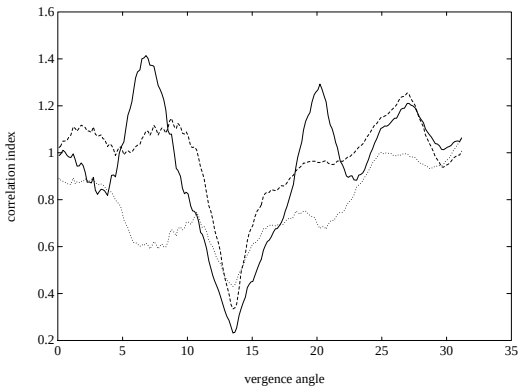


Figure 11: $\mathcal{C}(\nu)$ for varying values of $E$. Solid line: $E = 64$ , $\rho_{min} = 2.82$ pixels. Dashed line: $E = 32$ , $\rho_{min} = 13.43$ pixels. Dotted line: $E = 16$ , $\rho_{min} = 29.32$ pixels.

It is not possible to determine yet if the improvements are due to the increased resolution provided by the increase in the number of eccentricities, or to the smaller blind zone as the minimum radius decreases.

### c) Changing $\rho_{min}$ for fixed $A$ and $E$

In this experiment the value of $E$ is kept constant and equal to 32. The value of the innermost radius is increased which produces a reduction in the value of the logarithmic factor. The corresponding layouts are shown in Figure 12.

The results shown in Figure 13, reveal that the smaller the minimum radius (the larger the logarithmic factor) the deeper and wider is the shape of the vergence valley. According to our quality criteria, this should also improve the overall vergence system performance.

As in the former case, is not possible to decide which parameter ($k$ or $\rho_{min}$) is the ultimate responsible for the improvements.

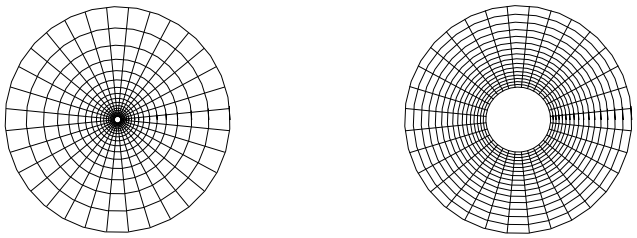

Figure 12: Effect of increasing the innermost radius while keeping the number of eccentricities constant, which leads to the decrease of the logarithmic factor.

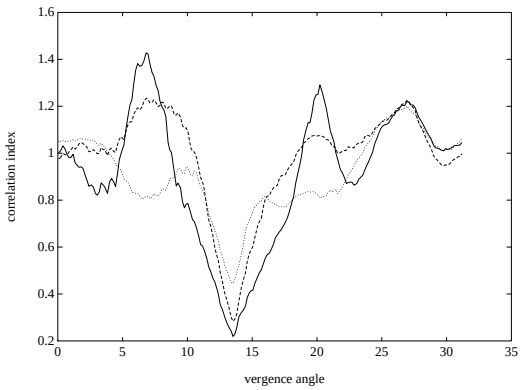


Figure 13: $\mathcal{C}(\nu)$ for changing values of $\rho_{min}$ Solid line: $\rho_{min} = 2.0$ pixels, $k = 1.11$. Dashed line: $\rho_{min} = 10.0$ pixels, $k = 1.06$. Dotted line: $\rho_{min} = 20.0$ pixels, $k = 1.04$.

**d) Changing $k$ for fixed $A$ and $\rho_{min}$**

This final experiment consists in keeping the value of the minimum radius constant, $\rho_{min} =$ 3pixels and changing the values of the parameters $E$ and $k$. As usual, the sensor layout is depicted in Figure 14.

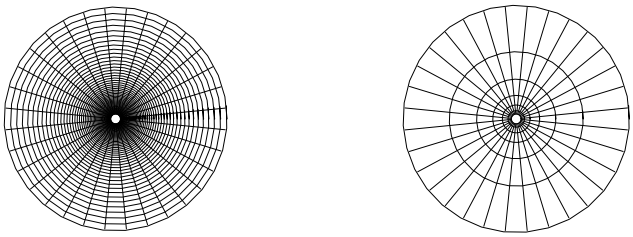

Figure 14: Effect of increasing the logarithmic factor while keeping the innermost radius constant, which leads to the decrease of the number of eccentricities.

The results, shown in Figure 15, reveal that a larger value for $k$ (smaller value for $E$) provides wider and deeper vergence valleys, therefore leading to improvements in the performance of the control algorithm.

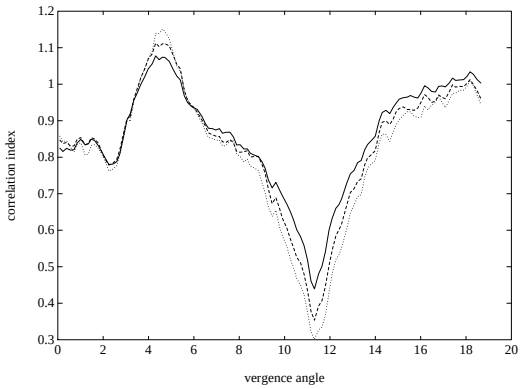


Figure 15: Plots of $\mathcal{C}(\nu)$ for constant inner radius, $\rho_{min}$. Solid line: $k = 1.02$, $E = 154$. Dashed line: $k = 1.15$, $E = 21$. Dotted line: $k = 1.30$, $E = 11$.

Each one of last three experiments, by themselves, are not decisive in the analysis of which parameters provide better performance. However, by considering the experiments

altogether it is possible to conclude how the parameters affect the shape of $\mathcal{C}(\nu)$ around the minimum.

The first observation refers to the experiments **b)** and **d)**. In the former, a improved shape was obtained with many eccentricities while the latter has shown the opposite case. According to this observation, it can be assumed that the improved resolution given by the increase in the number of eccentricities, is not determinant to the performance of the algorithm. This observation is coherent with the results of the experiment **a)**, where improved resolution due to the increase of the number of cells per eccentricity did not change substantially the shape of $\mathcal{C}(\nu)$.

With this assumption, the cause for wider an deeper valleys in $\mathcal{C}(\nu)$ is assigned to the reduction of the minimum radius in the experiment **b)**, and to the increase of the logarithmic factor in the experiment **d)**. This is also coherent with the results of the experiment **c)**.

Of course, as we want to verge on the object roughly located in the center of the retinas, this area is extremely important and a large void area is very inconvenient. In face of the experiments we can also say that the minimum radius is a determinant factor in the performance of the vergence system.

The logarithmic factor $k$ is not as important but it is highly recommended to increase its value not only because the performance of the algorithms slightly improve but mainly because the number of eccentricities is reduced, increasing the data reduction factor. This slight improvement in the shape of $\mathcal{C}(\nu)$ can be explained on the basis of the stronger weight given to the pixels near the origin relative to the pixels located in the periphery. The relative importance of the peripheral pixels is therefore very reduced.

To conclude, we can say that, as a rule of thumb, we should choose a low value for the minimum radius $\rho_{min}$ and a high value for the logarithmic factor $k$. Obviously these parameters should remain within certain limits. Very low values for the minimum radius increase the redundant information contained in the foveal area. This redundancy, measured in the percentage of redundant pixels in the total number of pixels in the log-polar image, should be kept low, for efficiency purposes, depending on the constraints of each particular application.

Very large values for the logarithmic factor tend to reduce excessively the number of eccentricities. Although we have not tested the effect of extremely reduced number of eccentricities, it should be kept under reasonable limits, or otherwise, the loss of resolution in the foveal area can be important.

## 5.2 Controlled Vergence

We have performed various experiments to test the closed loop vergence control system. A variety of objects, located at different distances from the stereo head, have been used and the system is able to correctly verge on a target. The current implementation runs with a speed of about 3Hz.

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 camera. The object is then suddenly moved to a different position and keeps moving back and forth, after a small time period. This input corresponds to a sequence of steps functions applied to the input of the vergence system. The closed loop controller is a simple proportional controller.

Figure 16 shows the results of this controlled experiment, illustrating the time evolution of the vergence angle, correlation index and control action. The time scale units are sampling periods (about 330ms).

The correct vergence angle is reached after a short transient period, similar to a first order dynamic system. The control goal is to minimize the correlation index, which is attained in stationary situations. When the target distance changes abruptly, a spike value in the correlation index appears, denoting an incorrect vergence. Then, a response spike in the control signal compensates for this effect, and brings the stereo head to the correct vergence again, in a few sampling periods. When the cameras approach the correct vergence, the activation signal decays, the same happening with the correlation index.

A point to notice is that the cameras move continuously, even when the desired vergence angle is achieved. The reason why, is that the system is always responding to the temporal difference cues, needed to react to depth changes. Although this could be considered a disadvantage, the camera motion can be kept small, in the equilibrium point.

# 6 Conclusions and further work

In this paper we have shown that robust vergence control can be achieved, through an active process, using log-polar images. Vergence can be controlled without the explicit measurement of disparity, by using simple visual information in the feedback loop. There is no need to calibrate the cameras.

In this work biological analogies can be established in two main aspects : First, human vergence system was shown to be a continuous process, as the method we propose here.

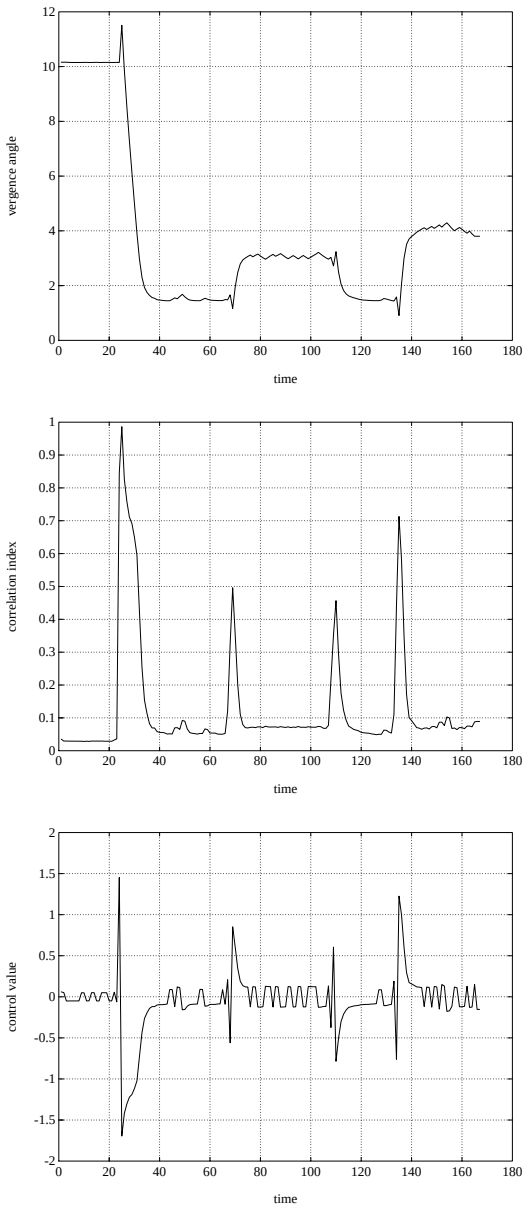


Figure 16: Response of the closed loop control system to step-like stimuli. The top curve is the evolution of the vergence angle, over time. The middle plot shows how the correlation index has evolved during the whole experiment. The bottom plot shows the evolution of the control action in degrees.

Second, a non-uniform sampling mechanism such as the log-polar mapping is based on the topology of the human retina.

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.

The closed loop control algorithm is based on a hill climbing procedure applied to the correlation index space. Differential measurements are avoided because of noise sensitivity. The control law is solely dependent on the value of the similarity index between two images.

The system was implemented, in real-time, on a stereo head Medusa. A large number of experiments are described in the paper, using realistic scenes and cluttered environments.

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 algorithm proved to be stable and robust. Furthermore, in dynamic situations it shows fast reaction and settling times.

Future work should address the problem of integrating the vergence system with the other similar ocular movements, following the same approach. Particularly, we envisage various other visual behaviors specialized in different parts of the visual field (foveal or peripheral), mutually cooperating in the task of following a moving target, or more complex tasks such as docking, grasping, etc.

# 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] A. Bernardino, C. Silva, J. Santos-Victor, and C. Pinto-Ferreira. Behaviour based oculomotor control architecture for stereo heads. In *Proc. of the 3rd Int. Symposium on Intelligent Robotic Systems*, Pisa, Italy, July 1995.

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

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

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

[7] 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.

[8] 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.

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

[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] J. Santos-Victor, F. van Trigt, and J. Sentieiro. Medusa - a stereo head for active vision. In *Proc. of the Int. Symposium on Intelligent Robotic Systems*, Grenoble, France, July 1994.

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

[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.