Correlation based Vergence Control using Log-polar Images *

Alexandre Bernardino** **José Santos-Victor *****

Instituto de Sistemas e Robótica
Instituto Superior Técnico
1096 Lisboa Codex
Portugal

Abstract. This paper describes a real-time vergence control mechanism based on log-polar images, developed for a robot head. We show that vergence behavior can be achieved at reduced computational cost using simple correlation measures on log-polar images.
The main advantages of using a non-uniform image sampling mechanism, such as the log-polar images, are related both to perceptual and algorithm complexity issues. We show that, when using correlation measures to control vergence, log-polar images give better results than cartesian images. Additionally, as log-polar images are smaller, the computation time is reduced.
Two algorithms for closed loop vergence control, using correlation measures over log-polar images, are proposed and compared. Their behavior in real situations is illustrated by test examples.

1 Introduction

The research interests in Active Vision have increased in the past few years providing efficient ways of combining the control of robotic systems and advanced visual sensing at a moderate computational cost [1, 2]. Various stereo heads have been designed and built in many research laboratories [5] and were often inspired after their biological analogues [4, 12]. Several ocular movements, like saccades, smooth pursuit and vergence, found in biological vision systems have been implemented and tested.

In nature, one of the main reasons for the existence of active head-eye control is the space-variant density of photo-receptors in the eye, decreasing from the fovea toward the periphery. Combining this feature with the ability to move the eyes over the visual field, it is possible to obtain the information needed to perform the desired tasks, without considering all the data contained in uniformly sampled images. Hence, the processing effort can be concentrated in the high resolution area - the fovea.

The main contributions of this paper are twofold. First we exploit the geometric properties of the space-variant sampled images to show that both performance and computational advantages arise when compared to the cartesian images. Correlation measurements computed in the space variant domain take advantage of the higher resolution at the image center as opposed to the image periphery. The second aspect deals with the design of control strategies that use correlation measures directly to control the vergence angle without requiring any intermediate step to reconstruct the environment or calibrate the cameras. We believe that these problems : purposive sensor geometry and control strategies which integrates directly the visual information in the control system, may have a significant impact on various applications of computer vision and robotics [13].

2 Log-Polar Mapping

Apart from few exceptions [10], most active vision systems use cartesian image representations. In humans, however, the retina exhibits a non-uniform resolution and has been found to follow a logarithmic-polar law,

* Alexandre Bernardino was 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.
** Email: ajmb@isr.ist.utl.pt
*** Email: jasv@isr.ist.utl.pt

taking into account the linear increment of the photo-receptors size with the distance to the center [15]. The vergence system described in this paper will be based on a log-polar space variant image sampling. The log-polar images simulate the sampling mechanism of the human retina and, additionally, provide an important data reduction, which is always desired in real-time applications. Various authors have pointed out advantages of log-polar images when compared to cartesian images, like shape invariance to scaling and rotation [18], improved binocular fusion [19], easy computation of time to contact [17], image compression [8] and adequate correlation properties to vergence control [3].

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), [15, 18, 19]. The mapping is described by :

$$\begin{cases} \xi = \log \rho \\ \eta = \theta \end{cases} \tag{1}$$

where ρ and θ denote the commonly used polar coordinates.

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 :

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

where ρ_{max} depends on the cartesian plane dimension.

Figure 1 shows the discretized cartesian and log-polar planes.

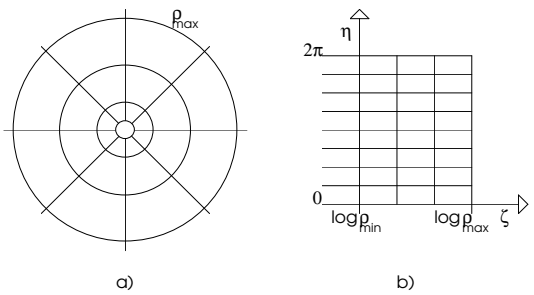

Fig. 1.: Space variant mapping from the cartesian plane, (a), to the log-polar plane, (b).

3 Vergence using correlation

Vergence is one of the basic eye movements to focus visual attention on a given scene point [12]. 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". This enables to register objects of interest on the fovea of each eye, allowing 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 [6]. Figure 2 illustrates the vergence angle ν, in a binocular system.

As the vergence angle is directly related to the target distance, depth cues such as binocular disparity and accommodation, can be used to help the vergence system in compensating depth changes. Traditional algorithms to calculate explicit disparity measures include *cepstral* filtering [20] and phase correlation [9], but these techniques are very time consuming, and therefore, not very adequate for real-time applications. In this paper we exploit the correlation of log-polar images to extract disparity cues. We will show that, by using these cues and relying on more active visual processes, it is possible to avoid the explicit computation of disparity and still attain correct vergence.

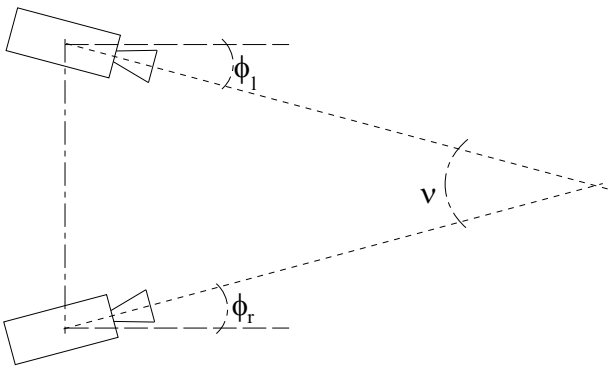

Fig. 2.: Stereo cameras verging at a point. The vergence angle, $\nu = \phi_l + \phi_r$, is shown.

In this paper we design and compare two algorithms for vergence control, which rely on closed loop control and on simple correlation measures of log-polar images. As both algorithms use correlation measures, we start by analyzing the differences between the correlation of images of a stereo pair on the cartesian and log-polar domains, for small deviations from vergence.

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

$$I\left(x + \delta_x, y\right) = I\left(x, y\right) + \frac{\partial I}{\partial x} \cdot \delta_x \tag{2}$$

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

$$c_c = \int\int_{x,y} \left[I\left(x + \delta_x, y\right) - I\left(x, y\right)\right]^2 dxdy \tag{3}$$

Substituting (2) in (3), the correlation value is given by:

$$c_c = \delta_x^2 \cdot \int\int_{x,y} \left(\frac{\partial I}{\partial x}\right)^2 dxdy \tag{4}$$

Analogous 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 is given by:

$$c_l = \int\int_{\xi,\eta} \left(\frac{\partial I}{\partial \xi} \cdot \delta_\xi + \frac{\partial I}{\partial \eta} \cdot \delta_\eta\right)^2 d\xi d\eta \tag{5}$$

The increments δ_ξ and δ_η can be expressed as:

$$\begin{cases} \delta_\xi = \frac{\partial \xi}{\partial x} \cdot \delta_x \\ \delta_\eta = \frac{\partial \eta}{\partial x} \cdot \delta_x \end{cases} \tag{6}$$

therefore:

$$c_l = \delta_x^2 \int\int_{\xi,\eta} \left(\frac{\partial I}{\partial \xi} \cdot \frac{\partial \xi}{\partial x} + \frac{\partial I}{\partial \eta} \cdot \frac{\partial \eta}{\partial x}\right)^2 d\xi d\eta \tag{7}$$

To compare the correlation of log-polar images with the correlation of the corresponding cartesian images, we need to change the coordinates of equation (5) from log-polar to the cartesian domain. We have:

$$\frac{\partial I}{\partial x} = \frac{\partial I}{\partial \xi} \cdot \frac{\partial \xi}{\partial x} + \frac{\partial I}{\partial \eta} \cdot \frac{\partial \eta}{\partial x} \tag{8}$$

and the determinant of the coordinate change jacobian matrix J is given by:

$$|J| = 1/\rho^2 \tag{9}$$

Hence the correlation function of the log-polar images, expressed in cartesian coordinates, is:

$$c_l = \delta_x^2 \iint_{x,y} \frac{1}{\rho^2} \left(\frac{\partial I}{\partial x} \right)^2 dxdy \tag{10}$$

Observing carefully the expressions for the correlation measures in the cartesian and log-polar domains (equations (4) and (10)), 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 over the cartesian geometry. Suppose that the cameras are verging on an object in the center of the visual field so that the area under vergence has 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.

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. For the remaining of this paper, the normalized cross-correlation is used. This function 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 μ_l and μ_r are the average gray levels for the left and right images, respectively.

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 sweep the full vergence space. The cartesian images size is 128x128 pixels whereas the log-polar images size is 64x32 pixels, thus achieving a reduction factor of 8 and a better computational efficiency. Figure 3 shows a plot of $\mathcal{C}(\nu)$, the NCC 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, with 8 times less data, the curve is much smoother and presents a sharp minimum in the vicinity of the correct vergence angle.

4 Control algorithm

In this section we propose and compare two strategies for vergence control, which work in a standard feedback control architecture, as represented in Figure 4.

Both vergence control mechanisms are based only on correlation cues acquired continuously over time. No calibration information is needed in order to accomplish vergence.

The difference among the two strategies remains on the processing step, as will be described in the next sub-sections.

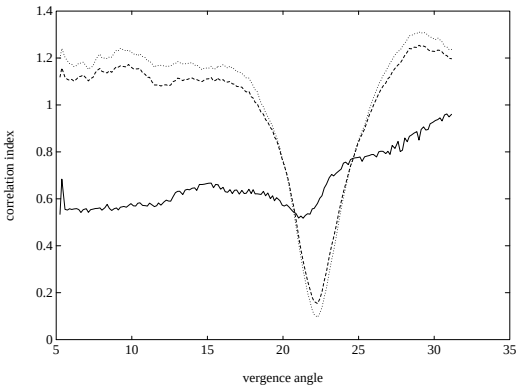

Fig. 3.: $\mathcal{C}(\nu)$: correlation index as a function of the vergence angle. Solid line: $\mathcal{C}(\nu)$ for a cartesian pair of images. Dashed and dotted lines: $\mathcal{C}(\nu)$ for log-polar images with different layouts. The minimum values represent the correct vergence angle.

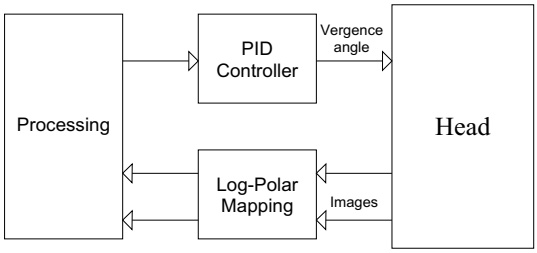

Fig. 4.: Vergence loop control diagram.

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 NCC between the two images, in a fashion similar to [10].

As we have seen in the former section, for log-polar images the correlation function $\mathcal{C}(\nu)$, exhibits a sharp minimum in correct vergence configurations. Therefore, to control vergence, we need to minimize the correlation function. Usual minimization techniques (eg. gradient descent) require the calculation of differential measures of $\mathcal{C}(\nu)$. In this case, however, differential measures exhibited severe instability due to sensor noise. Therefore, 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:

$$s_i = \mathcal{C}_i \cdot [-sgn(\Delta\mathcal{C}_i)] \cdot sgn(s_{i-1}) \tag{11}$$

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. In this algorithm two points should be noticed. First, the cameras must be always moving in order to detect variations in the correlation function needed to calculate the control value. Second the minimum value of the correlation function is not the same in all vergence configurations. Its value depends on the particular situation. Therefore, as the control signal depends on this value, the head behavior will be distinct for different vergence situations.

4.2 Preprogrammed vergence control

Several theories exist for vergence eye movement control. In [16] the theory of dual mode vergence eye movement is sustained and states that the initial portion of the movement is composed by a predetermined

motor force program, followed by a slower movement to attain accurate bifixation. 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 [11] 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.

Similarly, the second control strategy uses a channel based approach. Each channel is tuned for a specific disparity in the following way: the correspondence between the pixels in the log-polar images is established, for the respective disparity. These correspondences can be calculated off-line and stored in a look up table. In run time, the NCC between the corresponding pixels is calculated for each channel. The channel to be activated at each time is the one with the maximum correlation. In [11] it is argued that 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 Results

The vergence system described in the previous sections was implemented on the **Medusa** Stereo Head [14]. All the computations, including image mapping and processing, are done at a PENTIUM PC computer running at 133MHz. In the current implementation, the data transfer time from the image frame grabber to host computer memory limits the sampling rate to about 6 Hz. Without this constraint, for the data reduction achieved by using the log-polar mapping, it would be possible to run at video rates.

5.1 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 evaluate the system response to step-like stimuli, we placed an object at a given distance from the camera and then removed it. A simple proportional controller was used.

Figure 5 shows the results of this controlled experiment, illustrating the time evolution of the vergence angle. The time scale units are sampling periods (about 160ms).

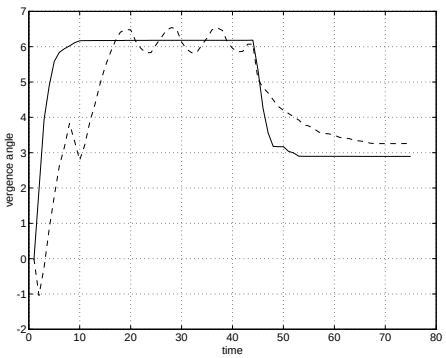

Fig. 5.: Response of the closed loop control system to step-like stimuli. Solid plot: angle of the preprogrammed vergence control. Dashed plot: angle of the iterative vergence control.

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. For a higher gain controller, the amplitude of the oscillations would be larger.

The control signal (equation 11), 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 are not perceivable because the lower correlation value leads to very small control signals. Both the oscillations in vergence situations and the direct dependence on the correlation value, are the main negative aspects of this approach.

The preprogrammed vergence control shows a much 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 bigger 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.

6 Conclusions

In this paper we have shown that fast vergence control can be achieved, using simple correlation measures over log-polar images. Vergence can be attained with the close integration of perception and control, without the explicit measurement of disparity and the need to calibrate the cameras.

The work described here is closely related to biological vision systems in at least two ways. On one hand the log-polar image sampling mechanism is inspired after the topology of the human retina. On the other hand the human vergence control system is based on a continuous stream of visual information, similarly to the approaches adopted.

The radial decrease of the resolution in the log-polar images was found to be more adequate for vergence control, when compared to cartesian images, and contributes to an important reduction of the amount of data to be processed.

Two vergence control strategies are proposed. The iterative vergence 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 preprogrammed vergence control algorithm follows a channel based approach. Depending on the level of disparity present in the log-polar images, a different channel is selected. The selection is made according to correlation values calculated over all channels defined. The selected channel determines the input to the controller module.

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 and cluttered environments.

Finally, results with controlled vergence are described. Both control algorithms confirm the usefulness of correlation measures to control vergence. However, major differences in their dynamic behavior are clear. The preprogrammed vergence control strategy was found to be more stable and robust. Furthermore, in dynamic situations it shows faster reaction and settling times.

To conclude, we would like to argue that the approach followed here addresses important issues that can have a wide impact on vision and robotic applications. In fact, by combining the purposive design of the sensor geometry with adequate control strategies, directly integrating perception and action, seems to be a powerful framework for successful applications of vision methodologies in the robotics field.

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 and J. Santos-Victor. Sensor geometry for dynamic vergence: characterization and performance analysis. In *Proc. of the Performance Characteristics of Vision Algorithms Workshop, ECCV96*, Cambridge, England, April 1996.
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. D. Fleet, A. Jepson, and M. Jenkin. Phase based disparity measurement. *CVGIP: Image Understanding*, 53(2):198–210, March 1991.
10. F. Panerai, C. Capurro, and G. Sandini. Space variant vision for an active camera mount. In *Proc. SPIE AeroSense95*, Florida, USA, April 1995.
11. M. Pobuda and C. Erkelens. The relation between absolute disparity and ocular vergence. *Biological Cybernetics*, 68:221–228, 1993.
12. D. Robinson. The oculomotor control system: A review. *Proc. of the IEEE*, 56(6), June 1968.
13. 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.
14. 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.
15. E.L. Schwartz. Spatial mapping in the primate sensory projection : Analytic structure and relevance to perception. *Biological Cybernetics*, 25:181–194, 1977.
16. J. Semmlow, G. Hung, J. Horng, and K. Ciuffreda. Disparity vergence eye movements exhibit preprogrammed motor control. *Vision Reserach*, 34(10):1335–1343, 1994.
17. 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.
18. R. Wallace, P. Ong, B. Bederson, and E. Schwartz. Space variant image processing. *Int. Journal of Computer Vision*, 13(1):71–90, September 1995.
19. C. Weiman. Log-polar vision for mobile robot navigation. In *Proc. of Electronic Imaging Conference*, pages 382–385, Boston, USA, November 1990.
20. Y. Yeshurun and E. Schwartz. Cepstral filtering on a columnar image architecture: A fast algorithm for binocular stereo segmentation. *IEEE Trans. on PAMI*, 11(7):759–767, July 1989.

This article was processed using the TEX macro package with SIRS96 style