Binocular stereo via log-polar retinas

Carl F. R. Weiman

Transitions Research Corporation
Shelter Rock Lane, Danbury, CT. 06810
weiman@trc.com

ABSTRACT

Log-polar pixel tessellation in the image plane improves binocular stereo perception compared to the familiar uniform Cartesian tessellation. This paper describes the advantages by analyzing the 3-D intersections of projections of dual retinas. The 3-D environment is divided into volume cells (voxels) which are the intersections of pixel projection cones. There are some interesting and useful differences between log-polar and Cartesian induced voxel distribution. Maximum stereo resolution for Cartesian voxels is inconveniently located at the outskirts of the field of view at the near point of intersection of the two fields of view, rapidly degrading therefrom. Log-polar stereo resolution is highest at the point of intersection of the optical axes. Active vision can steer this focus of attention like a spotlight to any point of interest in the 3-D environment. Within this focus of attention, stereo resolution is nearly uniform. Applications include active robot vision with close parallels to the human visual system and eye movements.

Keywords: binocular stereo, active vision, log-polar sensors, robot vision, visual servos, space-variant sensors.

1. INTRODUCTION - BINOCULAR STEREO VIA ACTIVE VISION

Vision, being the richest of human senses, has always been considered as potentially the richest of robot senses. Binocular stereo has been particularly alluring because of the apparently deterministic geometry of triangulation. In real world environments, however, the *correspondence problem*,[6] namely, associating two images of the same object, has proved formidable both algorithmically and computationally. The problem can be simplified by mounting stereo cameras on motion platforms to coordinate their convergence to a target. Once correspondence is achieved, by whatever initial means, camera motion can be servo controlled to maintain it, even while tracking a moving target. Such "active vision"[1] platforms also play an important role in eliminating motion blur on mobile robots and redirecting attention to new targets in a dynamic environment, whether the system is monocular or binocular.

Once the decision has been made to actively steer the camera(s), the environment is divided into that which is visible, and that which is not, for any particular position of the camera. The conflicting requirements to maximize field of view and resolution while minimizing pixel count can be reconciled by "foveation", that is, concentrating high resolution in the center and relegating low resolution to the periphery. Various schemes are available: multiresolution pyramids, uniform high resolution in the center, and warped coordinate systems such as fisheye and log-polar. Of these, log-polar is the most "harmonious" because it smoothly blends resolution from the center to the periphery with no discontinuities, and preserves isotropy of local neighborhood operators (e.g. edge detectors) throughout the field of view.

Foveated vision and active vision are mutually supportive.[2,9] Foveation requires steered cameras to direct high resolution resources to a target. Foveation aids the motion servoing process by reducing the computational load associated with pixel count, thereby reducing control delays. High central resolution also provides more accurate servo error signals for tight tracking of a locked-on target.

Below we describe the camera control configuration which underlies the binocular stereo geometry. A real-time binocular camera system was implemented in prototype and developed into a commercial product for vision researchers. In the following section we superimpose the log-polar coordinate system upon the image planes of this binocular camera configuration, quantify the improvement over traditional uniform Cartesian image plane pixels, and describe the contrasting binocular stereo disparity fields which are projected into the environment by these two different pixel patterns.

0-8194-1841-2/95/$6.00

2. BINOCULAR CAMERA CONFIGURATION AND CONTROL

A common strategy in the early days of binocular vision was to constrain cameras to parallel optical axes on a fixed baseline. However, for close viewing, disparity is so large that the correspondence problem becomes difficult. A preferred choice for active vision systems is a simplified four degree-of-freedom configuration, illustrated in figure 1. Two cameras are separated by baseline B with principal axes A_L (left) and A_R (right). Servo-controlled motors rotate cameras about parallel vergence axes V_L and V_R perpendicular to A_L and A_R through the nodal points (centers of perspective projection). The motors are in turn attached to a horizontal crossbar which tilts about a horizontal axis H. The central pan axis, V_C, is the vertical bisector of the H axis. The camera system can be aimed in any direction and verged to any range in the environment.

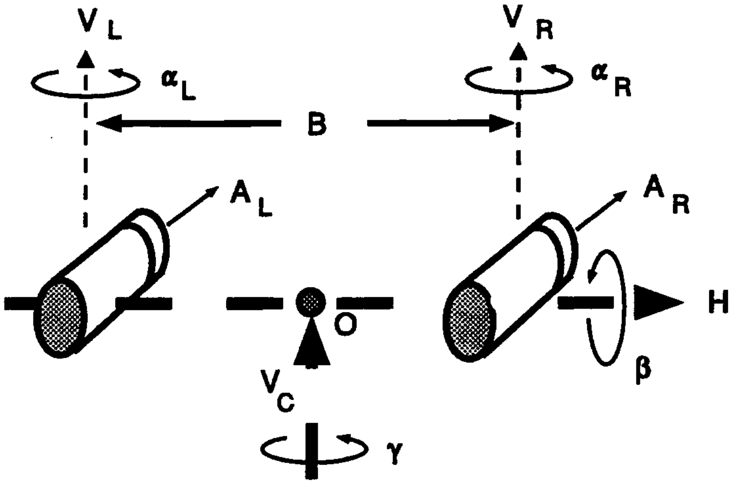


Figure 1. Articulation of camera mount

Symmetric vergence ($\alpha_L = -\alpha_R$ in figure 1) has important consequences for binocular stereo. It assures perfect correspondence at the centers of the fields of view, and eliminates asymmetric (left-right) warping in the disparity field. A simple control algorithm maintains symmetric convergence to a target as follows. Vergence cameras move to zero-out the horizontal component of angle between target image and optical axis. Independently, pan does likewise for the vertical component of target image. Beneath these, the head pan axis zeros out the average angle of vergence cameras, without reference to image position. Correspondence is maintained even while the target is moving. Figure 2 illustrates the control algorithm, with ψ_l and ψ_r referring to the horizontal components of the angle between target and optical axes A_L and A_R in the left and right camera images, and φ_l and φ_r the vertical components. Note that vergence and tilt are velocity outputs controlled by *image* position error, whereas neck pan is purely a function of vergence positions from encoders.

Figure 3 illustrates a plan view of the projections of two corresponding pixels at the center of the field of view of each camera of a symmetrically verged pair. The shaded quadrilateral in the center, the intersection of these projections, is a voxel (volume cell) which represents the position uncertainty of a 3-D point which is visible by both pixels simultaneously. We approximate this quadrilateral by a parallelogram to simplify the geometric analysis which follows, justifying the approximation by noting that for small angles, i.e. pixel subtense of a few milliradians, the error incurred is less than one percent; that is, pixel subtense is a differential quantity. From similar triangles in figure 3, the aspect ratio of such a voxel is the ratio of range Z to half-baseline, namely,

$$a = \frac{2Z}{B} \, . \qquad (1)$$

The length of the voxel is the magnitude of the range error for binocular stereo. Since this quantity is proportional to pixel subtense, we can quantify the log-polar range resolution advantage directly in terms of pixel size as described below.

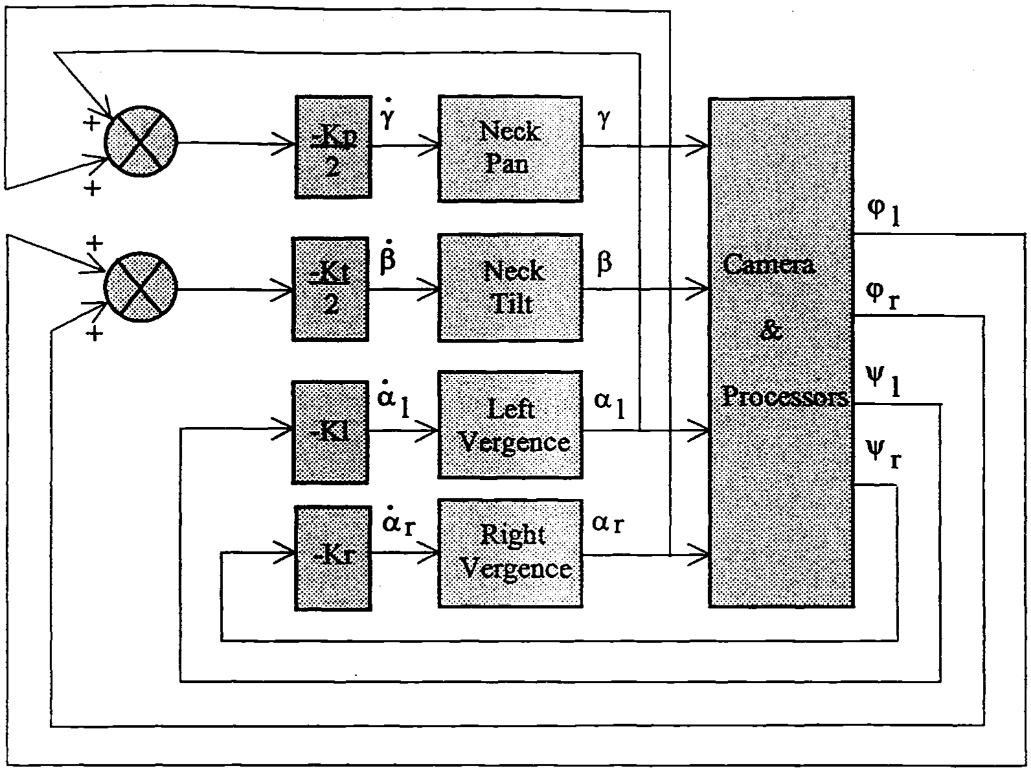


Figure 2. Tracking control for 4-DOF camera motion platform

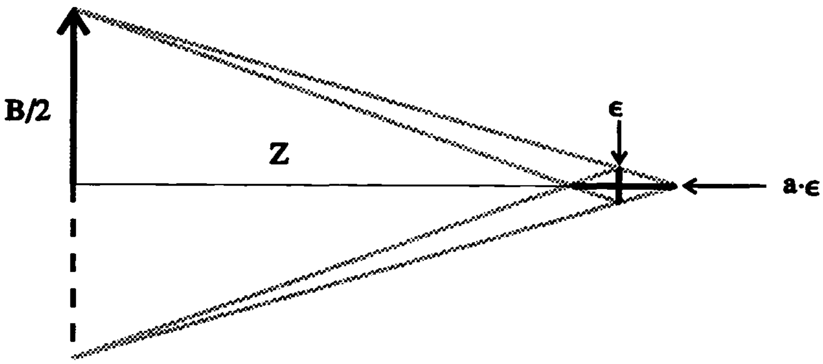


Figure 3. Voxel geometry at the binocular center

3. LOG-POLAR IMAGE PLANE COORDINATES

Log-polar image plane coordinates have been applied by a number of vision researchers.[2,4,5,8-15] In general, the logarithmically scaled coordinate system eliminates pixel addresses as arguments in computations involving transformations which scale imagery. Figure 4 represents the projection of a log-polar image plane into the 3-D environment, illustrating the zoom, rotation, and perspective symmetries of the pattern. In the following section we examine the intersection of two such configurations defined by convergent binocular viewing. First we quantify the resolution advantage of log-polar pixel arrays.

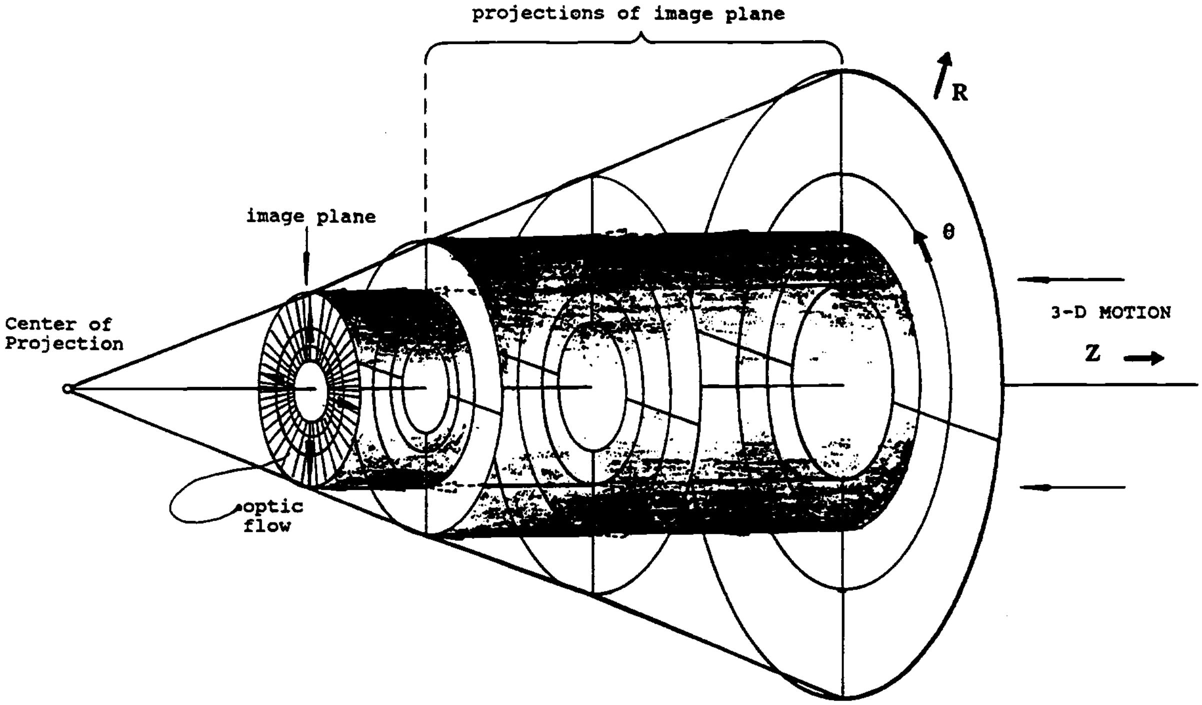
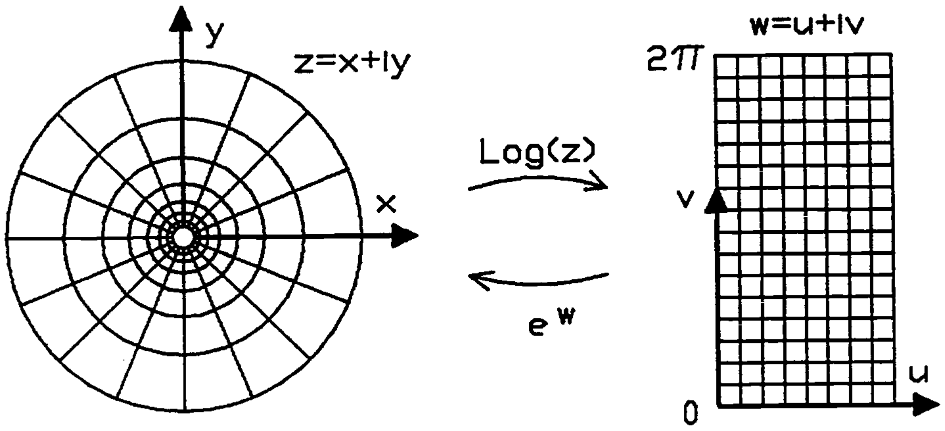


Figure 4. Projection of log-polar retina onto visual field

3.1 Definition of Log-Polar Pixel Array, and Fovea

Figure 5a illustrates the log-polar grid whose name arises from the $(log(r), \theta)$ coordinates which differ from polar coordinates (r, θ) in that the radial coordinate is logarithmic. This pattern is succinctly expressed as the complex exponential (conformal) mapping of a Cartesian grid (figure 5b). This simplifies expression of geometric transformations such as rotation, zoom, and perspectivity by reducing them to arithmetic (vector addition, sign reflection) on complex numbers. In vision, figure 5a corresponds to the image plane pixel layout, and figure 5b to pixel indexed data structure. Rays and rings of pixels map into rows and columns of data. NASA JSC[4] and Transitions Research Corporation[14] have built real time hardware which performs this transformation of coordinates on video data. IMEC[9,10] fabricated a "silicon retina" CCD chip which incorporates physical pixel layout similar to that in figure 5a.



a) Image plane pixel layout

b) Pixel index data structure

Figure 5. Log-polar image plane and mapping

Since pixels cannot become infinitesimally small at the center of the field of view in figure 5a, this region, called the "fovea" after its biological counterpart, must depart from the log-polar pattern. For example, the center can be uniformly filled with pixels the same size as pixels in the inner ring of the log-polar periphery. This is the approach used by IMEC for the world's first log-polar chip[10], illustrated in figure 6. One benefit of this approach is that linear coordinates are used in the center, permitting traditional translation-invariant pattern recognition and uniform neighborhood size. Alternatively, concentric scaled polar coordinates (that is, θ resolution is scaled by r) can be used in the fovea to blend rotation and zoom at the boundary, as illustrated in figure 7.

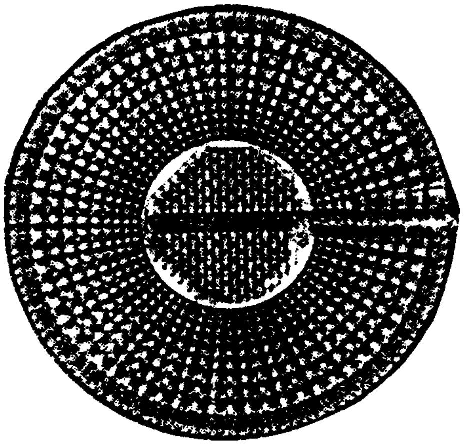


Figure 6. Fovea of IMEC sensor chip

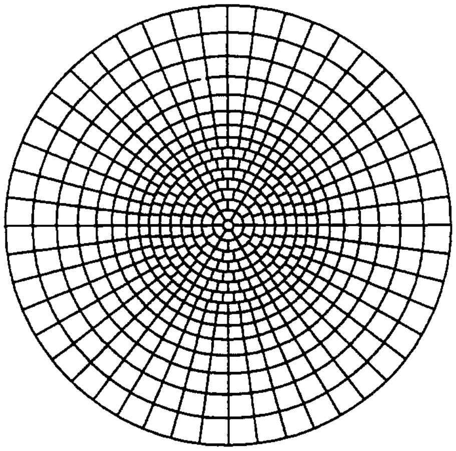
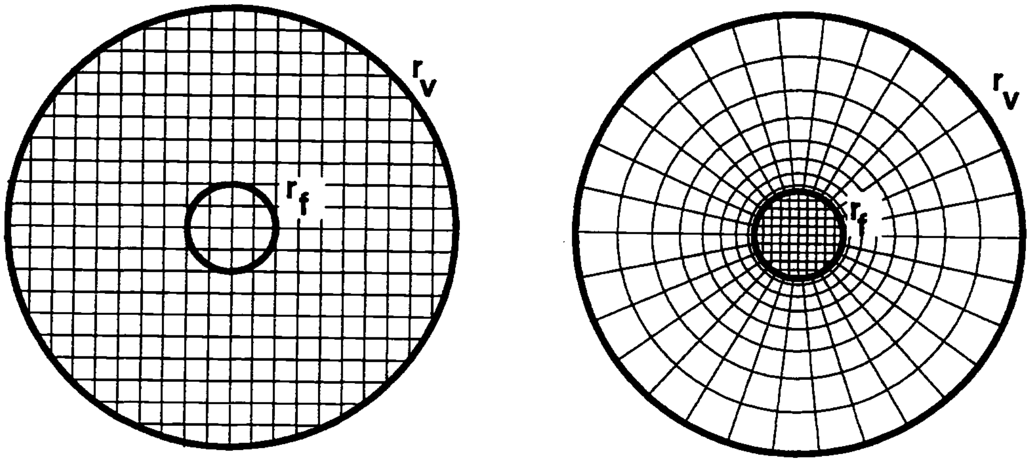


Figure 7. Concentric polar fovea

3.2 Fovea Resolution Ratio and Consequences for Range Resolution

Figure 8a illustrates a uniformly tessellated (Cartesian) disk of radius r_v in pixel units, and figure 8b a log-polar disk of the same radius and same *total* pixel count. That is, for equitable comparison, we constrain the number of pixels to be equal in both "retinas". Figure 8b depicts the fovea in Cartesian tessellation in order to allocate pixel count to disk area; the scaled polar fovea of figure 7 would have approximately the same pixel count, with more harmonious geometry.



a) uniform tessellation

b) log-polar tessellation

Figure 8. Cartesian vs. log-polar pixel tessellation

We now parametrize the concentration C of foveal resolution as the ratio of field-of-view radius r_v to fovea radius r_f in figure 8b,

$$C = \frac{r_v}{r_f} . \tag{2}$$

The number of pixels in the Cartesian retina is then the area of the disk,

$$npix = \pi \, r_v^2 \quad . \tag{3}$$

Now, the log-polar periphery of figure 8b is divided into n rays and q rings where

$$q = \frac{n}{2\pi} \log_e \left(\frac{r_v}{r_f} \right) = \frac{n}{2\pi} \log_e C \tag{4}$$

by virtue of conformality and unit aspect ratio pixels[12]. Thus, equating pixel count in Cartesian and log-polar retinas yields (the rightmost term is foveal pixel count)

$$npix = \pi \, r_v^2 \quad = \quad q \cdot n + \frac{n^2}{4\pi} \quad = \quad \frac{n^2 \log_e C}{2\pi} + \frac{n^2}{4\pi} \tag{5}$$

whence

$$n = \frac{2\pi \, r_v}{\sqrt{1 + 2 \, \log_e C}} \tag{6}$$

Substituting pixel counts into radius equations yields the resolution ratio

$$R_f = \frac{C}{\sqrt{1 + 2 \, \log_e C}} \tag{7}$$

That is, ***equation 7 expresses the ratio of improvement in foveal resolution, and hence improvement in range resolution, for log-polar foveas over Cartesian foveas, for any given pixel count per field of view.*** It is interesting to note that this quantity is *independent of absolute resolution* (total pixel count), depending only on the ratio of field of view to fovea. Figure 9, which plots the value of this function, shows that despite the transcendental appearance of the equation, the relation is nearly linear with respect to C because the denominator changes so slowly. This relation is approximated by a gain in range resolution roughly proportional to the ratio of field of view to fovea, with a constant of proportionality of about 1/3. Thus, foveas 1% the diameter of a log-polar retina (i.e. $C = 100$) yield a 30-to-1 gain in stereo resolution over Cartesian retinas with the same number of pixels.

Surprisingly, the penalty paid in decreased peripheral resolution is proportionally far less than the gain in foveal resolution. Equation 8, plotted in figure 10, gives the ratio of log-polar to Cartesian pixel diameter at the extreme periphery,

$$R_v = \frac{R_f}{C} = \frac{1}{\sqrt{1 + 2 \, \log_e C}} \quad . \tag{8}$$

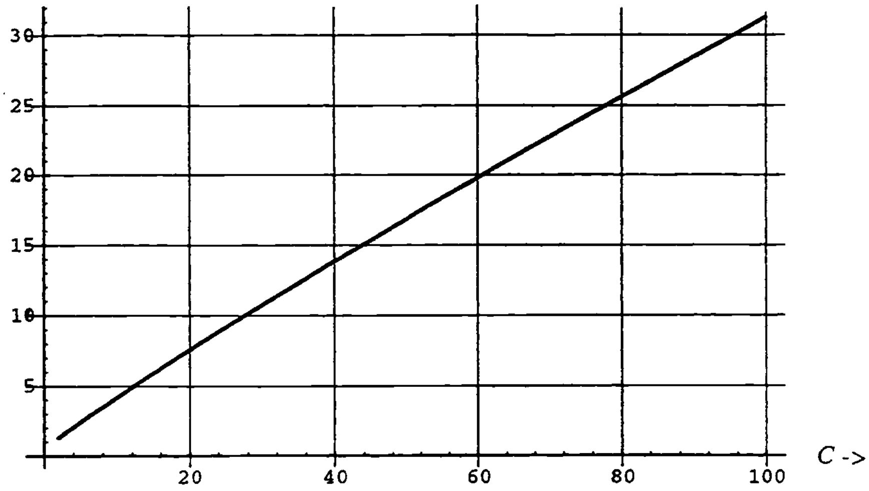


Figure 9. Range resolution ratio per fovea size relative to field of view

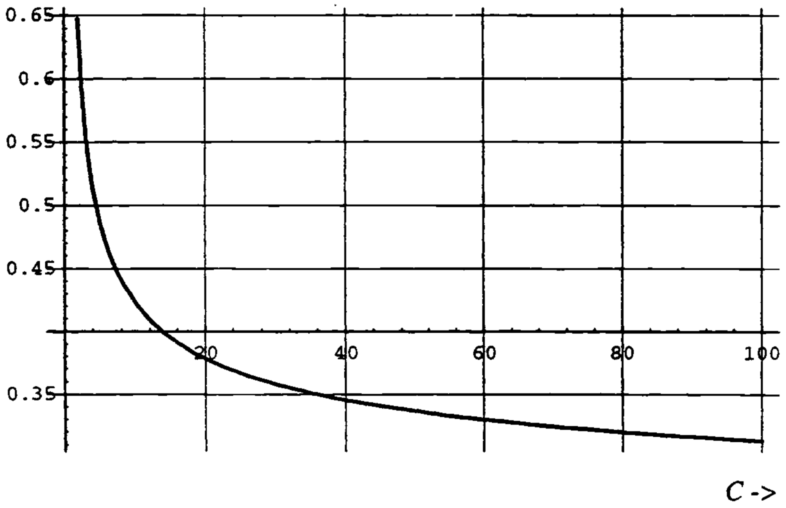


Figure 10. Peripheral loss of resolution

Let us now examine some practical examples using parameters in the range of typical applications. For example, consider a binocular camera system based on 512 pixel diameter ($r_v = 256$) image planes with 90° fields of view. For a baseline B of 25 centimeters, and a vergence distance of 2 meters, equation 1 tells us that voxel aspect ratio is 16-to-1. Pixel cross section is simply range times angular subtense in radians,

$$\frac{\pi}{2\cdot 512}\times 2\ meters = 6mm \tag{9}$$

whence range uncertainty is

$$16\times 6mm = 9.6cm \quad . \tag{10}$$

Now consider a log-polar retina with the same total pixel count in the disk of radius $r_v = 256$ and a fovea whose diameter is 1% of r_v, i.e., $r_f = 26$, or $C = 100$. From equation 7, the density ratio is 31.3-to-1, i. e., better than 30-to-1 improvement in range resolution to 3.26 mm. From equation 6 log-polar ray number n is thus 503. Peripheral resolution is only 3 times worse than a uniform Cartesian grid with the same total pixel count.

Another interesting example is the human eye whose cone distribution (color sensing pixels) can be closely modeled as log-polar, with a value of C close to 100 (fovea diameter is about 1% of retinal diameter) which yields a 31-to-1 improvement in stereo resolution over a uniformly tessellated retina with the same number of photoreceptors. With cone spacing of approximately 30 arc seconds and binocular baseline of 63 millimeters, the geometric length of a voxel is .4 millimeters (400 microns) for close eye-hand coordination (threading a needle, or delicate assembly at a range of 30 centimeters, elbows bent). Voxel width is approximately 42 microns. By way of example, fine thread is about 100 microns in diameter and human hair is in the small multiples of tens of microns.

4. BINOCULAR STEREO DISPARITY FIELDS

The vergence angles $\alpha_L = -\alpha_R$ yield range to a target centered in the fields of view of binocular cameras by simple triangulation,

$$Z = \frac{B}{2\tan\alpha_L} \quad . \tag{11}$$

Stereo perception is not limited to this single point, but should be effective throughout the field of view, including depth of field which includes deviations from zero disparity. We analyze the structure of the global stereo disparity field by slicing it in horizontal and vertical sections below, characterizing the distribution of pixel projection intersections which are the elementary volumetric units of the stereo sensor.

4.1 Horizontal Section Through Voxels

Figure 11 illustrates a horizontal slice through the fields of view at the level of the binocular axis, dissected by the rays projecting from symmetrically converging cameras whose image planes are divided into the uniform pixels of a traditional Cartesian coordinate system. The dissection pattern is a cross-section of the volume cells (voxels) defined by the binocular intersection of the prismatic cones projecting from image plane pixels. The size of voxels indicates the uncertainty of range measurement induced by pixel size. Note that voxels are arranged in concentric curves, but that the highest density (best depth resolution) is nearest the peripheral limits of the intersections of fields of view where disparities are maximum.

Figure 12 illustrates a horizontal slice for convergent cameras with log-polar image plane pixels. Note the high resolution focus of attention at the intersection of the centers of the fields of view. This voxel configuration differs markedly from that of the Cartesian case, which has no such differentiated center.

We can more clearly characterize the contrasting structure of the binocular disparity fields for Cartesian and log-polar retinas by examining the iso-disparity curves which are the moiré patterns visible in figures 11 and 12. Each moiré fringe represents a constant difference in pixel index between left and right viewplane projections [Oster, 1965]. That is, indexing pixels along the horizontal image plane axis (x-axis) in the left and right fields of view as j_L and j_R respectively, a fixed disparity curve is a solution to the indicial equation

$$j_{\mathrm{L}} - j_{\mathrm{R}} = k \tag{12}$$

where k is the disparity measured in pixel count. For example, $k = 0$ corresponds to the zero disparity locus, where pixel indices from both cameras are in exact correspondence. Iso-disparity curves are found by solving the equations for the intersections of rays whose indices satisfy the indicial equation for fixed values of k. Figures 13 and 14 illustrate these fixed-k moiré curves corresponding to the convergent Cartesian cameras of figures 11 and convergent log-polar cameras of figure 12, respectively. The "wave fronts" (which correspond to moiré fringes) emanating from the image planes represent loci of constant disparity, k, satisfying indicial equation 12. Note that in the Cartesian case (figure 13) the best range resolution is at the extreme periphery of the cameras' fields of view. In contrast, note that in the log-polar case, there is high resolution well within both fields of view, and that ***at the 3-D intersection of the binocular fovea projections, range-disparity voxels are nearly uniform and denser than the Cartesian case,*** as a result of the higher resolution allocated to the fovea. The improvement in range resolution is exactly proportional to the increase in resolution of fovea pixels over uniform Cartesian pixels, as described in section 3.2.

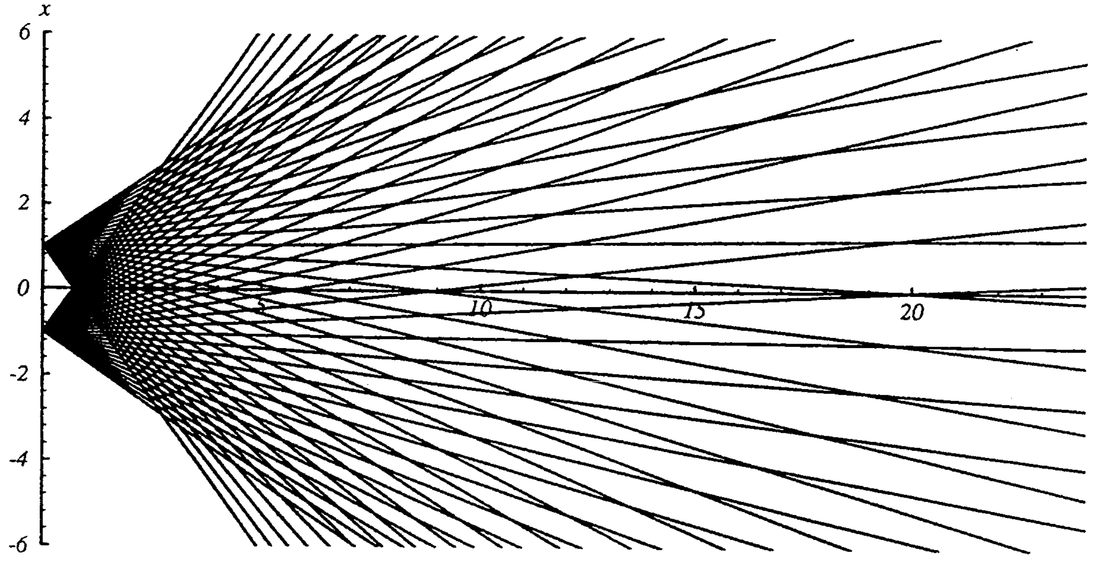


Figure 11. Top view of convergent Cartesian cameras

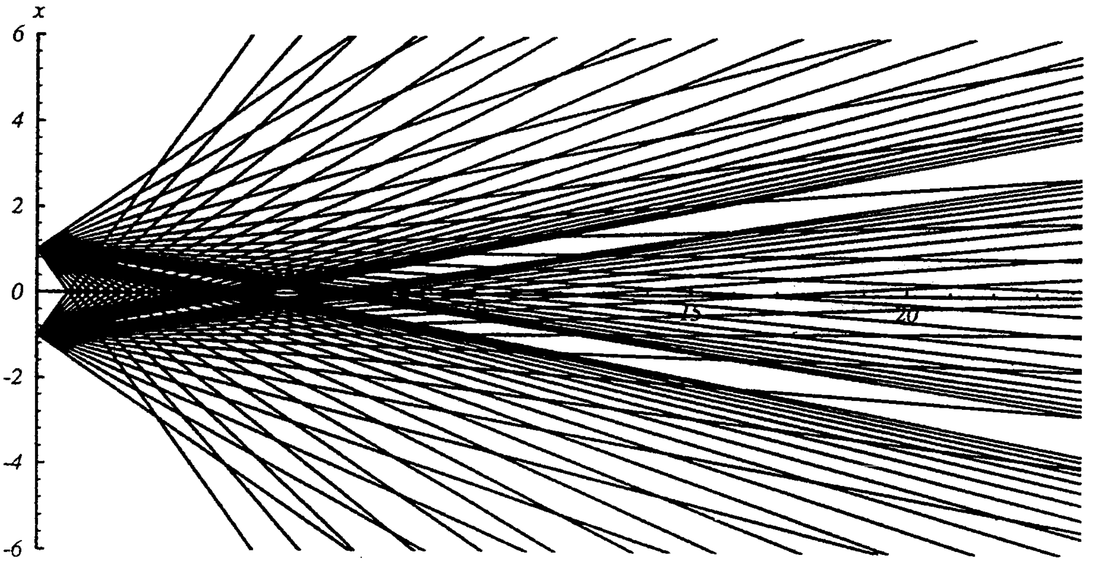


Figure 12. Top view of convergent log-polar cameras

4.2 Vertical Sections Through Voxels

Having examined the horizontal cross-section of the binocular field for symmetrically converging cameras, consider now the vertical longitudinal section which splits the binocular field of view. Figures 15 and 16 illustrate the vertical sections of the voxel field for convergent viewing through Cartesian and log-polar retinas, respectively. In the Cartesian case, note that the highest density (best range resolution) is at the periphery of the fields of view, closest to the cameras. In marked contrast, the log-polar convergent viewing case shown in figure 16 exhibits a well defined high-resolution center at the intersection of the principal viewrays. ***This focus of attention can be steered by active camera controls through the 3-D environment like the intersection of spotlights targeting any selected range and direction.***

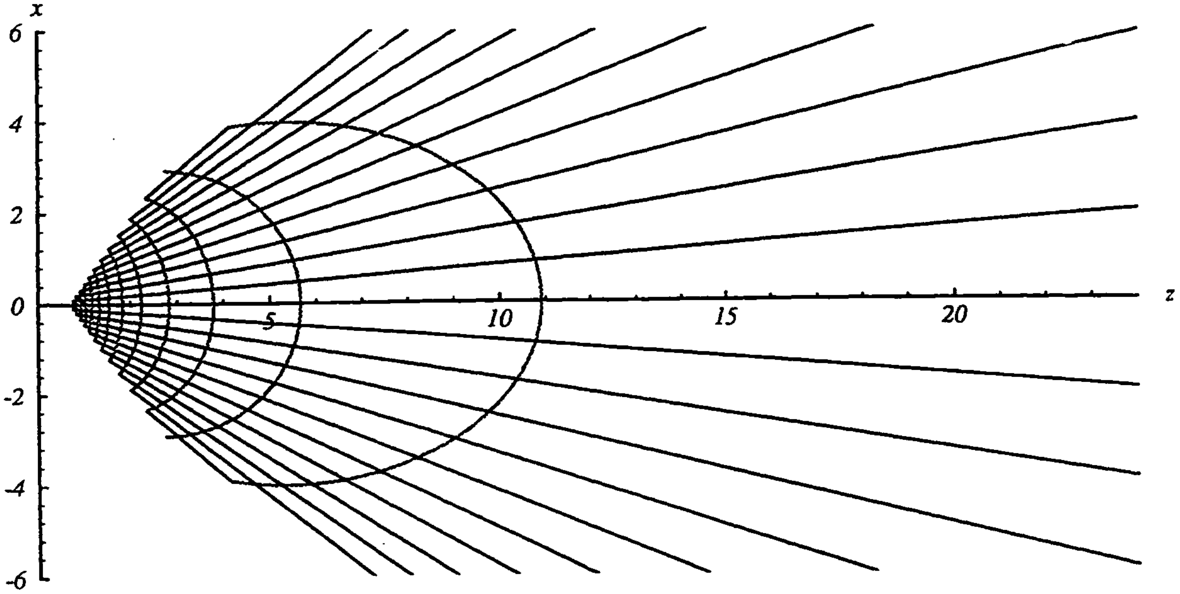


Figure 13. Moiré patterns for convergent Cartesian cameras

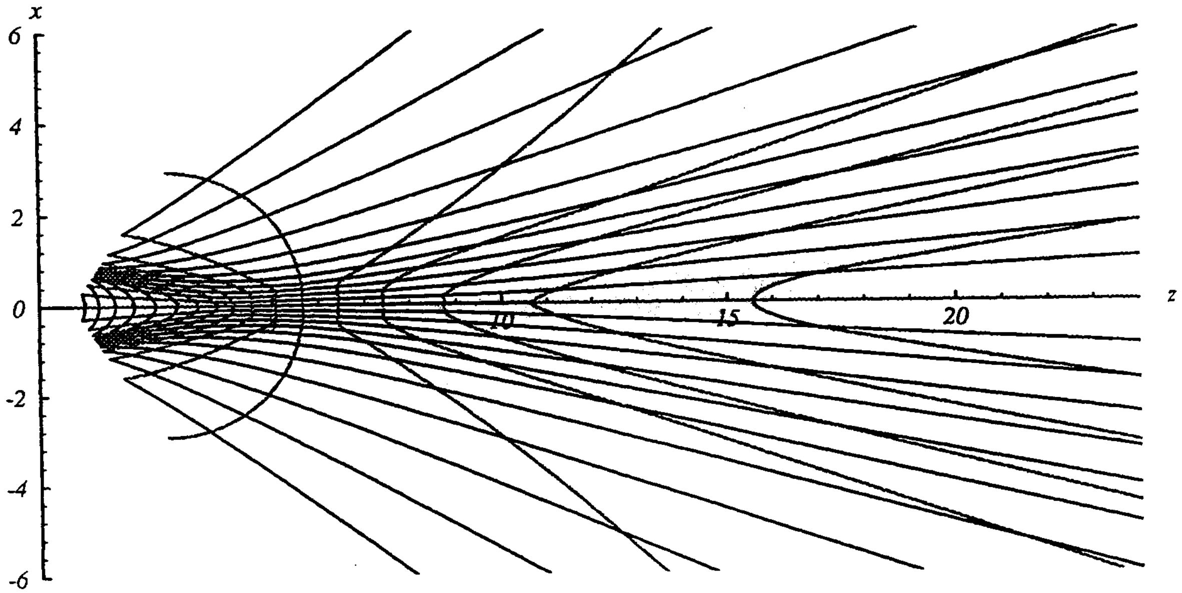


Figure 14. Moiré patterns for convergent log-polar cameras

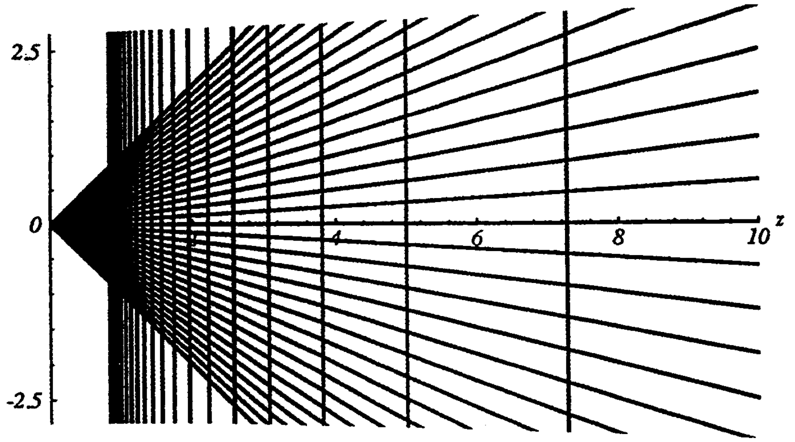


Figure 15. Vertical plane section for convergent Cartesian cameras

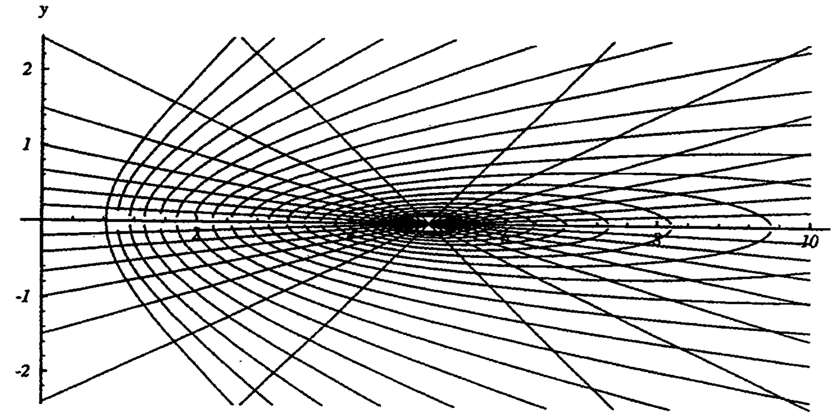


Figure 16. Vertical plane section for convergent log-polar cameras

5. CONCLUSIONS

Log-polar pixel distributions in the image plane concentrate resolution at the center of the field of view. In comparison with traditional uniform (Cartesian) patterns with the same number of pixels, 30-to-1 improvement can be reasonably achieved. For binocular stereo, this increase in image plane resolution translates directly into improvement in range resolution, provided the high resolution projections intersect on the target. Thus, log-polar binocular vision requires "active vision", i.e. motion control of camera direction of gaze. Conversely, the motion capability is symbiotic with foveated vision. Gaze control empowers the system to place and move the binocular focus of attention in three dimensions like the intersection of two searchlights. Within this intersection, stereo resolution is more than an order of magnitude better than that of conventional uniform rasters with comparable pixel populations. Wide fields of view are still available for attention and optic flow mechanisms, with modest sacrifice of peripheral resolution. The human visual system and its mechanical behavior (eye movements) exhibit just such a symbiosis between foveation and tracking.

6. ACKNOWLEDGEMENTS

This work was performed under NASA SBIR Phase II Contract # NAS 9-18637. The author wishes to express sincere appreciation to Timothy E. Fisher, NASA Johnson Space Center Contracting Officer Technical Representative, for his cogent suggestions, constant support, and patience.

7. REFERENCES

1. Aloimonos, Y. "Purposive and Qualitative Active Vision", *10th International Conference on Pattern Recognition,* IEEE, pp. 346-360, June 1990.

2. Bishay, M., Kara, A., Wilkers, D. M., Peters, R. A., and Kawamura, K. "An Active Vision Approach for Locating Salient Features of Objects Using Log-Polar Mapping with No Camera Motion", *Vision Interface '94*, pp. 64-72.

3. Burt, P. J., "Smart sensing with a pyramid vision machine", *Proc IEEE*, Vol 76, pp.1006-1014, August, 1988.

4. Fisher, T. E., and R. D. Juday, "A programmable video image remapper", Proc. SPIE Conf. on Patt. Recognition and Signal Proc, Vol. 938, *Digital and Optical Shape Representation and Pattern Recognition*, pp. 122-128, Orlando, 1988.

5. Griswold, N. C., Lee, J.S., and Weiman, C.F.R., "Binocular Fusion Revisited Utilizing a Log-Polar Tessellation", pp. 421-457 in Computer Vision and Image Processing, Academic Press, 1992.

6. Marr, D. and Poggio, T., "A Computational Theory of Human Stereo Vision", *Proceedings of the Royal Society, London B*, Vol 204, pp. 301-328, 1979.

7. Oster, G., The Science of Moiré Patterns, Edmund Scientific Company, Barrington, New Jersey, 1965.

8. Rojer, A. S. and Schwartz, E. L., "Design Considerations for a Space-Variant Sensor with Complex Logarithmic Geometry", *Proc. 10th International Conference on Pattern Recognition,* Atlantic City, pp. 278-285, 1990.

9. Sandini, G., and P. Dario, "Active vision based on space-variant sensing", in *Proceedings of the 5th International Symposium on Robotics Research,* Tokyo, MIT Press, 1989.

10. Van der Spiegel, J., G. Kreider, C. Claeys, I Debusschere, G. Sandini, P. Dario, F. Fantini, P. Bellutti, G. Soncini , "A Foveated Retina-Like Sensor Using CCD Technology", in Analog VLSI and Neural Network Implementations, C. Mead and M. Ismail, Eds., DeKluwer Pubs, Boston, 1989.

11. Weiman, C. F. R. and G. Chaikin, "Logarithmic Spiral Grids for Image Processing and Display", *Computer Graphics and Image Processing*, (11), pp. 197-226, 1979.

12. Weiman, C. F. R., "Exponential sensor array geometry and simulation", *Proc. SPIE Conf. on Pattern Recognition and Signal Processing*, Vol. 938, *Digital and Optical Shape Representation and Pattern Recognition*, pp. 129-137, Orlando, 1988a.

13. Weiman, C. F. R., "3-D sensing with exponential sensor arrays", *Proc. SPIE Conf. on Pattern Recognition and Signal Processing*, Vol. 938, *Digital and Optical Shape Representation and Pattern Recognition*, Orlando, 1988b.

14. Weiman, C. F. R., "Log Polar Vision for Mobile Robot Navigation", *Electronics Imaging: International Electronic Imaging Exposition and Conference*, pp. 382-386, Oct 1990.

15. Yeshurun, T., and Schwartz, E. L., "Shape Description With a Space Variant Sensor: Algorithms for Scan-Path, Fusion, and Convergence Over Multiple Scans", *IEEE Trans. PAMI*, Vol. 11, no. 11, pp. 1217-1222, November, 1989.