# THE UNIVERSITY OF MICHIGAN

# COMPUTING RESEARCH LABORATORY[1]

---

## AXIAL MOTION STEREO

### Nancy O'Brien and Ramesh Jain

### CRL-TR-11-84

JANUARY 1984

Room 1079, East Engineering Building
Ann Arbor, Michigan 48109
USA
Tel: (313) 763-8000

---

# Axial Motion Stereo[1]

**Nancy O'Brien**

and

**Ramesh Jain**[2]

2. Electrical and Computer Engineering

The University of Michigan

Ann Arbor, MI 48109

## Abstract

This paper presents a new stereo approach. Two or more images of a scene containing stationary objects may be obtained by moving a camera by a known distance along its optical axis. It is shown that the displacement of a point in the *Ego-motion Complex Logarithmic Space* depends only on its depth. This fact can be exploited to obtain depth of stationary surfaces. It was shown earlier that Ego-motion Polar Mapping allows easy segmentation of a dynamic scene. Thus, the approach of this paper, will allow depth recovery as a by-product in a dynamic scene analysis system. In this paper we present the mapping and demonstrate the feasibility of the proposed approach for recovering the depth.

**Index Terms:** Stereo, Axial Motion Stereo, Complex Log Mapping, Ego-motion Complex Log Mapping, Focus of Expansion.

---

## 1. Introduction

Several approaches have been proposed for recovering 3-D information about a scene [BaB82]. Passive methods, based on multiple images of a scene taken from different viewpoints, have remained popular due to their ability to recover the information in an unconstrained environment. The most popular and also the most obvious of these methods is the use of *common stereo*. Two or more images of a scene are obtained from known viewpoints at the same time instant, and the depth of the point is calculated using triangulation. It is also possible to obtain these images at different time instants using one camera. Nevatia [Nev76] used motion stereo to solve the correspondence problem. Bridwell and Huang [Bri83] proposed the use of lateral motion stereo to obtain 3-D information by moving the camera parallel to the image plane.

Schwartz [Sch77,Sch80,Sch81,Sch82] argued that the *retino-striate mapping* can be approximated using a Complex Logarithmic Mapping (CLM). He showed that this mapping is responsible for the scale, rotation and projection invariances in the human visual system. His claims are correct only if the direction of the gaze and the direction of the motion are the same [Cav78,Cav81]. Chaikin and Weiman [ChW79] have pointed out several advantages of the CLM in computer vision systems. The most attractive feature of the mapping is that it is conformal and hence, unlike most other commonly used transformations in image processing, does not loose spatial connectivity of points. One of the most important facts in information recovery from images is that a surface in the real world is mapped into a region in the image. The surface coherence is used in recovering structure of surfaces from the corresponding regions in images. The CLM preserves regions and hence allows recovery of surface structure from the regions in the CLM space.

For the analysis of dynamic scenes acquired using a moving camera, optical flow is attractive. It has been shown that optical flow carries information about the structure of the environment [Clo80,Gib79,Lee80,Pra80]. It appears, however, that the computation of acceptable quality optical flow is a very difficult problem [BrH83]. Jain tried to exploit characteristics of optical flow, without computing the optical flow [Jai83a,Jai83b]. If the motion of the camera is known,

then the *Focus of Expansion (FOE)* can be easily computed. It has been shown that by using the known translational motion of the camera, a transformation, called Ego-Motion Polar (EMP) transformation, can be used to detect moving objects in a scene acquired using a moving camera. Jain showed that the projection invariance can be obtained by using Ego-motion Complex Logarithmic Mapping (ECLM) for arbitrary translational motion of the observer [Jai83c]. It was also shown that the frame-to-frame displacement of a point in the ECLM space depends only on the depth of the point. The fact that all stationary points show only horizontal motion in the ECLM space will be useful in solving the correspondence problem.

In this paper we suggest that 3-D information about stationary objects may be obtained by a controlled motion of a camera along its optical axis. Though we move the camera along its optical axis, in principle any arbitrary translation of camera may be used in the proposed approach. We discuss relevant aspects of CLM and ECLM briefly and show the feasibility of recovering the 3-D information for synthetic scenes. We discuss some aspects of the mapping also and propose a computationally efficient and aesthetically pleasant method for the ECLM.

## 2. CLM and ECLM: Complex Log Mapping

First we will look at the mathematical definition of CLM. Complex log mapping is based on the equation

$$w = \ln z \tag{1}$$

with complex variables

$$z = x + iy = r(\cos\theta + i\sin\theta) = re^{i\theta} \tag{2}$$

and

$$w = u(z) + iv(z) \tag{3}$$

Hence, a function or image in z-space with coordinates $x$ and $y$ can be mapped to w-space with coordinates $u$ and $v$. This mapping results in:

$$u(r,\theta) = \ln r \tag{4}$$

$$v(r,\theta) = \theta \tag{5}$$

An attractive feature of this mapping is the fact that it is the only analytic function which maps a circular region into a rectangular region, which is desirable for the study and modelling of the human visual system. Also, several researchers have shown that rotation of surfaces in the z-space becomes vertical displacement in the w-space, and magnification of a surface in z-space corresponds to a horizontal displacement in w-space. It is through this last aspect that we wish to exploit CLM. If the observer is moving toward its fixation point, this results in a magnification of the stationary objects in the scene and thus a horizontal displacement of the mapped object.

The projection invariance is obtained only if the observer translates in the direction of its gaze. It has been shown [Jai83c], that to exploit this important feature of the CLM in case of arbitrary translation of the observer, the mapping should be obtained with respect to the *FOE*, not with the centre of the image. The mapping with respect to the *FOE* is the ECLM. We will now show mathematically that the depth of a stationary object in an image can be determined using the ECLM. This involves the use of two images of a scene, one initial and one after the observer has moved toward the scene along his line of sight, *i.e.* axial motion stereo.

Let us consider a stationary point $P$ in the environment whose real world coordinates with respect to the observer at a time instant are $(X, Y, Z)$. The projection $(x', y')$ of this point on the image plane, assuming that the projection plane is parallel to the $XY$ plane and is at $Z=1$, is given by

$$x' = \frac{X}{Z} \tag{6}$$

$$y' = \frac{Y}{Z} \tag{7}$$

If the observer moves, then the relationship between the distance $r$ of the projection of the point from the FOE and the distance $Z$ of the point from the observer is

$$\frac{dr}{dZ} = \frac{d\sqrt{x'^2 + y'^2}}{dZ} = -\frac{r}{Z} \tag{8}$$

The complex logarithmic mapping results in

$$\frac{du}{dZ} = \frac{du}{dr} * \frac{dr}{dZ} \tag{9}$$

and

$$\frac{dv}{dZ} = \frac{dv}{d\theta} * \frac{d\theta}{dZ} \tag{10}$$

From equation (4), we get

$$\frac{du}{dr} = \frac{1}{r} \tag{11}$$

Also,

$$\frac{d\theta}{dZ} = \frac{d(\tan^{-1}\frac{y'}{x'})}{dZ} = 0 \tag{12}$$

Therefore, we have

$$\frac{du}{dZ} = -\frac{1}{Z} \tag{13}$$

and

$$\frac{dv}{dZ} = 0 \tag{14}$$

These results show us that the depth $Z$ of a point can be determined from the horizontal change of the ECLM of that point and from the change in distance of the observer from the image. Furthermore, we can see that a change in distance of the observer will result in *only* a horizontal change in the mapping of the image points. There will be no vertical movement of the mapped points and thus correspondence of points between the two stereo pictures will become easier. Assuming that there is sufficient control of the camera to be able to determine the amount of movement, both variables necessary to detemine image depths are readily available.

Another interesting feature of this approach is that, if required, we can obtain more than two frames for resolving ambiguities that cannot be resolved based only on two frames. We have not investigated the use of more than two frames yet.
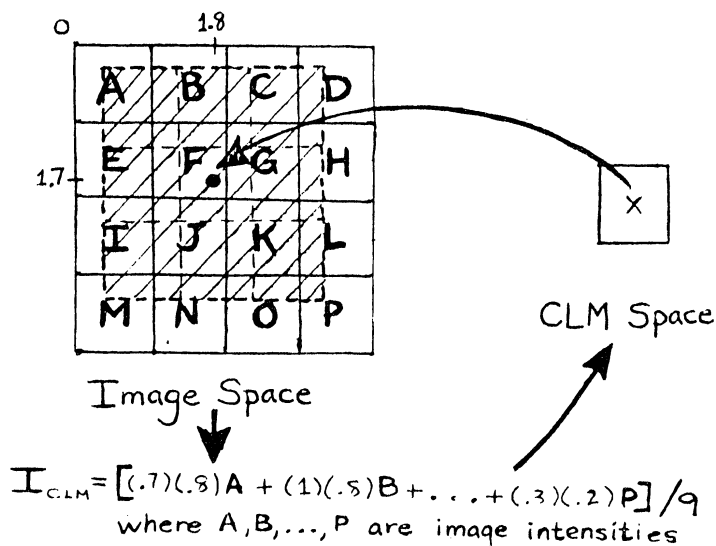
## 3. The Mapping

Mathematically, each point in the image space corresponds to exactly one point in the space transformed through complex logarithmic mapping. However, in computer vision systems where only a finite amount of memory space is available, an image can only be stored as a finite number of pixels and there are only a finite number of intensities representable. This quantization of the

image leads to ambiguity in the mapping, since an image pixel can map to a range of pixels in the transformed space. For example, for a pixel in the first quadrant with coordinates $(x,y)$ at the lower left corner, the $u$ coordinate in the transformed space will range from $\ln\sqrt{x^2+y^2}$ to $\ln\sqrt{(x+1)^2+(y+1)^2}$ and the $v$-coordinate will range from $\tan^{-1}\frac{y}{x+1}$ to $\tan^{-1}\frac{y+1}{x}$. These ranges can be quite wide for points around the origin or practically constant for points far from the origin.

We considered several different interpolations of the image pixels to produce the CLM. One very simple method we examined involved merely computing the range of each image pixel in the mapped space and setting each map pixel in this range to the corresponding image intensity This procedure resulted in a very broken, choppy mapping. We also tried working inversely from the mapped space. The image point corresponding to each pixel in the CLM space was determined and then various interpolations of the intensities of the image pixels around this point were tested. This method of inverse mapping resulted in a much smoother resulting CLM.

|Figure 1| The interpolation scheme used for the ECLM mapping uses the area of pixels that contribute to the intensity at the point in the ECLM space.



$$\mathcal{I}_{c.L.M} = \left[ (.7)(.8)A + (1)(.5)B + \ldots + (.3)(.2)P \right] / q$$
where $A, B, \ldots, P$ are image intensities

The combination of image pixels we found that resulted in the most continuous and pleasing mapping was surprisingly simple. It involved merely adding the intensities of the portions of the

image covered by a three pixel square around the point found from the inverse mapping. An indication of how this worked is shown in Fig. 1. The results of such a mapping are shown in Fig. 2. This method may be refined by assigning weights to the various areas of the square. In the future, if complex logarithmic mapping indeed becomes useful, it can easily be implemented in hardware.

[Figure 2] An image and its ECLM are shown in this figure.



## 4. Experiments

Through our experiments, we attempt to show the feasibility of the use of complex log mapping in depth recovery. Our first desire was to verify that the equations which have been derived actually work. For this, we selected random points in space, and calculated the projections of these points on the image plane perpendicular to the line of sight, which we assumed to be along the z-axis, at unit distance from the observer. From these projections we calculated the CLM for the points. Then we moved the observer back a unit distance, refigured the projections on the image plane, and calculated the new CLM. The horizontal $u$ displacement was easily figured and the inverse of this (because $dZ=1$) gave the depth value. A factor of $\dfrac{dZ}{2}$ was subtracted because the calculated depth was the average of the original depth and the depth after the observer moved back. Some results from this are in Table 1. The variable $K$ in the table is the amount

by which the calculated depth differed from the actual. We also see that, as expected, there is no vertical $v$ displacement involved in axial observer motion.

Table 1
Real Points

| Point | U1 | V1 | U2 | V2 | Z | K |
|-------|------|--------|--------|--------|------|--------|
| (12.0,9.4,7.3) | 0.763 | 38.07 | 0.608 | 38.07 | 7.3 | -0.011 |
| (5.6,14.9,28.1) | -0.568 | 69.40 | -0.603 | 69.40 | 28.1 | -0.003 |
| (-3.1,-8.4,10.8) | -0.187 | 69.74 | -0.276 | 69.74 | 10.8 | -0.007 |
| (-30.7,56.4,68.2) | -0.060 | -61.44 | -0.075 | -61.44 | 68.2 | -0.001 |
| (43.2,13.7,75.3) | -0.508 | 17.60 | -0.521 | 17.60 | 75.3 | -0.001 |

The next natural step was to attempt this same procedure with digitized points. These points actually now have an intrinsic associated size and therefore *exact* recovery of the depth is impossible. In the results shown here, all image pictures and complex logarithmic mappings are assumed to be 512×512 pixels. Test results are shown in Table 2. Again we see that there is no vertical displacement involved in the ECLM, therefore correspondence of the points from one stereo picture map to the next is extremely simple.

Table 2
Digitized Points

| Point | X1 | Y1 | U1 | V1 | X2 | Y2 | U2 | V2 | Recovered Point |
|-------|------|------|-----|------|-----|------|-----|------|-----------------|
| (1.5,1.5,2.0) | 192 | 192 | 487 | 128 | 128 | 128 | 452 | 128 | (1.5,1.5,2.0) |
| (1.2,1.3,2.5) | 123 | 133 | 452 | 134 | 88 | 95 | 423 | 134 | (1.2,1.3,2.5) |
| (-1.7,2.7,4.0) | -108 | 173 | 462 | -164 | -86 | 138 | 442 | -164 | (-1.6,2.6,3.8) |
| (3.4,-2.9,4.5) | 193 | -164 | 481 | -114 | 158 | -134 | 463 | -114 | (3.3,-2.8,4.3) |

Rather than work with arbitrary points in this experiment, we chose a number of coplanar points. From the data we entered about the image projections of the points and the depth information we were able to calculate through axial motion stereo, we found that we could determine, to a good degree of accuracy, the equation of the plane on which the points lie. A least squares algorithm was used to determine the plane equations. The equation corresponding to the coplanar points of Table 2 is

$$z = -x-y+5$$

From the points we determined with the axial motion stereo we were able to obtain the equation

$$z = -0.9x-0.9y+4.8$$

## 6. Further Research

The above experiments gave results for very simple cases. We are in the process of performing experiments using curved surfaces. We want to exploit surface coherence for determining the nature of the surface. For a given surface, the mapped regions in two frames in the ECLM space may have different shapes, depending on the nature of that surface, but will be displaced only horizontally. By detecting edges in two frames, we can find the depth of the boundary points of the surface. These points may be used in an interpolation scheme to find the depth at the interior points, using surface smoothness. If the nature of the surface is known, then we can use the least squares method to find the orientation and depth of the surface.

We plan to apply this approach to find the depth of surfaces in real world scenes. As shown in [Jai83c], if the camera motion is known then an ego-motion polar transform may be used to segment a scene. If we use ECLM then we may segment a dynamic scene into stationary and nonstationary components and may obtain the depth of stationary surfaces using the approach proposed in this paper. We are studying the plausibility of recovering 3-D motion parameters of a point from the rate of displacement of its projection in the ECLM space.

In many applications, such as robotics, where either the camera can be under computer control or the motion parameters of the camera may be obtained, the proposed approach for stereo will play very important role. A sequence obtained using a single moving camera may be used for segmentation of the scene, for determination of the motion parameters of objects, and for the determination of the depth of stationary objects. We are investigating the plausibility of such an integrated information recovery approach.

## 6. Conclusion

This paper presents a new method to obtain the depth of points. By moving a camera along its optical axis by a known amount, it is possible to obtain the depth of points using the ECLM proposed in this paper. Since views are obtained by moving a camera, if there are some ambiguities (maybe in correspondence), then more frames may be obtained to resolve them. Another advantage of this method is the fact that computations are performed in the CLM. As has been shown by several researchers [CaP77, ChW79, SaT80, Saw72, SWC81], the CLM space offers several advantages such as scale and rotation invariance, and a possibility of iconic processing. Thus, by working in the ECLM space, we can exploit those features, in addition to obtaining the depth using the proposed approach.

In this paper we presented a method for the transformation from image plane to the ECLM space using interpolation. It seems easy to put this transformation in hardware. In fact, such a device is being constructed for the National Bureau of Standardss [Ken83].

# REFERENCES

[BaB82]

Ballard, D.H. and C. M. Brown, *Computer Vision*, Prentice Hall, 1982.


[BGT79]

Braccini, C., G. Gamberdella, and V. Tagliasco, "A model of the early stages of human visual system," *Biological Cybernetics*, 44, 1982, pp.47-88


[BrH83]

Bruss, A.R. and B.K.P. Horn, "Passive Navigation", *Computer Vision, Graphics, and Image Processing*, Vol.21, 1983.


[Bri83]

Bridwell, N.J. and T.S.Huang, "A discrete spatial representation for lateral motion stereo," *Computer Vision, Graphics, and IMage Processing*, Vol. 21, pp.33-57, 1983.


[CaP77]

Casasent, D. and D. Psaltis, "New optical transforms for pattern recognition," *Proc. of IEEE*, vol. 65, pp.77-84, 1977.


[Cav78]

Cavanagh, P., "Size and position invariance in the visual system," *Perception*, vol. 7, pp.167-177, 1978.


[Cav81]

Cavanagh, P., "Size invariance: reply to Schwartz," *Perception*, col. 10, pp.469-474, 1981.


[ChW79]

Chaikin, G. and C. Weiman, "Log spiral grids in computer pattern recognition", *Computer*

*Graphics and Pattern Recognition,* vol.4, pp.197-226, 1979.

[Clo80]

Clocksin, W.F., "Perception of surface slant and edge labels from optical flow: A computational approach," *Perception,* vol. 9, 1980, pp.253-269.

[Gib79]

Gibson, J.J., *The ecological approach to visual perception,* Houghton Mifflen, Boston, 1979.

[Jai83a]

Jain, R., "Direct computation of the focus of expansion", *IEEE Trans. on PAMI,* vol.PAMI-5, pp.58-64, 1983.

[Jai83b]

Jain, R. "Segmentation of frame sequences obtained by a moving observer," *PAMI,* also to appear in *IEEE Trans.*

[Jai83c]

Jain, R., "Complex Logarithmic Mapping and the Focus of Expansion", SIGGRAPH/SIGART Workshop on MOTION: Representation and Perception, Toronto, April 1983.

[Ken83]

Kent, E., Personal Communication.

[Lee80]

Lee, D.N., "The optic flow field: The foundation of vision," *Phil. Trans. Royal Society of London,* vol. B290, 1980, pp. 169-179.

[Nev76]

Nevatia, R., "Depth Measurement by Motion Stereo", *CVGIP*, Vol.9, pp.203-214, 1976.


[Pra80]

Prazdny, K., "Egomotion and relative depth map from optical flow," *Biological Cybernetics*, vol. 36, 1980, pp. 87-102.


[SaT80]

Sandini, G and V. Tagliasco, "An anthromomorphic retin-like structure for scene analysis," *Computer Graphics and Image Processing*, vol.14, pp.365-372, 1980.

[Saw72]

Sawchuk, A. A., "Space-variant image motion degradation and restoration", *Proc. IEEE*, vol. 60, pp.854-861, 1972.


[Sch77]

Schwartz, E. L., "The development of specific visual connections in the monkey and goldfish: Outline of a geometric theory of receptotopic structure", *J. Theooretical Biology*, vol 69, pp.655-683.


[Sch80]

Schwartz, E. L., "Computational anatomy and functional architecture of striate cortex: a spatial mapping approach to coding," *Vision Research*, 20, 1980, pp.645-669.


[Sch81]

Schwartz, E. L., "Cortical anatomy, size invariance, and spatial frequency analysis," *Perception*, vol.10, pp.455-468, 1981.


[Sch82]

Schwartz, E.L., "Columnar architecture and computational anatomy in primate visual cortex: Segmentation and feature extraction via spatial frequency coded difference mapping,"

*Biological Cybernetics,* vol. 42, pp.157-168, 1982.


[SWC81]

Schenker, P.S., K.M. Wong, and E.G.Cande "Fast adaptive algorithms for low-level scene analysis: Application of polar exponential grid (PEG) representation to high-speed, scale-and-rotation invariant target segmentation", *Proc. SPIE,* Vol. 281, Techniques and Applications of Image Understanding, pp.47-57, 1981.