JPL Publication 87-1, Rev. 1

Stereo Depth Distortions in Teleoperation

Daniel B. Diner
Marika von Sydow

May 15, 1988

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No. 87-1, Rev.1	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Stereo Depth Distortions in Teleoperation		5. Report Date May 15, 1988 6. Performing Organization Code
7. Author(s) D. B. Diner and M. von Sydow		8. Performing Organization Report No. JPL PUB 87-1,Rev.1
9. Performing Organization Name and Address JET PROPULSION LABORATORY California Institute of Technology 4800 Oak Grove Drive Pasadena, California 91109		10. Work Unit No. 11. Contract or Grant No. NAS7-918 13. Type of Report and Period Covered JPL Publication
12. Sponsoring Agency Name and Address NATIONAL AERONAUTICS AND SPACE ADMINISTRATION Washington, D.C. 20546		14. Sponsoring Agency Code RE 159 BK-549-02-51-01-00

15. Supplementary Notes

16. Abstract

In teleoperation, a typical application of stereo vision is to view a work space located short distances (1 to 3 meters) in front of the cameras. The work presented in this report treats converged camera placement and studies the effects of intercamera distance, camer-to-object viewing distance, and focal length of the camera lenses on both stereo depth resolution and stereo depth distortion. While viewing the fronto-parallel plane 1.3 meters in front of the cameras, we have measured depth errors on the order of 2 centimeters.

A geometric analysis was made of the distortion of the fronto-parallel plane of convergence for stereo TV viewing. The results of the analysis were then verified experimentally. The objective was to determine the optimal camera configuration which gave high stereo depth resolution while minimizing stereo depth distortion.

We find that for converged cameras at a fixed camera-to-object viewing distance, larger intercamera distances allow higher depth resolutions, but cause greater depth distortions. Thus with larger intercamera distances, operators will make greater depth errors (because of the greater distortions), but will be more certain that they are not errors (because of the higher resolution).

The analysis predicts camera configurations and a camera motion strategy that minimize stereo depth distortion without sacrificing stereo depth resolution.

17. Key Words (Selected by Author(s))	18. Distribution Statement
Teleoperation Stereo Depth Distortion	Unclassified, unlimited distribution

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified		

HOW TO FILL OUT THE TECHNICAL REPORT STANDARD TITLE PAGE

Make items 1, 4, 5, 9, 12, and 13 agree with the corresponding information on the report cover. Use all capital letters for title (item 4). Leave items 2, 6, and 14 blank. Complete the remaining items as follows:

3. Recipient's Catalog No. Reserved for use by report recipients.

7. Author(s). Include corresponding information from the report cover. In addition, list the affiliation of an author if it differs from that of the performing organization.

8. Performing Organization Report No. Insert if performing organization wishes to assign this number.

10. Work Unit No. Use the agency-wide code (for example, 923-50-10-06-72), which uniquely identifies the work unit under which the work was authorized. Non-NASA performing organizations will leave this blank.

11. Insert the number of the contract or grant under which the report was prepared.

15. Supplementary Notes. Enter information not included elsewhere but useful, such as: Prepared in cooperation with... Translation of (or by)... Presented at conference of... To be published in...

16. Abstract. Include a brief (not to exceed 200 words) factual summary of the most significant information contained in the report. If possible, the abstract of a classified report should be unclassified. If the report contains a significant bibliography or literature survey, mention it here.

17. Key Words. Insert terms or short phrases selected by the author that identify the principal subjects covered in the report, and that are sufficiently specific and precise to be used for cataloging.

18. Distribution Statement. Enter one of the authorized statements used to denote releasability to the public or a limitation on dissemination for reasons other than security of defense information. Authorized statements are "Unclassified—Unlimited," "U. S. Government and Contractors only," "U. S. Government Agencies only," and "NASA and NASA Contractors only."

19. Security Classification (of report). NOTE: Reports carrying a security classification will require additional markings giving security and downgrading information as specified by the Security Requirements Checklist and the DoD Industrial Security Manual (DoD 5220.22-M).

20. Security Classification (of this page). NOTE: Because this page may be used in preparing announcements, bibliographies, and data banks, it should be unclassified if possible. If a classification is required, indicate separately the classification of the title and the abstract by following these items with either "(U)" for unclassified, or "(C)" or "(S)" as applicable for classified items.

21. No. of Pages. Insert the number of pages.

22. Price. Insert the price set by the Clearinghouse for Federal Scientific and Technical Information or the Government Printing Office, if known.

Jet Propulsion Laboratory

Interoffice Memorandum

June 14, 1988

To: Distribution

From: Daniel Diner

Subject: JPL Publication 87-1, Rev. 1

Please replace your copy of JPl Publication 87-1 with Publication 87-1, Rev. 1 and return the discarded copy to Dan Diner, M/S 278.

JPL Publication 87-1, Rev. 1

Stereo Depth Distortions in Teleoperation

Daniel B. Diner
Marika von Sydow

May 15, 1988

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

The research described in this publication was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

ABSTRACT

In teleoperation, a typical application of stereo vision is to view a work space located short distances (1 to 3 meters) in front of the cameras. The work presented in this report treats converged camera placement and studies the effects of intercamera distance, camera-to-object viewing distance, and focal length of the camera lenses on both stereo depth resolution and stereo depth distortion. While viewing the fronto-parallel plane 1.3 meters in front of the cameras, we have measured depth errors on the order of 2 centimeters.

A geometric analysis was made of the distortion of the fronto-parallel plane of convergence for stereo TV viewing. The results of the analysis were then verified experimentally. The objective was to determine the optimal camera configuration which gave high stereo depth resolution while minimizing stereo depth distortion.

We find that for converged cameras at a fixed camera-to-object viewing distance, larger intercamera distances allow higher depth resolutions, but cause greater depth distortions. Thus with larger intercamera distances, operators will make greater depth errors (because of the greater distortions), but will be more certain that they are not errors (because of the higher resolution).

The analysis predicts camera configurations and a camera motion strategy that minimize stereo depth distortion without sacrificing stereo depth resolution.

FOREWORD

This report originally appeared in the Proceedings of the Twenty-Second Annual Conference on Manual Control, AFWAL-TR-86-3093, Wright-Patterson AFB Aeronautical Labs, Ohio, USA, 1986.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

ALIGN	horizontal distance between the camera convergence point and the center point between two test bars
C	camera convergence point
D	camera-to-object viewing distance
dL	distance between the camera Vieth-Mueller circle and the left test bar
dR	distance between the camera Vieth-Mueller circle and the right test bar
f	focal length of the TV camera lenses
FPP	fronto-parallel plane including the camera convergence point
ICD	intercamera distance
ITD	intertarget distance
k	horizontal distance from the camera convergence point to a point in the fronto-parallel plane
L	distance from first nodal point of the TV camera lens to the camera convergence point
Lf'	projection of distance k on the line of equidistant projection for far left camera
Ln'	projection of distance k on the line of equidistant projection for near left camera
MTERM	multiplicative product of ICD and ALIGN
Pl'	projection of distance k on the line of equidistant projection for the left camera
Pr'	projection of distance k on the line of equidistant projection for the right camera
Rf'	projection of distance k on the line of equidistant projection for far right camera
Rn'	projection of distance k on the line of equidistant projection for near right camera
w	one-half of the intercamera distance
WP	width per pixel at the CCD camera

1. INTRODUCTION

In teleoperation, one typical application of stereo vision is the viewing of a work space located 1 to 3 meters away from the cameras. We have investigated such close stereo viewing and, over the range of parameters tested, we have explored the trade-off between stereo depth resolution and stereo depth distortion as a function of camera configuration.

When selecting a stereo camera configuration, it is necessary to choose between parallel and converged camera configurations. Parallel configurations, which may have certain advantages for far stereo viewing, have inherent undesirable aspects for near stereo viewing. First of all, the two views of the cameras do not overlap entirely in the work space. Thus some of the image on the monitor screen will not be presented in stereo. Second, an object located exactly in front of the stereo camera system will be seen to the left of center by the right camera, and to the right of center by the left camera. This may force uncomfortable viewing conditions upon the observer, and may reduce performance drastically.

For this reason, we have focused our attention on converged camera configurations. Properly converged camera configurations do not suffer either of the undesirable aspects mentioned above.

However, converged camera configurations can induce stereo depth distortion. For example, with widely converged cameras, an observer stereoscopically viewing a meter stick (located in the fronto-parallel plane including the camera convergence point) reports that the meter stick appears to be curved away from the observer. As the intercamera distance is decreased, and thus the camera convergence angle is decreased, the apparent curvature of the meter stick decreases, but with a loss of stereo depth resolution. This distortion/resolution trade-off is the subject of this report.

This distortion changes with intercamera distance, viewing distance, and focal length of the camera lenses. Unfortunately, for a fixed viewing distance, widely converged camera configurations, which yield higher stereo depth resolution, also yield larger stereo depth distortions.

Camera configurations which are similar to natural human viewing conditions are called orthostereoscopic; unnaturally wide camera separation configurations are called hyperstereoscopic. In the literature on stereo imaging, some researchers advocate orthostereoscopic camera alignments, and other researchers advocate hyperstereoscopic camera alignments.

Shields, Kirkpatrick, Malone and Huggins (1) found no gain in performance with hyperstereopsis on a stereo depth comparison task, and recommended orthostereopsis. This does not surprise

us, as the depth distortion of hyperstereopsis may well have overridden the advantage of the increased depth resolution.

Grant, Meirick, Polhemus, Spencer, Swain, and Tewell (2) found no gain in performance with hyperstereopsis on a peg-in-hole task, and recommended orthostereopsis. This result does surprise us, in that a peg-in-hole task requires high depth precision only in a small region of the work space. The depth distortion of hyperstereopsis only becomes significant for objects which are separated horizontally. Thus the performance of the insertion of the peg into the hole should increase with the increased depth resolution of hyperstereopsis. Perhaps the depth distortions hurt the performance of the long range motions (such as moving towards the peg and moving the peg towards the hole) enough to overshadow the increase in performance of the insertions.

Upton and Strother (3) reported that hyperstereopsis greatly enhanced depth detection of camouflaged buildings from helicopter-mounted stereo cameras. This result is expected. The critical point here is that the accurate detection of depth is a different phenomenon from the accurate estimate of the magnitude of a true depth. Hyperstereopsis artificially magnifies the perceived magnitude of a true depth difference, making that depth difference easier to detect, but much harder to perform accurate teleoperation upon. For example, hyperstereopsis might make a one-story camouflaged building appear to be four stories tall.

Zamarian (4) reported that hyperstereopsis improved performance over orthostereopsis on a three-bar depth adjustment task. He used converged cameras. The three-bar depth adjustment task insures that the depth distortions will play a role in his experiment. He states, "...it was found that performance improved with increasing [camera] separation but at a decreasing rate of improvement." We suspect that he was experiencing the trade-off between increased resolution and distortion.

Pepper, Cole, and Spain (5) reported that hyperstereopsis improved performance on a two-bar depth adjustment task. They used parallel camera configurations, and therefore introduced no stereo depth distortions. These results, therefore, should not apply directly to our work.

Spain (6) reported that hyperstereopsis improved performance on a two-bar depth adjustment task. He converged the cameras so that the camera convergence point was half-way between the two bars when the bars were located at equal depth. We feel that each bar experienced the same depth distortion. The net effect then would have been that the relative distortion between the two bars cancelled out. In that case, the increased stereo depth resolution of hyperstereopsis would have improved performance.

Bejczy (7) reported surprisingly poor performance with a stereo TV viewing system of a task which required the positioning and orienting of an end-effector in an almost static visual scene. Operators were required to pick up one block and place it upon another block. Although the thrust of this work was to evaluate the effect of short-range proximity sensors in conjunction with mono and stereo camera systems on the performance of this task, the surprisingly poor performance with stereo viewing must be noted.

In reviewing the literature, we noticed that most analyses of stereo TV viewing use small angle approximations. However, the actual stereo distortion of the fronto-parallel plane of convergence is such that small angle approximations obscure the relationship between this distortion and the key parameters of the camera configurations.

To investigate this question more rigorously, we have used a geometric analysis of the distortion of the fronto-parallel plane of convergence (FPP) for stereo TV viewing, without any small angle approximations.

This report explores the following question. Will human observers' responses follow the predictions of our geometric analysis, despite internal perceptual corrections and/or distortions? If so, we may use our geometric analysis to predict optimal camera configurations, which can then be tested and verified. We wish to find camera configurations which give high stereo depth resolution without large stereo depth distortions.

This is not a trivial question. We humans surely have perceptual corrections and distortions. Each time we converge our eyes on a flat wall, for example, we experience similar distortions to those described above for converged cameras. We should therefore perceive flat walls as curved away from us. The fact that, in general, we do not, indicates the existence of these corrections and distortions. However, the distortions and corrections may not be so powerful as to negate the predictions of our geometric analysis.

Our ultimate goal is to determine the best trade-off between stereo resolution and distortion per performance task, for work spaces limited to 3 meters depth. A necessary first step is to minimize all non-stereo depth cues. Then we can measure how the observers react to the stereo depth distortion cues in the absence of other possible interfering cues. Once we understand the factors determining the optimal stereo camera configuration for each specific task, we plan to integrate this understanding into experimentation involving visual scenes rich in the other depth cues.

2. GEOMETRIC ANALYSIS

Most geometric analyses of the stereo camera system use small angle approximations, which, as previously noted, obscure the relationship between the stereo depth distortion and the key parameters of the camera configuration. Therefore, we have made a geometric analysis of the distortion of the fronto-parallel plane of convergence (FPP), without using small angle approximations. For the derivation, see Appendix 1.

This analysis predicts distortions for converged camera configurations, but not for parallel camera configurations. Figure 1 shows that parallel cameras, when viewing two objects separated by a horizontal distance, k, will see the same distance between the objects. That is, Pl' = Pr'. Therefore, no stereo depth distortion will be produced by the camera geometry.

In contrast, consider the converged camera configuration in Figure 1, viewing the same two objects where one object is now located at the camera convergence point. The left camera will see a greater distance between the two objects than the right camera. That is, Pl' > Pr'. Therefore the two cameras will present different distances between the two objects to the monitor. We call the difference between the distances on the monitor the spatial monitor disparity between the two camera images. The stereo system presents the left camera image to the left eye, and the right camera image to the right eye. Figure 2 shows that if the eyes see different distances between two objects, the objects will be perceived at different depths.

Static Depth Distortions

Figure 3 shows the nature of the static stereo depth distortions. By static, we mean the distortion that is present when we do not move the cameras. It stems from the camera alignment geometry.

In a quantized TV system, the spatial monitor disparity can be analyzed as the number of pixels difference between the two camera images. The quantized TV system separates space into regions within which motion is invisible. Figure 3 represents two CCD cameras converged and viewing a work space. Each diamond-like shape, which we shall call a lozenge, represents the region in space that is seen by a pair of pixels, one on each camera. If a point source of light is moved within a lozenge, no change will be registered by the TV cameras. The stereo depth resolution will be defined by the lozenge size. Specifically, an object must move at least half a lozenge length in depth for any change to be registered. The stereo depth distortion of the FPP can be understood as the difference in spatial monitor disparity of the various points on the plane. The camera convergence point, which is on the FPP, has zero spatial monitor

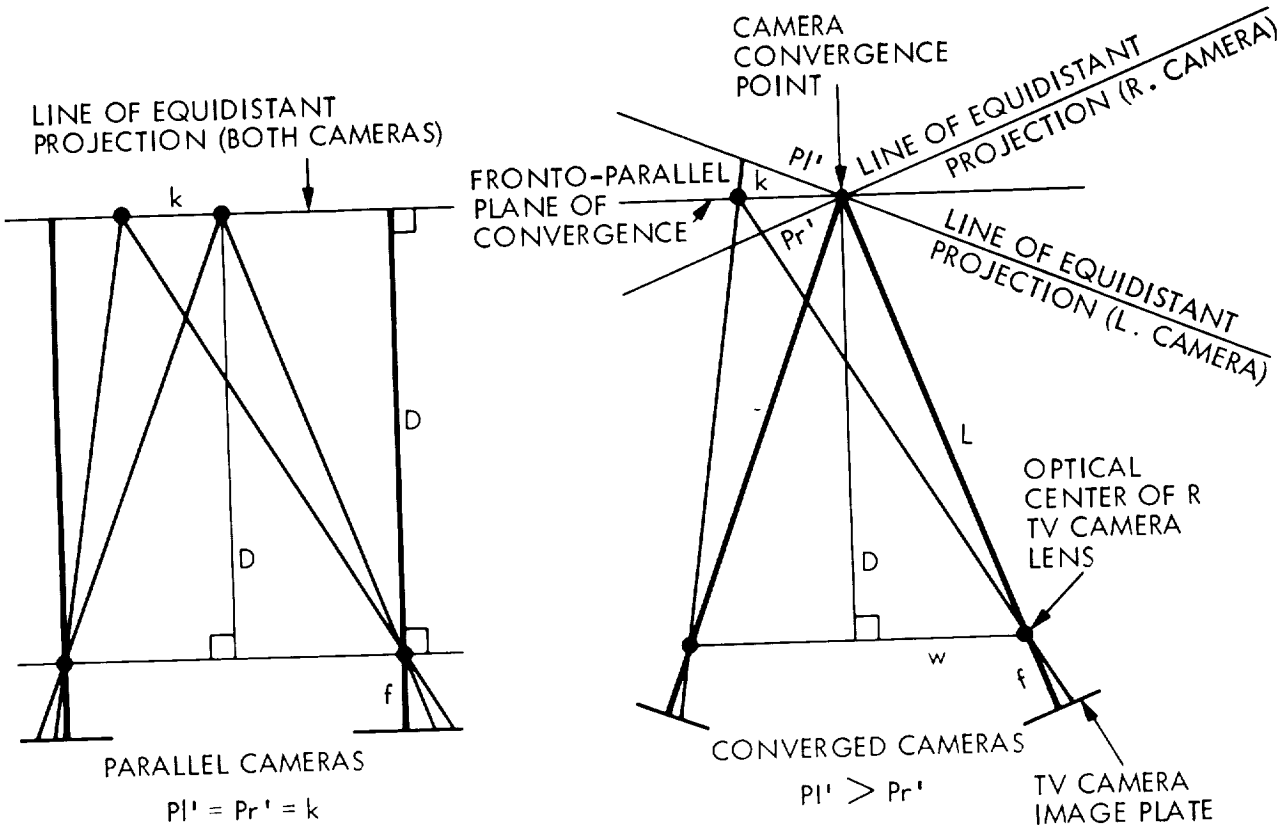

Figure 1. The geometry of parallel and converged CCD camera configurations. On the lines of equidistant projection, every pixel sees a unit length segment. This segment length is (D/f) * (width/pixel at CCD) for the parallel cameras, and (L/f) * (width/pixel at CCD) for the converged cameras. The # pixels difference presented to the monitor by the two cameras will be proportional to (Pl' - Pr'). Consider an object located a horizontal distance k from the camera convergence point. For converged cameras, Pl' > Pr', while for parallel camera configurations, Pl' = Pr'.

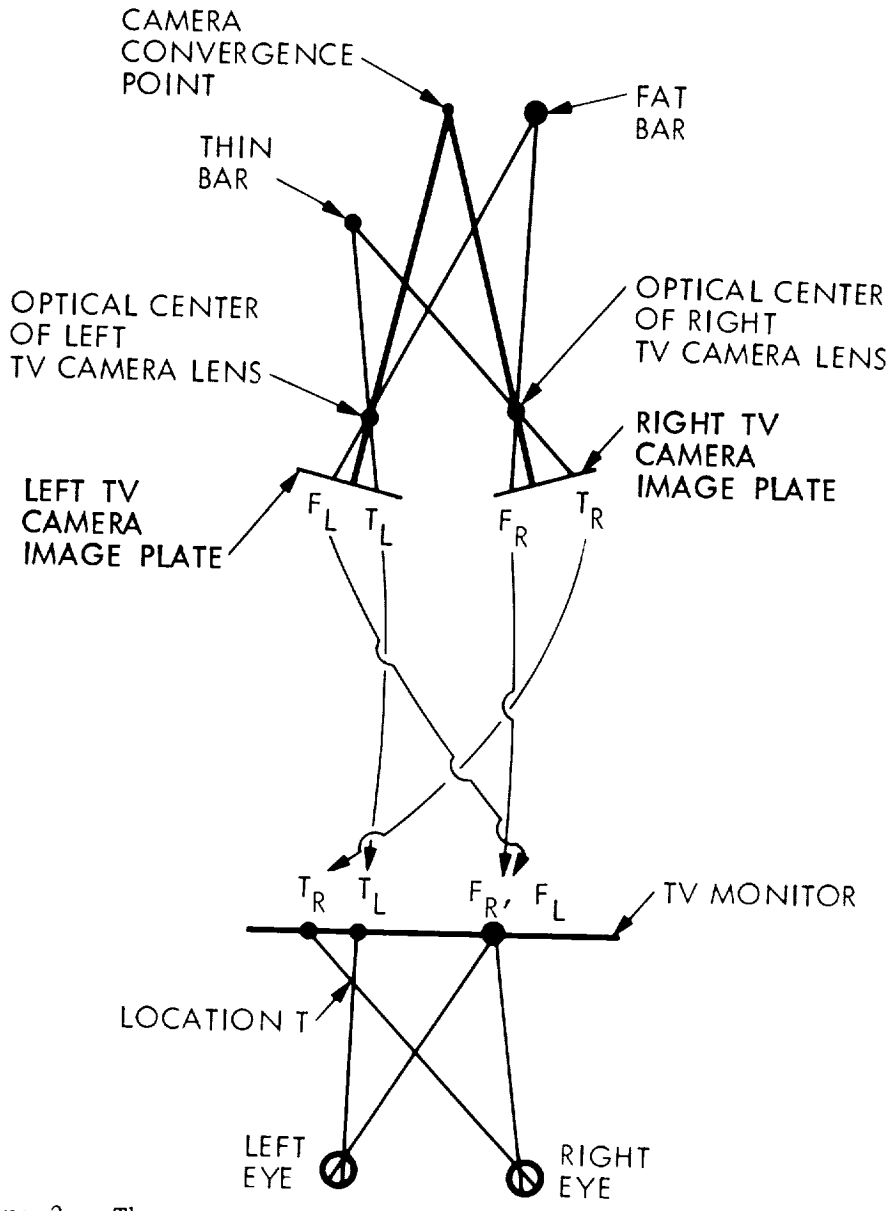

Figure 2. The stereo depth cues. The right camera records a greater # pixels between the thin and fat bars than the left camera, and displays them on the TV Monitor. The observer's right eye sees only the right camera image on the TV monitor, and the observer's left eye sees only the left camera image. Location T is the intersection of the left and right eyes' lines of sight for the thin bar. This is the only place in space that the thin bar could be, and still be seen by the two eyes on those particular lines of sight. The pixel information (# pixels difference between the two camera views as presented on the TV monitor) that determines this location includes both the true stereo depth cues and the stereo depth distortion cues. Note: we did not use bars of different thickness in our experiments.

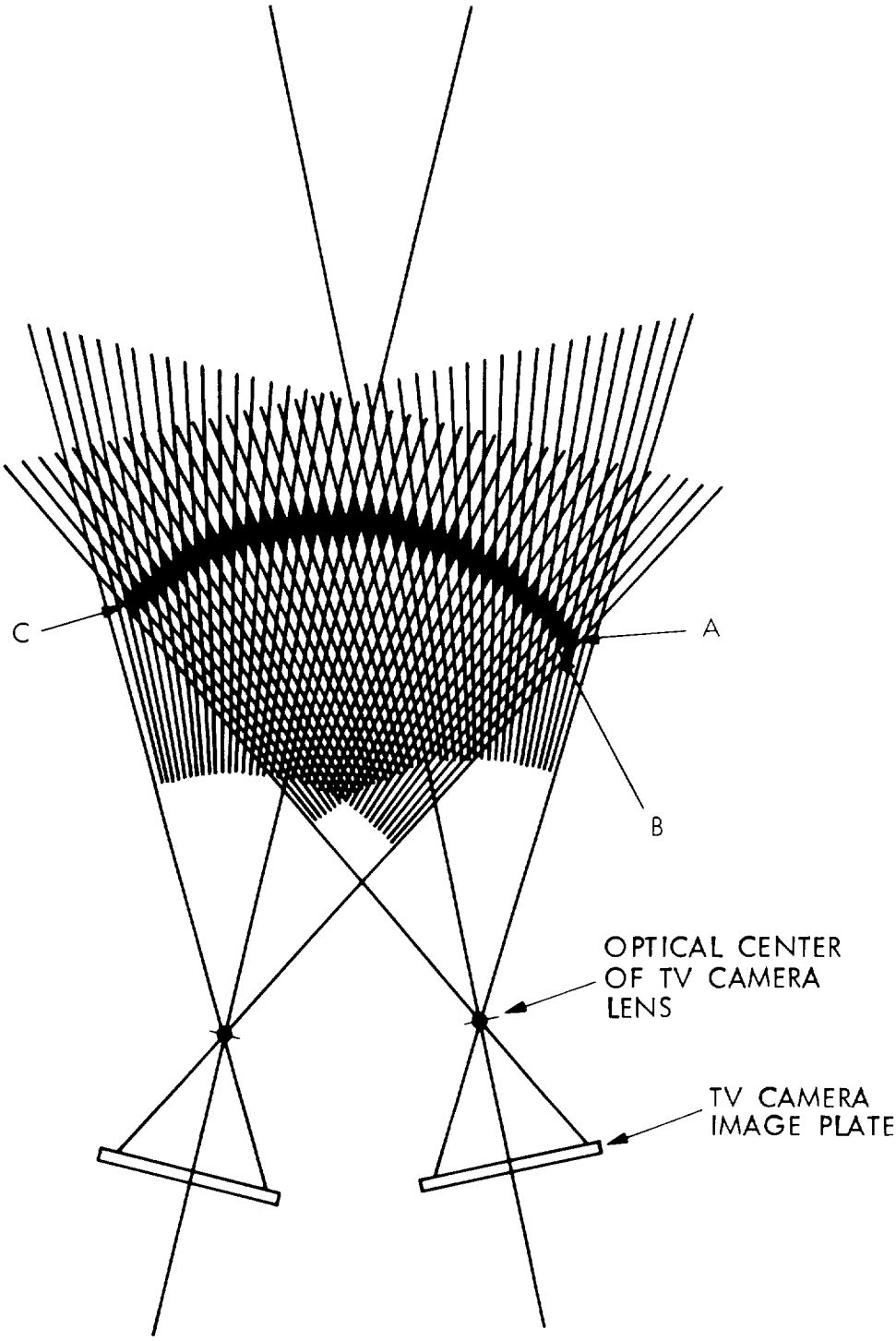

Figure 3. The geometry of the work space as viewed by converged stereo cameras. Shaded lozenges all present the same number of pixels difference to the monitor screen. Adapted from a drawing by Stephen P. Hines, HinesLab, Glendale, California.

disparity. Therefore, the depth distortion of any point on the FPP can be reduced to its spatial monitor disparity.

For two points on the FPP, one located at the convergence point, and the other a horizontal distance, k, from the convergence point, spatial monitor disparity, expressed as a number of pixels, will be:

$$\text{number of pixels} = \frac{2 * k^2 * D * f * w}{(D^4 + w^4 + 2 * D^2 * w^2 - k^2 * w^2) * (WP)} \quad (1)$$

where D = camera viewing distance (from the convergence point to the point equidistant between the first nodal points of the camera lenses)
f = focal length of the lenses (equal for both cameras)
w = ICD/2
WP = the width/pixel at CCD

For the ranges we are interested in, $k^2 * w^2$ can always be restricted to less than $D^4/1000$, and thus can be ignored.

Formula (1) can be generalized for two points located anywhere in the FPP at arbitrary distances from the camera convergence point. Consider two vertical bars held a fixed distance apart. Let us call the horizontal distance between the camera convergence point and the center point between the two bars ALIGN, and the distance between the bars the inter-target distance (ITD). The values of k in Formula (1) will then be ITD/2 + ALIGN and ITD/2 - ALIGN. The number of pixels difference we expect is the difference between these squared values which equals 2 * ITD * ALIGN. Therefore,

$$\text{number of pixels diff (2 bars)} = \frac{2 * D * f * ITD * ALIGN * ICD}{(D^2 + (ICD/2)^2)^2 * (WP)} \quad (2)$$

Here we have replaced w with ICD/2.

By moving the bars horizontally in the FPP, and measuring observers' perceptions of relative depth between the bars, the apparent shape of the FPP can be determined. For example, if an object in space is located within a lozenge with three pixels difference between camera views, the three pixel difference presented on the monitor will be the stereo depth cue the observer will see. If the object happens to be in the FPP, then the perceived depth associated with the three pixel difference will be purely distortion. In Figure 3, lozenges A and B have the same number of pixels difference. That is because lozenge A is seen by a pair of pixels which is one pixel to the left (on

each camera) of the pair of pixels which sees lozenge B. In fact, all the shaded lozenges in Figure 3 have the same number of pixels difference. Therefore objects located within these lozenges will appear in the same plane when viewed on the stereo monitor. This is because all such objects will have the same angular disparity when viewed by the human eyes, and angular disparity is the human stereo depth cue. Equal disparity leads to equal depth, which we interpret as flatness. If this curve in space appears flat, the FPP will appear convexly curved.

For the ranges we are interested in, ICD/2 never exceeds D/4 and the denominator will never be larger than $1.2 * D^4$. Thus Formula (2) can be approximated by a $1/D^3$ relation. This will lead to a camera configuration technique which significantly reduces the stereo depth distortion without reducing the stereo depth resolution, and will be discussed later.

The results of this analysis may be surprising at first. It is well known that when the two eyes converge on a point, the points in space that are at equal angles to both eyes lie on a circle. This circle passes through the convergence point and the first nodal points of the two eyes. This circle is known as the Vieth-Mueller circle. Analogously, a Vieth-Mueller circle can be defined for two converged TV cameras. The circle will pass through the convergence point and the first nodal points of the two lenses. See Figure 4. The equal angles imply that the number of pixels difference between the left and right images will be zero for all points on the camera Vieth-Mueller circle.

For a fixed viewing distance D, a smaller ICD yields a Vieth-Mueller circle with smaller radius, that is sharper curvature.

$$\text{Radius (Vieth-Mueller circle)} = \frac{D^2 + (ICD/2)^2}{2 * D} \qquad (3)$$

Thus, less spatial distortion could be expected for the larger ICD, because a bar need move less distance from the FPP to the location of 0 pixel difference. However, with the larger ICD, Formula (2) predicts a larger number of pixels difference, and thus, a larger stereo depth distortion.

The solution is as follows:

A larger ICD enhances the stereo monitor disparity, and hence the stereo percept of depth for a given physical separation of two objects in space. Thus the depth difference between the FPP and the Vieth-Mueller circle is enhanced. Calculations for two bars 15 cm apart in the FPP, aligned off-center by 5.5 cm, at a viewing distance D = 1.30 meters, and for three typical ICDs are presented in Table 1.

Table 1

Pixel characteristics of depth distortion of converged cameras at three intercamera distances

ICD	Depth (FPP to V.-M. C.†)	Depth / pixel diff	# pixels
16 cm	1.277 cm	0.515 cm	< 2.5
38 cm	1.255 cm	0.219 cm	> 5.7
60 cm	1.217 cm	0.141 cm	> 8.6

†V.-M. C. = Vieth-Mueller circle

Table 1 shows that by increasing the ICD by a factor of 3.75, (i.e., 60cm/16cm), we enhance the depth signal (number of pixels difference) by a factor of more than 3.4, (i.e., 8.6/2.5), even though the actual distance a bar would have to move from the FPP to reach a location of 0 disparity would be smaller.

The detection of a depth difference is a threshold phenomenon. The number of pixels difference must exceed the threshold, or no depth difference will be perceived. For the purposes of this discussion, let us assume a threshold of two pixels difference. Table 1 shows that for the 16 cm ICD, two pixels difference would represent 1.030 cm of depth. For the 60 cm ICD, two pixels would represent only 0.282 cm of depth.

If one bar were located in the FPP and a horizontal distance, k, from the camera convergence point, and a second bar were located at the camera convergence point, then the distance the first bar would have to be moved forward in order to lose the percept that it is behind the second bar is a measure of the depth distortion of the FPP.

For the viewing configuration described by Table 1, and the 16 cm ICD, the first bar need only be moved 0.247 cm, (i.e., 1.030 cm behind the Vieth-Mueller circle,) and the observers would not see it as behind the second bar. However, for the 60 cm ICD, the first bar would have to be moved forward 0.935 cm (i.e., 0.282 cm behind the Vieth-Mueller circle,) before the observers would no longer see it behind the second bar. Clearly, the 60 cm ICD camera configuration will suffer more distortion than the 16 cm ICD configuration.

The stereo depth resolution for the 60 cm ICD configuration will be higher than for the 16 cm ICD configuration. This is because, with the 60 cm ICD, the first bar need be moved a shorter depth distance before the number of pixels difference changes, than with the 16 cm ICD. For example, with the 60 cm ICD, the first bar would be perceived at equal depth with the second bar when it is anywhere between 0.282 cm behind and

0.282 cm in front of the Vieth-Mueller circle. With the 16 cm ICD, the first bar would be perceived at equal depth with the second bar when it is anywhere between 1.030 cm behind and 1.030 cm in front of the Vieth-Mueller circle. Thus when attempting to measure the perceived depth distortions, observers would be expected to be more certain of their perceptions of depth with the 60 cm ICD.

The conclusion here should be stressed. The larger ICDs produce higher depth resolutions, but at the expense of producing greater depth distortions. Thus with larger ICDs, we expect the operator to make larger depth errors (because of the greater distortions), and to be more certain that they are not errors (because of the higher resolution).

Dynamic Depth Distortions

In order to inspect the work space horizontally by moving the cameras, one can either translate (as shown in Figure 4) or pan (as shown in Figure 5) the cameras. Any other horizontal motion can be described as a combination of these two. Motion of either type will cause additional distortion, which we shall call dynamic depth distortion. By comparing Figure 4 with Figure 5, it can be seen that the depth difference, dL-dR, is smaller in Figure 5. This is because the rotated Vieth-Mueller circle is closer to the left bar and further from the right bar, than the translated Vieth-Mueller circle. The camera configurations are otherwise identical, and therefore the depth per pixel difference (and stereo depth enhancement) will be the same in both configurations. We therefore expect that panning the cameras will produce less depth distortion than horizontally translating the cameras.

All of the above predictions of the geometric analysis were tested with four human observers under controlled laboratory conditions.

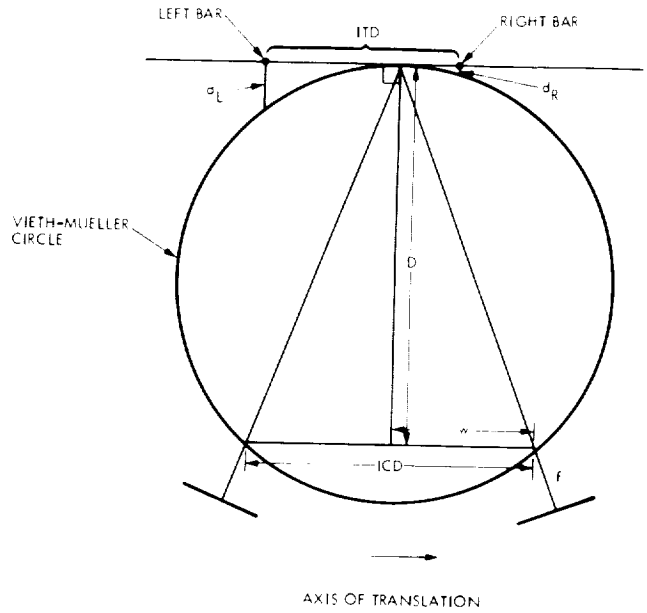

Figure 4. Depth distortion between 2 bars as stereo camera pair is translated to the right. The left bar must be moved distance dL - dR to be equidistant, behind the Vieth-Mueller circle, with the right bar. Those points on the Vieth-Mueller circle which are visible to the cameras present 0 pixels difference to the monitor screen.

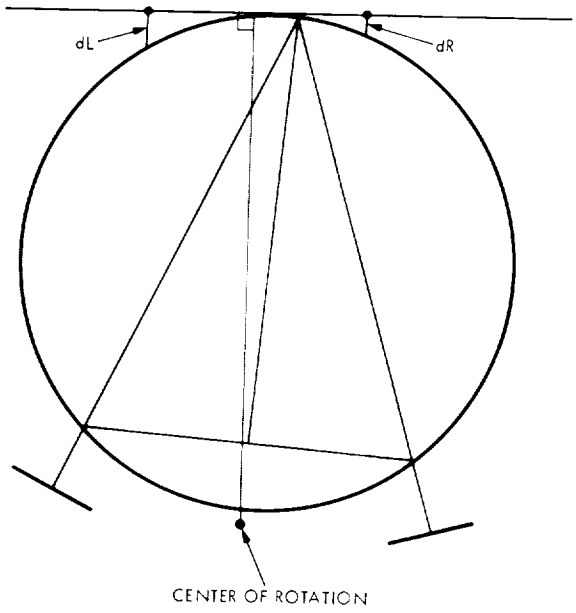

Figure 5. Depth distortion between 2 bars as stereo camera pair is panned to the right. Note: dL - dR is smaller here than in Figure 4.

3. THE EXPERIMENTS

Equipment

Two black vertical rods (0.9 cm diameter) were viewed at 1.3 meters distance by a stereo pair of RCA TC1004 videcon cameras with Vicon V17-102M auto-iris, zoom lenses. A plain white background was located about 2 meters behind these test bars. The background gave no depth cues. All non-stereo depth cues were minimized. For example, the size cue (closer bars appear larger) was minimized by adjusting the cameras and bar motions so that the tops of the bars always appeared at the same height on the monitor. The bottoms of the bars were not visible on the monitor. Thus closer bars did not appear taller. The focus cue (sharply focused bars appear closer) was minimized by limiting bar motions so that no bar ever appeared out of focus.

Stereo images were presented via a Honeywell field-sequential PLZT Stereo Viewing System, through a Dynair series 10 video switcher, to a 19-in. Toshiba 'Blackstripe' color shadow-mask monitor. The monitor has 600 horizontal pixels (triads) per line, and was the limiting factor in horizontal resolution. The shadow mask monitor breaks the screen into 600 discrete image windows. Thus, our system optically and mathematically emulates a system with CCD cameras.

The right bar was mounted on a tripod, and did not move during the experiment. The left bar was mounted on a Unimate Puma 560 robot arm. An IBM/AT Personal Computer was used to control the experiment and collect the data. Parallel ports and co-axial switches were used to enable the computer to turn on and off the information flow to the viewing monitor. When the co-axial switches were turned off, the viewing monitor appeared blank. The monitor was blanked to prevent the observers from seeing any motion of the test bars.

The two TV cameras were mounted on a precision-machined, stereo-camera mounting apparatus which could be manually adjusted to move both cameras symmetrically about the viewing axis. The stereo pair of cameras could be manually translated horizontally, precisely perpendicular to the viewing axis, or they could be panned (rotated) about a point between the cameras. See Figure 6.

A computer keyboard was masked off so that only the top row keys 1, 2, 3, 4, and 5 could be depressed. The computer read this keyboard through a serial port. This keyboard and the stereo monitor were set up in a control room where the experimental observers sat. Observers sat with their eyes about 75 cm from the stereo TV monitor. They could not see the experimental bars directly from the control room. See Figure 7.

A 20-line/inch removable transparent plastic grid was fitted to the monitor screen to aid in the precision alignment of the cameras. The grid was not present during experimentation.

ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH

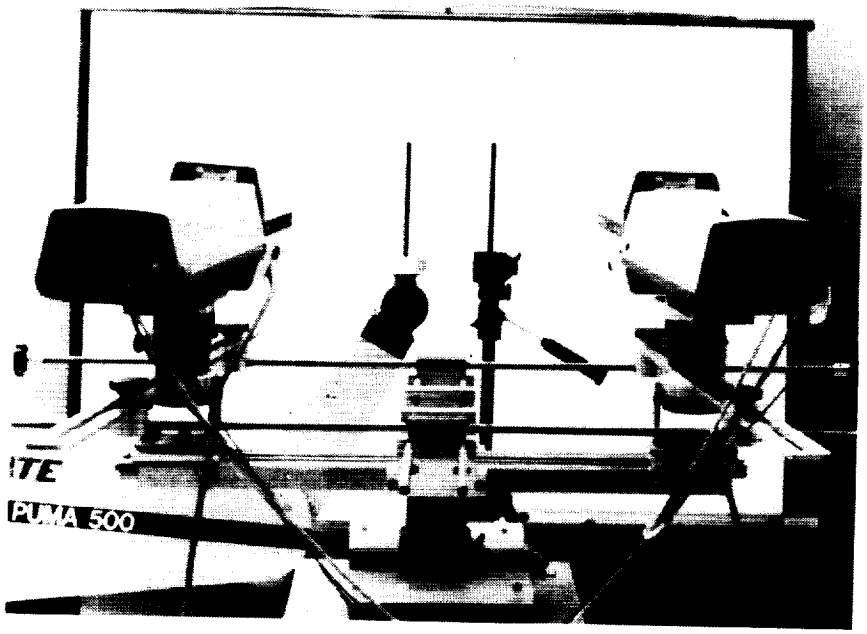

Figure 6. Experimental workspace.

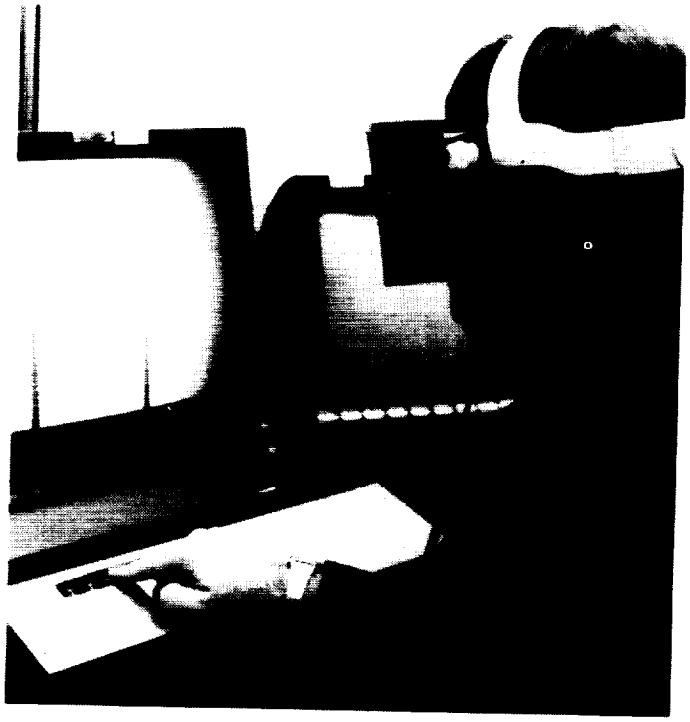

Figure 7. Experimental control room.

Experiment 1

Procedure

In experiment 1, we tested three ICDs of 16, 38 and 60 cm, and five locations of the camera convergence point in the FPP, for each ICD. The two test bars were separated horizontally by 15 cm, and presented in the FPP.

The curvature of the apparent fronto-parallel plane (AFPP) can be measured by placing the right test bar in several locations of the FPP, maintaining a fixed horizontal ITD, and determining the location of the left test bar that appears equal in depth. To do this, the left bar was moved by the robot arm to one of 19 test locations located on a line perpendicular to the plane of convergence, and parallel to the axis of symmetry between the cameras. See Figure 8. These locations were numbered 0 to 18, with location 9 in the plane of convergence. Locations 0 to 18 were -6.0, -5.0, -4.0, -3.0, -2.5, -2.0, -1.5, -1.0, -0.5, 0.0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 4.0, 5.0, and 6.0 cm from the plane of convergence, where negative values are behind the plane of convergence and positive values are in front of the plane of convergence. By "in front", we mean closer to the cameras.

The left bar was presented at each of these 19 locations five times in random order.

The experimental observers were instructed to report their perceptions of relative depth as follows:

"1" if the left bar is surely in front of the right bar
"2" if the left bar is probably in front of the right bar
"3" if the the observer is not sure which bar is closer
"4" if the left bar is probably behind the right bar
"5" if the left bar is surely behind the right bar.

In addition, if the observer perceived the bars at equal depth, he/she was instructed to report "3".

We actually moved the cameras horizontally, instead of moving the bars horizontally. These two procedures are optically and mathematically identical. The five horizontal camera alignments tested for each ICD were, in this order, 0.0, 5.5, -5.5, -3.0, and 3.0 cm. Positive numbers mean the cameras were moved to the left. Thus positive numbers mean the images were moved to the right on the monitor.

The experiment proceeded as follows.

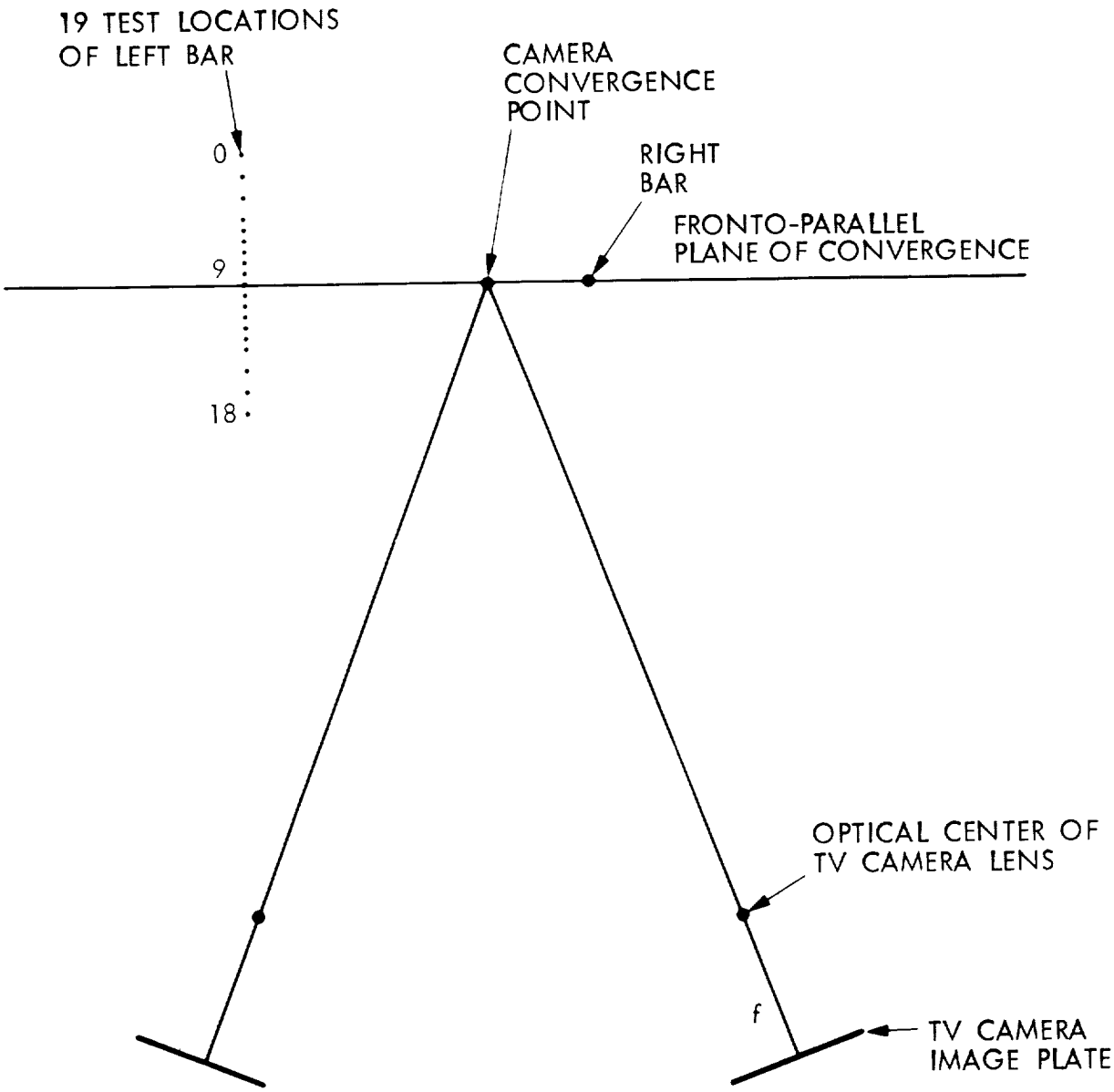

Figure 8. Experimental set-up showing the fixed right bar and the 19 possible test locations of the movable left bar.

The cameras were adjusted to the first ICD, and aligned at 0.0 cm. This proved to be a delicate task. We therefore normalized our data to control for possible adjustment inaccuracies. This is discussed below. The bars were placed in the plane of convergence (i.e., the right bar in its place, and the left bar at position 9). The alignment grid was placed on the monitor screen to measure the distance between the images of the two bars. The cameras were aligned so that each camera presented the same distance between the images of the two bars to the monitor. The adjustment grid was then removed.

The observer was seated in the control room and was asked to don the stereo visor. The experimental run then started.

The computer blanked the monitor screen. The robot moved the left bar to a randomly selected test location. After 2 seconds, the computer presented the stereo image to the monitor screen and then waited for the response from the keyboard.

The observer viewed the monitor screen until reporting a response by pressing a key ("1" to "5").

The computer recorded the response, blanked the screen, and selected the next test location. The experiment continued until all 19 locations had been presented 5 times each.

At this point, the screen was blanked for 9 seconds, the left bar was moved to position 9, the data was printed out (see Figure 9), and the experimenter was informed that the run had been completed.

The observer left the room without seeing the experimental setup. The experimenter moved the cameras horizontally to the next alignment, and the observer re-entered the control room.

After the 5 alignments had been tested, the observer rested for 15 minutes while the experimenter adjusted the cameras to the next ICD. A maximum of 10 experimental runs (2 ICDs with 5 alignments) was run each day on any one observer. Usually, only 5 experimental runs (1 ICD) were run per observer per day. The total time for 5 runs, including adjusting time, was about 25-40 minutes per observer.

As discussed later, each ICD was tested twice, in the following counterbalanced order:

16, 38, 60, 60, 38, 16 cm.

Experiment 2

In experiment 2, the stereo cameras were rotated about a point between the cameras, instead of translated, as in experiment 1. Otherwise, experiments 1 and 2 were identical.

4. DATA ANALYSIS

For each experimental run, we computed an observed depth distortion and a measure of the observer's uncertainty of that distortion. The calculation procedures are detailed in Figures 9 and 10 and Appendix 2.

Tables 2 and 3 show the computed distortions and uncertainties for experiments 1 and 2, respectively.

Next, we normalized the computed distortions and uncertainties to the 0.0 cm camera alignment value. This controlled for initial adjustment inaccuracies and enabled us to better see the effects of the camera alignments at each ICD. In other words, the data were shifted to the 0.0 cm aligned position as origin. Quite simply, for each experimental run, we subtracted the measured depth distortion of the 0.0 cm aligned position from all the measured depth distortions of that run. We adjusted the uncertainty values accordingly. These shifted data are presented in Tables 4 and 5 for experiments 1 and 2, respectively.

Our geometric analysis predicts the main independent variable to be the product of ICD and image alignment, which we shall call MTERM. In order to test if our observers' responses followed the predictions of the geometric analysis, an analysis of variance of the data in Tables 2 through 5 (both shifted and non-shifted data) was performed using the Statistical Package for the Social Sciences (SPSS) Regression program. This analysis was performed with the following 4 combinations of independent variables:

ICD and image alignment (ALIGNMENT)
ICD, ALIGNMENT and observer (OBSERVER)
MTERM, ICD, and ALIGNMENT
MTERM, ICD, ALIGNMENT and OBSERVER.

A

Experiment: 1
Condition: A
Cameras: converged
Intercamera Distance: 60 cm
Inter-Target Distance: 15 cm
Zoom: 38 mm
Depth range tested: 12 cm
Work location (from cameras): 1.30 meters

Position	Response	Position	Response	Position	Response
7	5	10	4	3	5
10	4	14	1	16	1
15	1	0	5	15	1
9	4	7	5	5	5
10	4	8	5	14	1
10	4	16	1	6	5
17	1	8	5	10	4
7	5	7	5	0	5
18	1	7	5	1	5
3	5	9	5	8	5
14	1	5	5	5	5
8	5	16	1	15	1
5	5	6	5	9	5
16	1	1	5	0	5
0	5	18	1	4	5
2	5	3	5	0	5
8	5	13	1	5	5
13	1	16	1	18	1
1	5	4	5	15	1
14	1	14	1	18	1
1	5	17	1	15	1
12	2	12	2	13	1
9	5	4	5	3	5
9	5	11	3	13	1
12	2	13	1	3	5
6	5	12	2	6	5
17	1	2	5	4	5
4	5	11	3	11	3
6	5	18	1	11	3
12	2	2	5	11	3
1	5	2	5	2	5
17	1	17	1		

The Observer's Responses

Response	Target Locations																		
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
" 1 "	0	0	0	0	0	0	0	0	0	0	0	0	0	5	5	5	5	5	5
" 2 "	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0
" 3 "	0	0	0	0	0	0	0	0	0	0	0	(5)	0	0	0	0	0	0	0
" 4 "	0	0	0	0	0	0	0	0	0	1	5	0	0	0	0	0	0	0	0
" 5 "	5	5	5	5	5	5	5	5	5	4	0	0	0	0	0	0	0	0	0

DEPTH DISTORTION ± UNCERTAINTY
1.00 ± 0.088

B

Experiment: 1
Condition: A
Camera: converged
Intercamera Distance: 16 cm
Inter-Target Distance: 15 cm
Zoom: 38 mm
Depth range tested: 12 cm
Work location (from cameras): 1.30 meters

Position	Response	Position	Response	Position	Response
6	5	13	1	10	1
12	1	11	1	2	5
7	5	10	3	17	1
17	1	18	1	9	5
16	1	1	5	16	1
16	1	18	1	15	1
14	1	13	1	4	5
18	1	8	5	10	5
14	1	3	5	9	5
6	5	6	5	16	1
14	1	8	5	17	1
16	1	14	1	5	5
9	5	5	5	2	5
6	5	1	5	10	5
17	1	3	5	5	5
15	1	18	1	14	1
15	1	18	1	15	1
12	1	7	5	7	5
6	5	3	5	13	1
11	1	4	5	15	1
5	5	4	5	13	1
11	1	0	5	12	1
10	5	3	5	12	1
7	5	12	1	13	1
4	5	0	5	5	5
11	1	17	1	3	5
2	5	0	5	8	5
0	5	7	5	11	1
1	5	1	5	8	5
8	5	2	5	9	5
4	5	2	5	0	5
9	5	1	5		

The Observer's Responses

Response	Target Locations																		
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
" 1 "	0	0	0	0	0	0	0	0	0	0	1	5	5	5	5	5	5	5	5
" 2 "	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
" 3 "	0	0	0	0	0	0	0	(2	4	5	4)	0	0	0	0	0	0	0	0
" 4 "	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
" 5 "	5	5	5	5	5	5	5	3	1	0	0	0	0	0	0	0	0	0	0

DEPTH DISTORTION ± UNCERTAINTY
-0.10 ± 0.167

Figure 9. Actual data for two experimental runs of one observer. Data for 60 cm ICD in A. Data for 16 cm ICD in B. Note wider spread of "3" responses (circled) in B, yielding greater estimate of uncertainty.

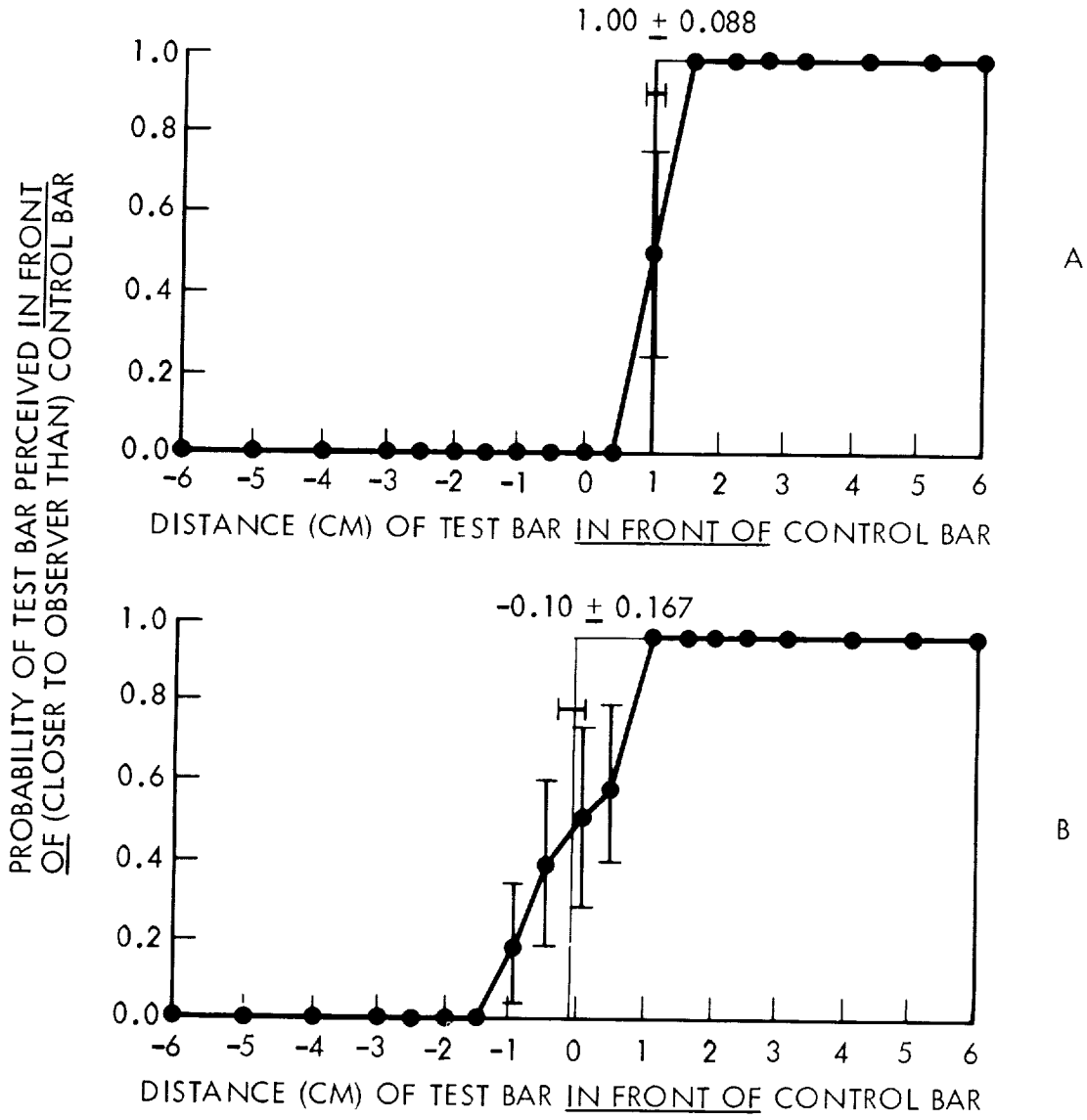

Figure 10. Probability right bar is perceived in front of left bar as a function of distance of right bar in front of left bar. Heavy line shows rectangles of equal area. Measured distortions and corresponding uncertainties were computed from the left edges of the rectangles of equal area. Data from Figure 9.

STEREO VISION IN TELEOPERATION

EXPERIMENT 1

Measured Distortion & Corresponding Uncertainty

for

Intercamera Distances of 16cm, 38cm, and 60cm. Alignments A=aligned, B=far right, C=far left, D & E= midpoints left and right resp.

OBSERVER	1		2		3		4	
16cm	RUN 1	6	1	6	1	6	1	6
C	.10 +- .322	-1.10 +- .169	-.50 +- .325	-1.05 +- .135	1.45 +- .177	.35 +- .205	1.25 +- .122	.65 +- .071
D	.05 +- .351	-1.45 +- .164	-.20 +- .336	-.95 +- .160	1.15 +- .168	.15 +- .168	.90 +- .124	.90 +- .119
A	-.38 +- .267	-1.35 +- .170	.75 +- .394	-1.35 +- .148	.20 +- .153	-.10 +- .167	.35 +- .164	-.25 +- .075
E	-.60 +- .402	-.45 +- .200	-.65 +- .233	-1.30 +- .205	.25 +- .190	-.55 +- .184	.65 +- .120	.05 +- .097
B	.38 +- .292	-.65 +- .232	1.35 +- .312	-1.85 +- .238	-.15 +- .156	-1.00 +- .174	.60 +- .081	.15 +- .122
38cm	2	5	2	5	2	5	2	5
C	.70 +- .249	.25 +- .180	.15 +- .191	.30 +- .133	2.15 +- .135	1.00 +- .133	2.30 +- .108	.85 +- .071
D	1.15 +- .198	-.60 +- .210	-.20 +- .245	-.30 +- .170	1.95 +- .187	.70 +- .155	1.80 +- .053	.75 +- .000
A	-.05 +- .266	-.35 +- .168	-1.20 +- .237	-.70 +- .119	.95 +- .139	.00 +- .088	.80 +- .053	-.05 +- .087
E	.35 +- .284	-1.10 +- .210	-1.73 +- .213	-1.45 +- .148	.50 +- .151	-.95 +- .189	.80 +- .053	-.30 +- .053
B	.15 +- .247	-.95 +- .190	-1.75 +- .254	-1.38 +- .141	.25 +- .125	-1.00 +- .088	.60 +- .113	-.55 +- .125
60cm	3	4	3	4	3	4	3	4
C	1.00 +- .234	.30 +- .217	.60 +- .323	1.70 +- .288	1.65 +- .120	1.00 +- .088	1.05 +- .087	1.25 +- .010
D	1.05 +- .218	.13 +- .210	.25 +- .232	-.63 +- .233	1.65 +- .120	.55 +- .103	1.00 +- .088	.75 +- .000
A	.85 +- .205	-.05 +- .194	-.25 +- .211	-.25 +- .164	.95 +- .103	1.00 +- .088	1.25 +- .000	.35 +- .071
E	.20 +- .188	-.70 +- .198	-.35 +- .181	-.70 +- .179	.00 +- .088	-.85 +- .120	-.05 +- .088	-.50 +- .088
B	-.05 +- .212	-.90 +- .191	-1.18 +- .258	-.93 +- .177	-.45 +- .103	-1.00 +- .088	.05 +- .100	-.35 +- .071

Table 2. Measured distortions and corresponding uncertainties of experiment 1 for four observers. Note the counterbalanced order of presentation. For example, the first two columns are data for Observer 1. Runs 1, 2, and 3 (down the first column) were followed by runs 4, 5, and 6 (up the second column).

STEREO VISION IN TELEOPERATION

EXPERIMENT 2

Measured Distortion & Corresponding Uncertainty

for

Intercamera Distances of 16cm, 38cm, and 60cm. Alignments A=aligned, B=far right, C=far left, D & E= midpoints left and right resp.

OBSERVER	1		2		3		4	
16cm	RUN 1	6	1	6	1	6	1	6
C	-1.40 +- .174	-.16 +- .166	-1.58 +- .196	-1.40 +- .151	-.25 +- .170	-.70 +- .201	.30 +- .119	.20 +- .108
D	-1.15 +- .192	-.70 +- .129	-1.65 +- .223	-1.60 +- .149	-.35 +- .197	-.15 +- .190	.55 +- .087	.25 +- .100
A	-.95 +- .144	-1.05 +- .132	-1.70 +- .108	-1.75 +- .194	-.35 +- .173	-.35 +- .184	.15 +- .112	.20 +- .053
E	.30 +- .155	-.75 +- .141	-.55 +- .200	-1.05 +- .154	-.40 +- .161	-.05 +- .211	.50 +- .088	.75 +- .075
B	.05 +- .226	.15 +- .158	-.30 +- .163	-1.22 +- .220	-.65 +- .200	-.15 +- .217	.60 +- .081	1.10 +- .081
38cm	2	5	2	5	2	5	2	5
C	.15 +- .112	-.45 +- .087	.35 +- .244	-.35 +- .135	.50 +- .145	.10 +- .113	.45 +- .087	.55 +- .087
D	-.05 +- .087	-.30 +- .119	.25 +- .125	-.95 +- .152	.25 +- .125	-.05 +- .103	.25 +- .000	.60 +- .081
A	-.60 +- .143	-.50 +- .088	.05 +- .187	-.65 +- .071	.05 +- .125	-.25 +- .125	.10 +- .091	.20 +- .053
E	.15 +- .146	.00 +- .108	-.10 +- .081	-.90 +- .129	-.05 +- .148	-.25 +- .135	.10 +- .081	.50 +- .088
B	.70 +- .139	-.55 +- .132	.25 +- .214	-1.15 +- .103	-.25 +- .125	-.50 +- .116	-.25 +- .000	.10 +- .102
60cm	3	4	3	4	3	4	3	4
C	.35 +- .112	-.10 +- .124	-.05 +- .115	.05 +- .097	.45 +- .125	-.30 +- .124	.95 +- .087	.75 +- .000
D	.15 +- .071	-.25 +- .141	-.25 +- .120	-.40 +- .088	.20 +- .124	.35 +- .112	.45 +- .087	.80 +- .088
A	.30 +- .133	-.35 +- .132	-.50 +- .108	-.40 +- .088	.05 +- .103	.05 +- .103	.20 +- .053	.20 +- .053
E	-.20 +- .083	-.25 +- .158	-.65 +- .171	-.35 +- .139	-.45 +- .125	-.20 +- .124	.45 +- .087	.35 +- .071
B	.40 +- .113	.35 +- .112	-.85 +- .075	-.70 +- .053	-.55 +- .138	-.35 +- .120	.05 +- .087	-.05 +- .087

Table 3. Measured distortions and corresponding uncertainties of experiment 2 for four observers.

STEREO VISION IN TELEOPERATION

EXPERIMENT 1

Measured Distortion & Corresponding Uncertainty

for

Intercamera Distances of 16cm, 38cm, and 60cm. Alignments A=aligned, B=far right, C=far left, D & E=midpoints left and right resp.

SHIFTED TO ALIGNED POSITION AS ORIGIN

OBSERVER	1		2		3		4	
16cm	RUN 1	6	1	6	1	6	1	6
C	.48 +- .418	.25 +- .240	-1.25 +- .503	.30 +- .200	1.25 +- .234	.45 +- .264	.90 +- .204	.90 +- .103
D	.43 +- .441	-.10 +- .236	-.95 +- .510	.40 +- .218	.95 +- .227	.25 +- .237	.55 +- .206	1.15 +- .141
A	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000
E	-.22 +- .483	.90 +- .262	-1.40 +- .449	.05 +- .253	.05 +- .244	-.45 +- .248	.30 +- .203	.30 +- .115
B	.76 +- .396	.70 +- .288	.60 +- .495	-.50 +- .280	-.35 +- .219	-.90 +- .241	.25 +- .183	.40 +- .143
38cm	2	5	2	5	2	5	2	5
C	.75 +- .364	1.10 +- .246	1.35 +- .304	1.00 +- .176	1.20 +- .194	1.00 +- .159	1.50 +- .120	.90 +- .112
D	1.20 +- .332	.25 +- .269	1.00 +- .341	.40 +- .208	1.00 +- .233	.70 +- .178	1.00 +- .075	.80 +- .087
A	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000
E	.40 +- .389	-.25 +- .269	-.53 +- .319	-.75 +- .190	-.45 +- .205	-.95 +- .208	.00 +- .075	-.25 +- .102
B	.20 +- .363	-.10 +- .254	-.55 +- .347	-.68 +- .185	-.70 +- .187	-1.00 +- .124	-.20 +- .125	-.50 +- .152
60cm	3	4	3	4	3	4	3	4
C	.15 +- .311	.35 +- .291	.85 +- .386	1.95 +- .331	.70 +- .158	.00 +- .124	-.20 +- .087	.90 +- .072
D	.20 +- .299	.18 +- .286	.50 +- .314	-.38 +- .285	.70 +- .158	-.45 +- .135	-.25 +- .088	.40 +- .071
A	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000
E	-.65 +- .273	-.65 +- .277	-.10 +- .278	-.45 +- .243	-.95 +- .135	-1.85 +- .149	-1.30 +- .088	-.85 +- .113
B	-.90 +- .295	-.85 +- .272	-.93 +- .333	-.68 +- .241	-1.40 +- .146	-2.00 +- .124	-1.20 +- .100	-.70 +- .100

Table 4. Measured distortions and corresponding uncertainties shifted to the 0.0 cm aligned position as origin. Data from experiment 1.

STEREO VISION IN TELEOPERATION

EXPERIMENT 2

Measured Distortion & Corresponding Uncertainty

for

Intercamera Distances of 16cm, 38cm, and 60cm. Alignments A=aligned, B=far right, C=far left, D & E=midpoints left and right resp.

SHIFTED TO ALIGNED POSITION AS ORIGIN

OBSERVER	1		2		3		4	
16cm	RUN 1	6	1	6	1	6	1	6
C	-.45 +- .226	.99 +- .212	.12 +- .224	.35 +- .246	.10 +- .243	-.35 +- .273	.15 +- .163	.00 +- .120
D	-.20 +- .240	.35 +- .185	.05 +- .248	.15 +- .245	.00 +- .262	.20 +- .264	.40 +- .142	.05 +- .113
A	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000
E	1.25 +- .212	.30 +- .193	1.15 +- .227	.70 +- .248	-.05 +- .236	.30 +- .280	.35 +- .142	.55 +- .092
B	1.00 +- .268	1.20 +- .206	1.40 +- .196	.53 +- .293	-.30 +- .264	.20 +- .285	.45 +- .138	.90 +- .097
38cm	2	5	2	5	2	5	2	5
C	.75 +- .182	.05 +- .124	.30 +- .307	.30 +- .153	.45 +- .191	.35 +- .169	.35 +- .119	.35 +- .102
D	.55 +- .167	.20 +- .148	.20 +- .225	-.30 +- .163	.20 +- .177	.20 +- .162	.15 +- .081	.40 +- .097
A	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000
E	.75 +- .204	.50 +- .139	-.15 +- .204	-.25 +- .147	-.10 +- .194	.00 +- .184	.00 +- .115	-.30 +- .103
B	1.30 +- .199	-.05 +- .159	.20 +- .284	-.50 +- .125	-.30 +- .177	-.25 +- .171	-.35 +- .081	-.10 +- .115
60cm	3	4	3	4	3	4	3	4
C	.05 +- .174	.25 +- .181	.45 +- .158	.45 +- .131	.40 +- .162	-.35 +- .161	.75 +- .102	.55 +- .053
D	-.15 +- .151	.10 +- .193	.25 +- .161	.00 +- .124	.15 +- .161	.30 +- .152	.25 +- .102	.60 +- .103
A	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000	.00 +- .000
E	-.50 +- .159	.10 +- .206	-.15 +- .202	.05 +- .165	-.50 +- .162	-.25 +- .161	.25 +- .102	.15 +- .089
B	.10 +- .175	.70 +- .173	-.35 +- .131	-.30 +- .103	-.60 +- .172	-.40 +- .158	-.15 +- .102	-.25 +- .102

Table 5. Measured distortions and corresponding uncertainties shifted to the 0.0 cm aligned position as origin. Data from experiment 2.

5. RESULTS

The Depth Distortions

Tables 6 and 7 show the effects of the independent variables on the observers' responses.

In experiment 1 (Table 6), for the non-shifted data, the depth distortions are significantly influenced by the ALIGNMENT, the OBSERVER, and the ICD. When we include MTERM as the first independent variable, the residual effects of ICD and OBSERVER are seen to be significant, although the residual effects of the ALIGNMENT are not. These results agree with Formula (2), which has the term ALIGN * ICD in the numerator and an ICD term in the denominator.

Shifting the data greatly reduces the significance of the effect of OBSERVER and increases the significance of the effect of the other independent variables. This suggests that much of the variability in our non-shifted data stems from inaccuracies in our initial adjustments. We repeated the initial adjustment each run so that each observer, each day, may have seen a different initial adjustment. Had the variability in our non-shifted data stemmed mostly from the effect of OBSERVER, the significance of the OBSERVER effect would not have been reduced so drastically by shifting the data. All the statements in the above paragraph about MTERM, ALIGN and ICD remain true for the shifted data.

In experiment 2 (Table 7), for the non-shifted data, the depth distortions are significantly influenced by the OBSERVER and the ICD, but not by the ALIGNMENT. When we include MTERM as the first independent variable, the residual effects of ICD, OBSERVER, and also ALIGNMENT, are seen to be significant. Note that the effect of ALIGNMENT is not seen to be significant until MTERM is introduced as the first independent variable. This occurs in both the shifted and non-shifted data, and stands in marked contrast to the results of the same test in experiment 1.

Perhaps image alignment has two cancelling effects in experiment 2. One is an MTERM effect, and one is not an MTERM effect. This makes sense logically, as image alignment here is the result of panning the cameras, thus causing both the MTERM effect of experiment 1 and the cancelling effect of rotating the fronto-parallel plane of convergence. See Figures 4 and 5.

Shifting the data in experiment 2 reduces the significance of the effect of OBSERVER and ICD and increases the significance of the effect of MTERM and ALIGNMENT. This once again suggests that much of the variability in our non-shifted data stems from inaccuracies in our initial adjustments.

Table 6

F and p values from Regression analysis.
Experiment 1, non-shifted and shifted data.

A. Non-shifted				
Independent Variables	Depth Distortions		Uncertainties	
	F	p	F	p
ICD	3.803	<0.05	7.665	<0.001
ALIGNMENT	40.042	<0.001	0.001	NS
ICD	4.716	<0.01	17.975	<0.001
ALIGNMENT	49.649	<0.001	0.002	NS
OBSERVER	29.072	<0.001	158.364	<0.001
MTERM	7.668	<0.001	0.368	NS
ICD	4.020	<0.05	7.624	<0.001
ALIGNMENT	0.077	NS	0.314	NS
MTERM	9.667	<0.001	0.866	NS
ICD	5.068	<0.01	17.954	<0.001
ALIGNMENT	0.097	NS	0.740	NS
OBSERVER	31.244	<0.001	158.181	<0.001
B. Shifted				
Independent Variables	Depth Distortions		Uncertainties	
	F	p	F	p
ICD	11.446	<0.001	3.830	<0.05
ALIGNMENT	83.666	<0.001	0.001	NS
ICD	11.350	<0.001	4.955	<0.01
ALIGNMENT	82.964	<0.001	0.001	NS
OBSERVER	0.018	NS	35.355	<0.001
MTERM	17.265	<0.001	0.116	NS
ICD	13.037	<0.001	3.802	<0.05
ALIGNMENT	0.173	NS	0.092	NS
MTERM	17.119	<0.001	0.150	NS
ICD	12.927	<0.001	4.919	<0.01
ALIGNMENT	0.171	NS	0.119	NS
OBSERVER	0.021	NS	35.096	<0.001

NOTE: p values > 0.05 are reported as NS.

Table 7

F and p values from Regression analysis.
Experiment 2, non-shifted and shifted data.

A. Non-shifted

Independent Variables	Depth Distortions		Uncertainties	
	F	p	F	p
ICD	11.261	<0.001	28.768	<0.001
ALIGNMENT	0.168	NS	0.051	NS
ICD	14.251	<0.001	35.373	<0.001
ALIGNMENT	0.212	NS	0.063	NS
OBSERVER	32.075	<0.001	27.864	<0.001
MTERM	10.994	<0.001	0.041	NS
ICD	12.223	<0.001	28.532	<0.001
ALIGNMENT	7.927	<0.001	0.007	NS
MTERM	14.288	<0.001	0.050	NS
ICD	15.884	<0.001	35.083	<0.001
ALIGNMENT	10.301	<0.001	0.009	NS
OBSERVER	35.749	<0.001	27.635	<0.001

B. Shifted

Independent Variables	Depth Distortions		Uncertainties	
	F	p	F	p
ICD	8.381	<0.001	8.308	<0.001
ALIGNMENT	0.374	NS	0.006	NS
ICD	8.477	<0.001	8.800	<0.001
ALIGNMENT	2.335	NS	0.006	NS
OBSERVER	0.379	NS	7.937	<0.001
MTERM	27.810	<0.001	0.007	NS
ICD	10.302	<0.001	8.237	<0.001
ALIGNMENT	20.051	<0.001	0.002	NS
MTERM	28.261	<0.001	0.008	NS
ICD	10.489	<0.001	8.725	<0.001
ALIGNMENT	20.376	<0.001	0.002	NS
OBSERVER	2.883	<0.05	7.869	<0.001

NOTE: p values > 0.05 are reported as NS.

The depth distortions in experiment 1 were significantly greater than the depth distortions in experiment 2. This can be shown in two ways.

The first way is to simply compare the depth distortions of experiment 1 with those of experiment 2. The SPSS analysis showed the depth distortions to be larger in experiment 1 than in experiment 2 ($p < 0.001$).

The second way to study the magnitudes of the distortions of experiments 1 and 2 is to compare the difference in observed distortions between the negative and positive 5.5 cm camera alignment test conditions. This data is presented in Table 8, and graphed in Figures 11 and 12, for experiments 1 and 2, respectively.

The SPSS analysis of variance was run on this data, and once again, ICD was found to be a significant factor ($p < 0.002$ and $p < 0.001$ for experiments 1 and 2, respectively). The values for experiment 1 were significantly greater than the values for experiment 2, ($p < 0.001$). Neither ALIGNMENT nor ALIGNMENT * ICD could be tested here as we chose the two most extreme alignments to compare, thus eliminating ALIGNMENT as a variable.

TABLE 8

Statistics of differences in perceived depth distortions of the -5.5 cm and 5.5 cm camera alignment test conditions

Experiment Number	ICD	Group Mean Distortion Difference	Standard Error of the Mean	Regression Co-efficient	F	p
1	16	0.29	0.395	0.5892	12.65	<0.002
	38	1.54	0.171			
	60	1.67	0.202			
2	16	-0.57	0.202	0.6178	14.91	<0.001
	38	0.37	0.163			
	60	0.48	0.168			

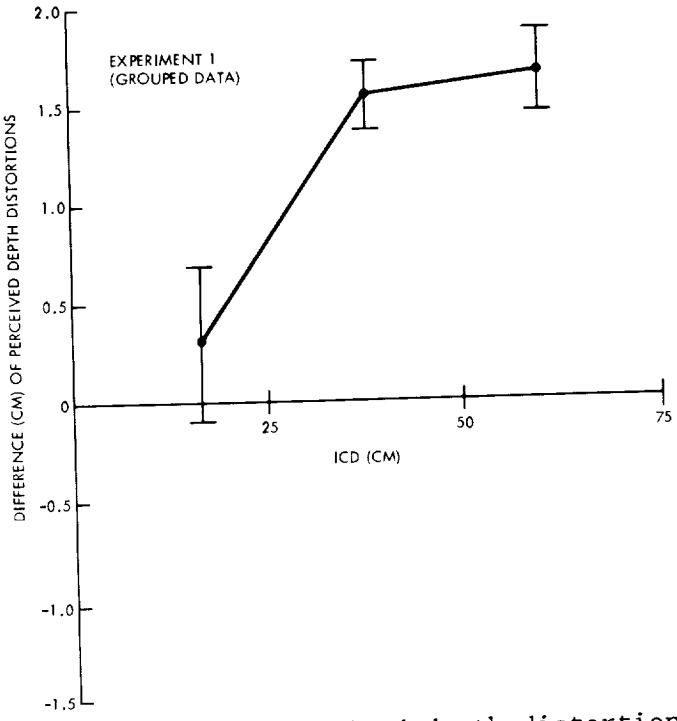

Figure 11. Difference in perceived depth distortion of the -5.5 cm and 5.5 cm camera alignment test conditions as a function of intercamera distance for experiment 1, grouped data.

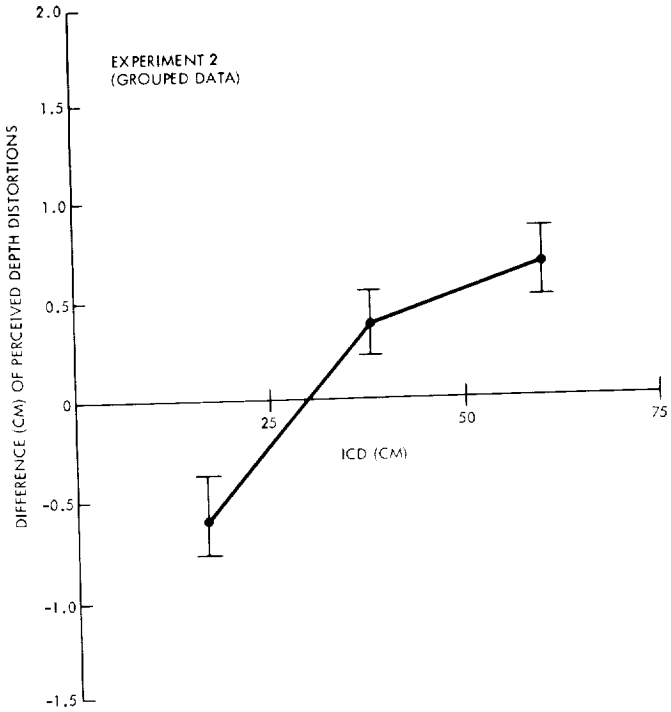

Figure 12. Difference in perceived depth distortion of the -5.5 cm and 5.5 cm camera alignment test conditions as a function of intercamera distance for experiment 2, grouped data.

The Uncertainties

The computed uncertainties in Tables 2 through 5 relate theoretically to the size of the lozenges in Figure 3, and the depth/pixel difference in Table 1. In Tables 6 and 7, in all cases, ICD and OBSERVER are the only independent variables with significant effects on the uncertainties. Specifically, uncertainty decreases with increasing ICD ($p < 0.007$ and $p < 0.0001$ for experiments 1 and 2, respectively). This agrees with expectation. However, the effect is much smaller than expected.

Table 1 predicts that the measured uncertainty of the 60 cm ICD would be less than 30% of the measured uncertainty of the 16 cm ICD. However, we found the 60 cm ICD uncertainty to be about 70% of the 16 cm ICD uncertainty. This could be due to the double meaning of the response "3", which always contributes to the calculation of the uncertainty, although it is only an uncertain answer some of the time. Specifically, when the bars are truly at the same depth, and the observer so perceives them with absolute certainty, he/she responds "3"; but, our uncertainty statistic computes this as an uncertain response. This artificially increases all the estimates of uncertainty, thus adding a roughly constant amount to all conditions. This may well explain the difference between the expected 30% and the observed 70%.

This problem arose during the actual data collection. The observers asked what response to give when they were sure the bars were at equal depth. We decided they should respond "3" as that would yield an accurate value for the perceived depth distortion. The proper reaction should have been to redesign the response keyboard to allow a separate response button to be pressed. Then both our perceived depth distortions and our uncertainty measures would have been accurate. This shall be done in all future work. Nevertheless, despite this bias against us in our measurement, we have successfully measured a significant drop in uncertainty with increasing ICD.

Time-order effects, including practice, must be considered in experiments of this type. We were able to tease out the time-order effects from the effects of the ICD by counterbalancing the presentation of the ICD tests, (16, 38, 60, 60, 38, 16 cm).

We have plotted the uncertainty values in Figures 13 and 14 for experiments 1 and 2, respectively. An SPSS linear regression analysis was run with time as the only independent variable, and then with ICD as the only independent variable. In experiment 1, time was a factor ($p < 0.0007$) and ICD was a factor ($p < 0.007$). In experiment 2, time was not a factor ($p > 0.40$) but ICD was a factor ($p < 0.0001$). We therefore estimate that the time-order effects, including practice, were completed during experiment 1.

This was not expected, as we allowed our observers to practice for about one hour per day, five days a week, for one month, prior to the start of experiment 1.

In summary, one result of this work is that the criterion of certainty varies between our observers, although the actual depth distortions they perceive do not.

The main result of this work is that the observers' responses follow the geometric predictions of the stereo information (number of pixels difference) on the TV monitor. Thus, the observers' internal corrections and/or distortions do not invalidate the usefulness of our geometric analysis to predict optimal camera configurations.

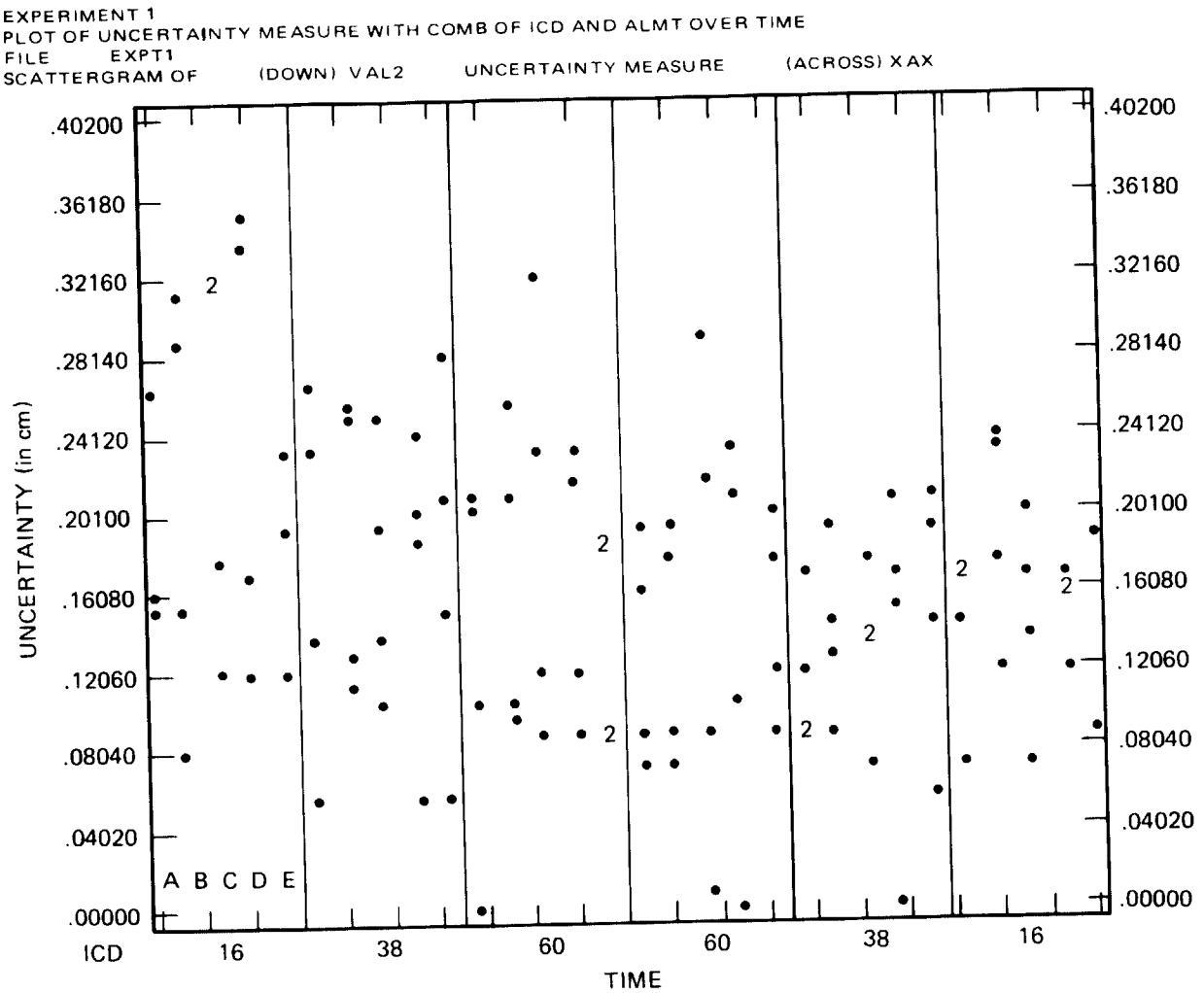

Figure 13. Uncertainty measurements as a function of order of presentation of ICD for experiment 1. Each of the 6 main columns includes 5 minor columns of camera alignments (A = 0.0, B = 5.5, C = -5.5, D = -3.0, and E = 3.0 cm.) Each camera alignment column contains four points, one for each subject's uncertainty measurement. A "2" represents 2 data points at that location.

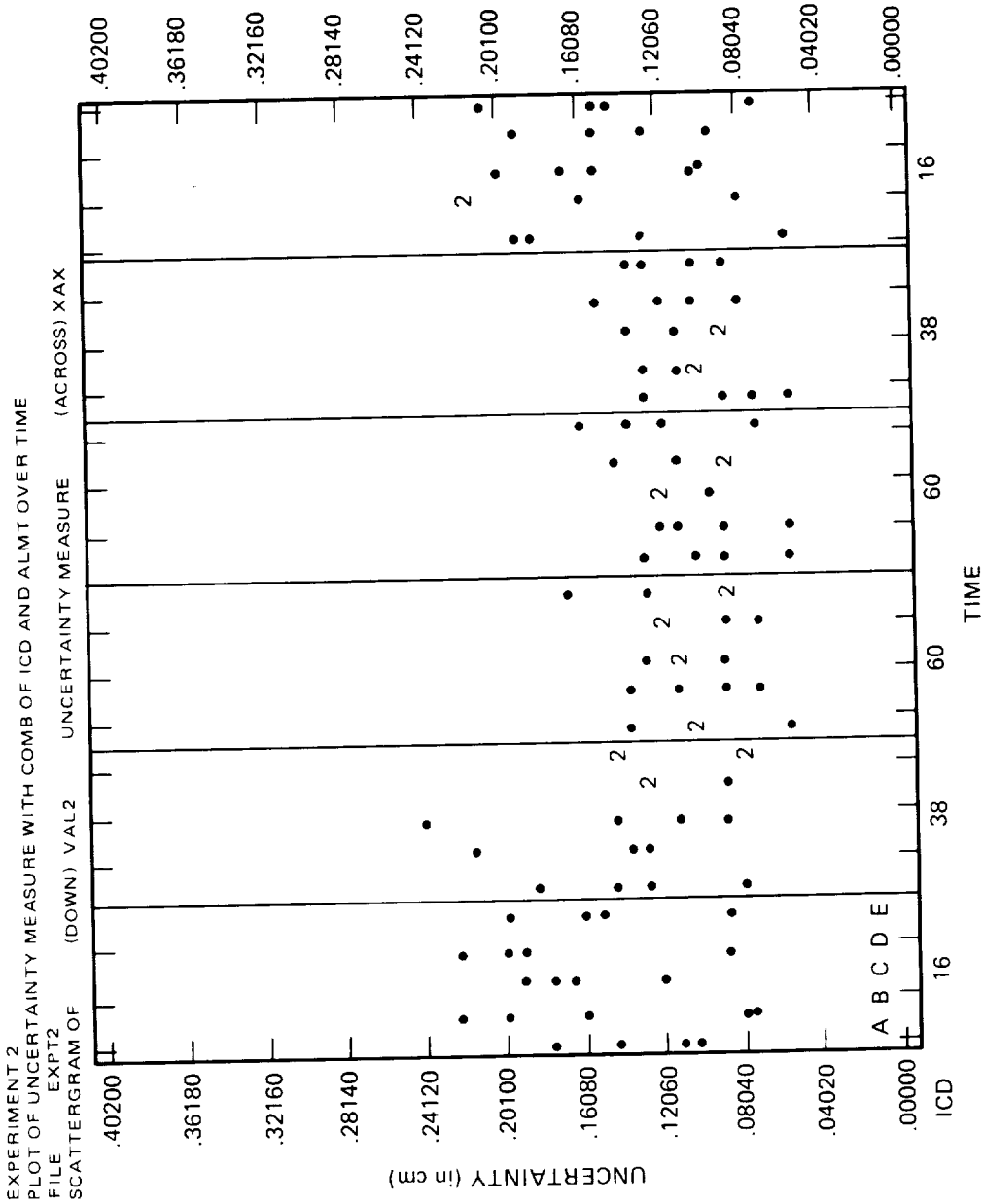

Figure 14. Uncertainty measurements as a function of order of presentation of ICD for experiment 2.

6. DISCUSSION

The stereo depth distortion can be analyzed by breaking it into static and dynamic components. By static, we mean the distortion that is present when we do not move the cameras. It comes from the camera alignment geometry. By dynamic, we mean the change in the static distortion as the stereo camera system scans the work space.

Figure 3 shows the nature of the static stereo depth distortions. Figure 3 represents two CCD cameras converged and viewing a work space. Each lozenge represents the region in space that is seen by a pair of pixels, one on each camera.

In Figure 3, all the shaded lozenges have the same number of pixels difference. Lozenges with equal number of pixels difference will present equal depth cues to the human observer.

The centers of the lozenges with 0 pixels difference lie on a circle. This circle goes through the convergence point and the first nodal points of the lenses of the cameras. We shall refer to it as the Vieth-Mueller circle of the cameras.

Consider now the lozenges with a fixed, non-zero, number of pixels difference (for example, 3). The centers of these lozenges lie on a curve. This curve also goes through the first nodal points of the lenses of the cameras. However, this curve and all other curves with a non-zero number of pixels difference are not circles.

Minimization of the Static Depth Distortion

Consider now the $1/D^3$ relation which resulted from Formulas (1) and (2). This shall lead us to a way to greatly minimize static depth distortions without loss of stereo depth resolution.

Let us look at Formula (2). Suppose we viewed one bar at the convergence point, and a second bar at k = ITD. In this case, Formula (2) = Formula (1), (with the exception of the $k^2 * w^2$ term, which we can ignore) because ALIGN = ITD/2. Now let us ask what would happen if we double the viewing distance D, and double the ICD (which of course doubles w), and also double the focal length. In this case, our cameras would now view the work space from the same angle as before the doubling. We leave k unchanged (which of course leaves ITD unchanged), and we converge on the same convergence point (which leaves ALIGN unchanged). What happens to the depth signal at the monitor? In other words, what is the effect on the number of pixels difference?

Formulas (1) and (2) predict the number of pixels difference would be halved. That is, the distortion would be halved.

Consider now Figure 15. Here we have the two camera configurations in question. We have labelled the cameras Rn, Rf, Ln, and Lf for Right camera in Near configuration, Right camera in Far configuration, etc. We have also drawn two lines parallel to the camera CCD chips which we shall call the lines of equidistant projection. On these lines, every pixel sees a unit length segment of (L/f) * (width/pixel at CCD), where

$$L^2 = D^2 + w^2.$$

Because we doubled D, w and f, for cameras Rf and Lf, every pixel on each of the 4 cameras sees the same size unit length segment for the line of equidistant projection parallel to its CCD chip.

We have labelled the projection points on the corresponding lines of equidistant projection as Rf', Rn', Lf', and Ln'.

Consider first the near cameras. Clearly, the length Ln' to C is larger than Rn' to C. The number of pixels difference will be strictly proportional to (Ln' - Rn').

Consider next the far cameras. Clearly the length Lf' to C will be less than Ln' to C. Also, the length Rf' to C will be greater than Rn' to C. Thus, the number of pixels difference, which will be proportional to (Lf' - Rf'), is less than (Ln' - Rn').

We have qualitatively shown that the number of pixels difference for the far cameras will be less than for the near cameras. The quantitative demonstration of this is exactly Formulas (1) and (2).

The importance of this point must not be overlooked. By increasing the camera-to-object viewing distance, the ICD, and the focal lengths of the camera lenses, we can maintain image field size and stereo depth resolution, while significantly decreasing the static stereo depth distortion!

Minimization of Dynamic Depth Distortion

We have shown that panning about point A in Figure 16 produces less distortion than translating horizontally. However, it is easy to see theoretically that panning about point B in Figure 17 (the center of the V.-M. circle) should produce hardly any distortion at all. If the curves of equal number of pixels difference were circles with center B, no dynamic distortion at all would be so produced. As is, the only dynamic distortion produced would be the difference between circles with center at B and the actual curves. The center of the Vieth-Mueller circle is less than half the distance between the cameras and the convergence point. For close teleoperation, it would be easy to compute this point and devise a method to pan about it.

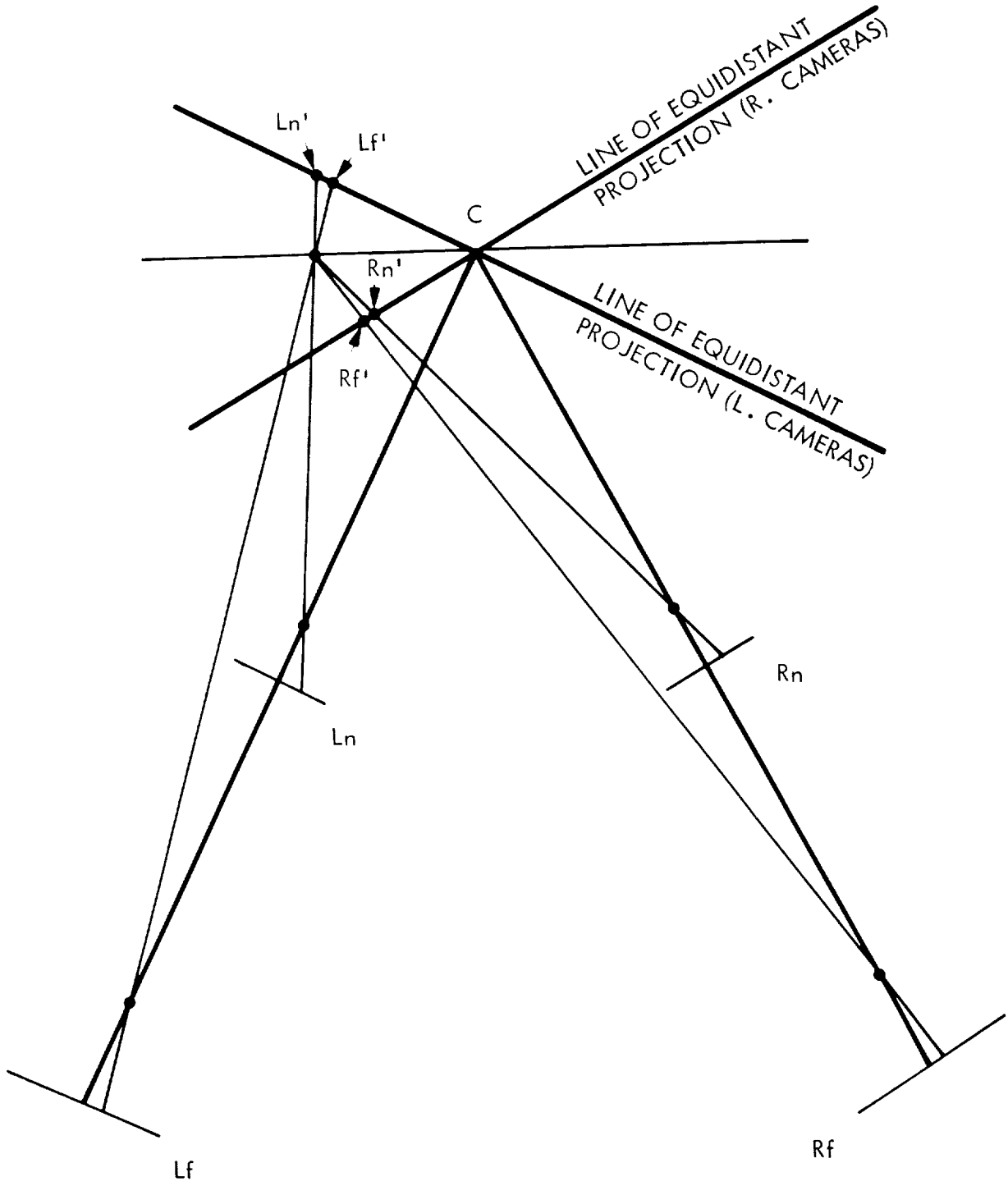

Figure 15. Minimization of static depth distortion. By doubling the camera-to-object viewing distance, the intercamera distance and the focal length of the camera lenses, one can maintain image field size and stereo depth resolution, while cutting the static stereo depth distortion in half.

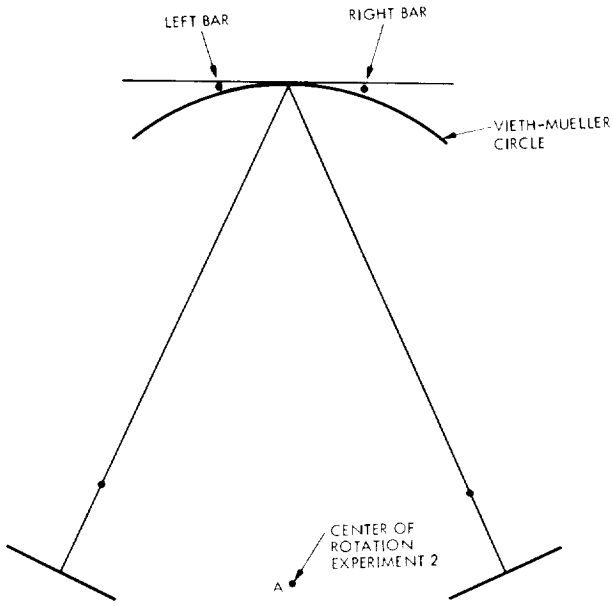

Figure 16. Partial minimization of dynamic depth distortion. Point A represents the center of rotation for experiment 2.

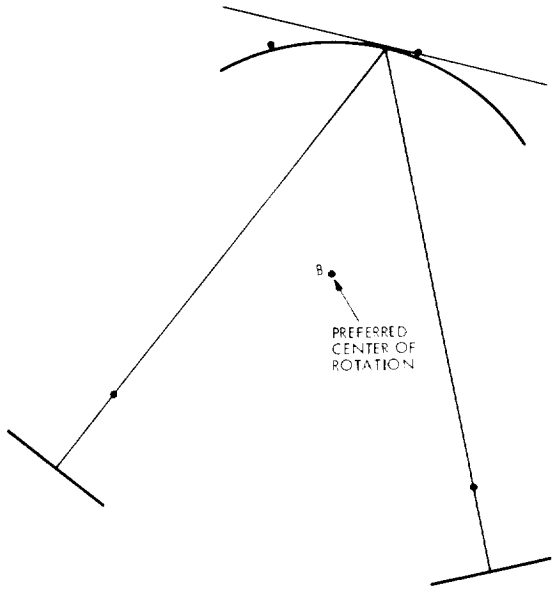

Figure 17. Minimization of dynamic depth distortion. Point B represents the preferred center of rotation, i.e., the center of the Vieth-Mueller circle. Panning the camera pair about this point will minimize dynamic distortion.

7. CONCLUSION

A geometric analysis without small angle approximations has been shown to predict distortions of the FPP which otherwise might not be adequately predicted. These distortions have been demonstrated to be perceived by four human observers.

Our human observers' responses follow the stereo information on the TV monitor. Internal perceptual corrections and/or distortions do not invalidate the usefulness of our geometric analysis to predict optimal camera configurations.

Our analysis predicts that static stereo depth distortion may be greatly decreased, without decreasing the stereo depth resolution, by increasing the camera-to-object viewing distance, the intercamera distance, and the focal length of the TV camera lenses.

Our analysis further predicts that dynamic stereo depth distortion may be greatly reduced by rotating about the center of the Vieth-Mueller circle when panning the stereo cameras.

8. POTENTIAL APPLICATIONS

In the final approach and close-up work of free-flying or stationary teleoperation, stereo TV vision systems may be used to provide necessary depth information.

In order to eliminate the stereo depth distortion errors from teleoperation task performance, a supervised automated system can be built which will adjust the stereo camera configuration on line as the end effector moves through the work space. Different tasks and different people may require different depth resolutions and may tolerate different depth distortions. This may well entail on-line adjustments of the intercamera distance. As the intercamera distance between converged stereo cameras is changed, different distortions of the three spatial axes may be produced. The system should provide the optimal trade-off between stereo depth resolution and stereo depth distortions for a specific task and operator and should automatically adjust the translational axes gains of the hand controller to counteract any remaining visual distortions. For example, if an operator were viewing a meter stick stereoscopically, and the meter stick appeared to be curved convexly away from the operator, the operator need move the hand controller along an identical convex curve, and the end effector would move along the surface of the truly uncurved meter stick.

This translational axes gain adjustment technique has been employed in stereo microscopes with joystick-driven microsurgery tools, and has demonstrated remarkable improvement in the performance of trained personnel. (D. H. Fender, personal communication.)

Other adjustment or compensation procedures are also possible.

Such an automated system should be designed to allow the operator to function with a distorted percept of space as if it were not distorted at all.

9. ACKNOWLEDGEMENTS

We should like to thank several people who contributed to this project including: Antal K. Bejczy, Roy Chafin, Kevin Corker, Michel Delpech, Derek H. Fender, Stephen P. Hines, Michael Hyson, Ed Kan, Dan Kerrisk, Gerhardt Knieper, Cosette Lare, Camille Marder, Doug McAffee, William Rosar, and Marcos Salganicoff.

10. REFERENCES

1. Shields, N.L., Kirkpatrick, M., Malone, T.B., and Huggins, C.T. Design parameters for a stereoptic television system based on direct depth perception cues. Washington, D.C.: Proceedings of the Human Factors Society 19th Annual Meeting, 1975, pp. 423-427.

2. Grant, C., Meirick, R., Polhemus, C., Spencer, R. Swain, D., and Tewell, R. Conceptual design study for a teleoperator visual system report. Denver, CO: Martin Marietta Corporation Report NASA CR-124273, April 1973.

3. Upton, H.W. and Strother, D.D. Design and flight evaluation of a helmet-mounted display and control system. In Birt, J.A. & Task, H.L. (Eds.) A symposium on visually coupled systems: Development and application (AMD-TR-73-1). Brooks Air Force Base, TX, September, 1973.

4. Zamarian, D.M. Use of stereopsis in electronic displays: Part II - Stereoscopic threshold performance as a function of system characteristics. Douglas Aircraft Company Report MDC J7410, December, 1976.

5. Pepper, R.L., Cole, R.E., and Spain, E.H. The influence of camera separation and head movement on perceptual performance under direct and TV-displayed conditions. Proceedings of the Society for Information Display, 1983, 24, pp. 73-80.

6. Spain, E.H. A psychophysical investigation of the perception of depth with stereoscopic television displays. University of Hawaii doctoral dissertation, May 1984.

7. Bejczy, A.K., Performance evaluation of computer-aided manipulator control. IEEE International Conference on Systems, Man and Cybernetics, 1976.

8. Fender, D.H., Professor, Biological Information Systems California Institute of Technology, Pasadena, Calif. 91125. Personal communication.

APPENDIX 1

In Figure 18, the lines of equidistant projection are drawn for both cameras. For a point on the fronto-parallel plane of convergence located a distance k horizontally to the left of the camera convergence point, its projection on the left camera line of equidistant projection will be Pl' from the camera convergence point, where

$$Pl' = \tan(\text{alpha}) * L$$

$$= \tan [\ \arctan(w/D) - \arctan(\ (w-k)/D\ )\ ] * L$$

$$= \tan\left[\arctan\left(\frac{w/D - (w-k)/D}{1 + w/D * (w-k)/D}\right)\right] * L$$

$$= \frac{[w/D + (k-w)/D] * L}{1 - w * (k-w)/D^2}$$

$$= \frac{k * L * D}{D^2 + w^2 - k * w}$$

Similarly, Pr', the projection on the right camera line of equidistant projection, will be:

$$Pr' = \frac{k * L * D}{D^2 + w^2 + k * w}$$

The difference between the two projections will be:

$$Pl' - Pr' = \frac{2 * k^2 * D * L * w}{(D^2 + w^2 - k * w) * (D^2 + w^2 + k * w)}$$

The number of pixels difference will be:

$$\text{number of pixels diff} = \frac{(Pl' - Pr') * f}{L * (\text{width per pixel at camera plate})}$$

$$= \frac{2 * k^2 * D * f * w}{(D^4 + w^4 + 2 * D^2 * w^2 - k^2 * w^2) * \left(\begin{matrix}\text{width per pixel} \\ \text{at camera plate}\end{matrix}\right)}$$

NOTE: Small angle approximations (x = tan x), would yield

$$Pl' = Pr' = \frac{k * L}{D}$$

or, equivalently, Pl' - Pr' = 0. This is how the small angle approximations can obscure the nature of the stereo depth distortion of the fronto-parallel plane of convergence.

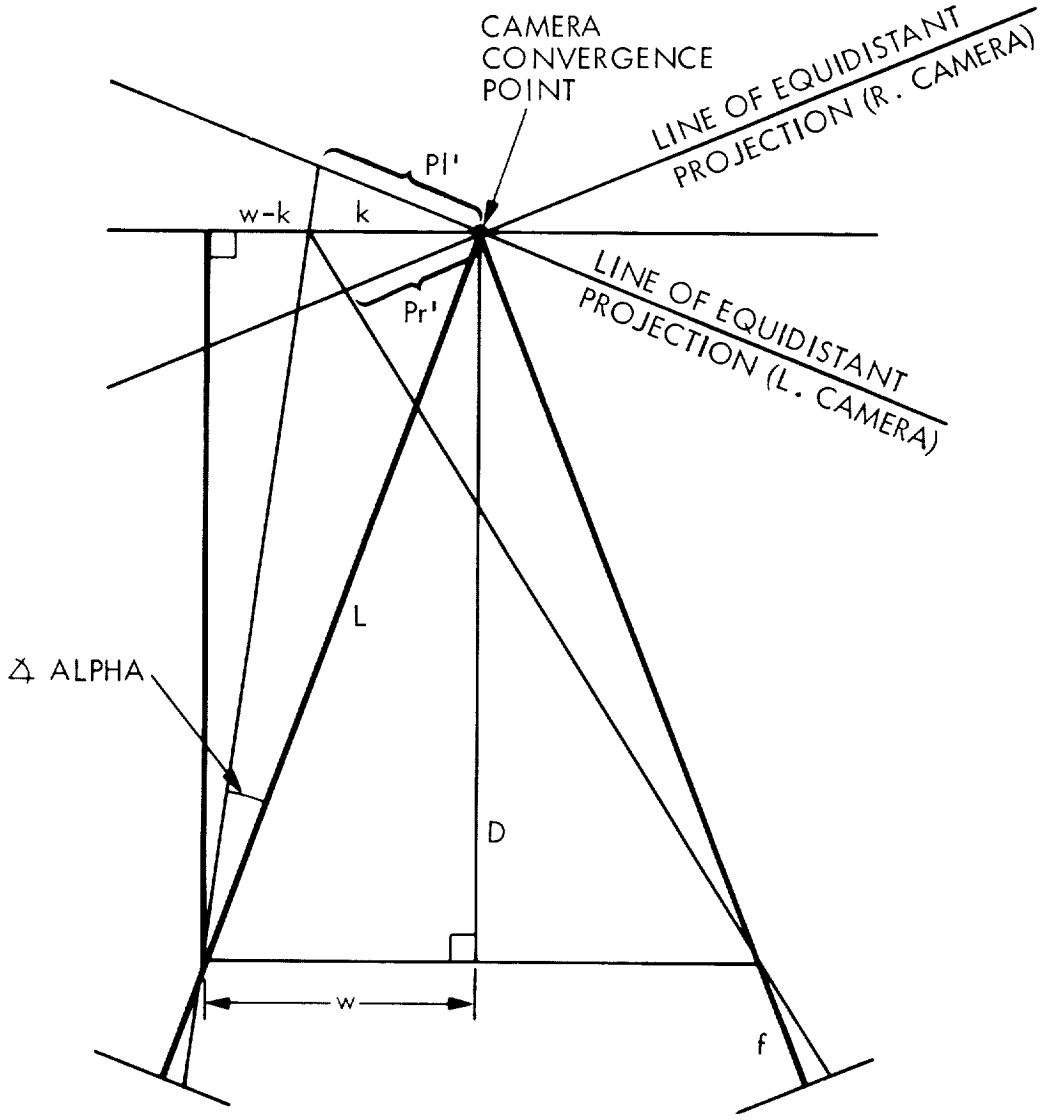

Figure 18. The geometry of converged stereo cameras. On the lines of equidistant projection, every pixel sees a unit length segment of (L/f) * (width/pixel at CCD). The # pixels difference presented to the monitor by the two cameras will be proportional to (Pl' - Pr').

APPENDIX 2

In each experimental run (one ICD and one alignment), 19 test locations were judged 5 times to be in front of, behind, or equal to a fixed location. This gave us a measurement of the probability that each position would be perceived in front of the fixed location. We computed that probability as follows:

$$P(\text{front}) = \frac{N("1") + N("2") + N("3")/2}{N("1") + N("2") + N("3") + N("4") + N("5")}$$

where N("I") is the number of responses of "I" for I = 1 to 5.

Thus, if an observer answered all "1" and "2" for location 18, we would compute P(front) for location 18 to be 1.0. If an observer answered "3" twice, and "5" three times, for location 7, we would compute P(front) for location 7 to be 0.2. (See Figs. 9B and 10B, where location 7 in Figure 9B corresponds to -1 cm on Figure 10B.) NOTE: we count each "3" response as 1/2 in front and 1/2 behind. We count "4" and "5" responses as behind, and therefore they do not show up in the numerator.

By breaking our responses into two categories, we had a binomial distribution of P(front) about each location. We estimated the uncertainty about this point by $(P * (1 - P))/N$, where N is the number of responses at that location, (in this case 5). The only time an uncertainty could be non-zero is when P is not equal to 0 or 1. This can only occur when a particular location was either reported as "3" (equal depth or the observer is uncertain) or when that location was reported as sometimes in front and sometimes behind. (NOTE: we did not count reports of "probably" as adding to the uncertainty).

We next graphed the P(front) as a function of the distance between the test location and the right (fixed) bar location, and computed the area under the curve. We computed a rectangle of equal area and probability 1.0, which gave an estimate of the depth distortion between the two bars for that ICD and at that particular alignment. See Figure 10.

Using the uncertainties of each of the 19 P(front) measurements, we approximated the uncertainty of the width of a rectangle of equal area. We first found the area and standard deviation of each trapezoid under the curve. We summed the areas, and used the sums-of-squares rule to combine the standard deviations. The uncertainty bars on the rectangles of equal area may, at first glance, appear too small. They are not. To see this, one must realize that the Y axis is probability, with a maximum value of 1.0. Thus an error bar of $\pm$ 1.0 (twice the height of the Y axis) would contribute between 1/2 cm and 1 cm (depending on the test location) to the standard deviation of a rectangle of equal area.