MASSACHUSETTS INSTITUTE OF TECHNOLOGY
A. I. LABORATORY

Artificial Intelligence
Memo No. 289 March 1973

VISUAL POSITION EXTRACTION USING STEREO EYE SYSTEMS WITH A RELATIVE ROTATIONAL MOTION CAPABILITY*

Daniel W. Corwin

ABSTRACT

This paper discusses the problem of context-free position estimation using a stereo vision system with moveable eyes. Exact and approximate equations are developed linking position to measureable quantities of the image-space, and an algorithm for finding these quantities is suggested in rough form. An estimate of errors and resolution limits is provided.

Work reported herein was conducted at the Artificial Intelligence Laboratory, a Massachusetts Institute of Technology research program supported in part by the Advanced Research Projects Agency of the Department of Defense and monitored by the Office of Naval Research under Contract Number N00014-70-A-0362-0005.

---

*This paper was originally issued as A.I. Vision Flash 22, January 1972.

I. Introduction

The intent of this study is two-fold:

i. to determine quantitatively the nature and amount of additional information presented by a stereo (as opposed to monoscopic) visual apparatus;

ii. to investigate qualitatively some useful ways of incorporating this additional information in an artificial visual scene analyser.

It may be noted at once that the only real distinction between stereoscopic and monocular vision is that the latter presents a single visual image of a scene, while the former provides us with two images. This distinction becomes quickly meaningless however, unless a practical method exists of comparing the two images, and determining the differences between them. For this reason, I am forced right at the beginning to address the question of "difference-measuring" between visual images, and to state explicitly the assumptions I have made concerning it.

My first assumption is that at some level of even current vision programs, the image seen by a single eye is represented as a 2-D matrix of measured light intensity values, or could be so represented without much difficulty.

My second assumption is that if a stereo eye system were to be used, it would be mechanically* constrained so that the "center points" of the 2-D image matrices were never

---

* a variety of feedback control systems, or even digital control systems can be imagined which might do this, and yet allow the constraint to be removed if desired.

representative of different points in 3-space; i.e., that a "point-of-trigonometric-focus" existed, towards which both eyes always "pointed". This focal point could freely shift in distance away from the eyes, or closer, but the forward axes of the eyes could not become significantly skew relative to the limits of angular resolution. The eyes would be capable of only single-degree-of-freedom motion with respect to each other, about their vertical axes.

My third assumption, which is difficult to justify just yet, is that if an element in one eye's image matrix were selected, its counterpart in the other image could be found from local evidence such that both represented the same point in 3-space. This is rather a difficult exercise in pattern matching in the general case, particularly since I understand that high noise levels are present in the visual images, but I will offer some results in Part III that can help quite a bit in limiting the search. I'll come back to this problem later; for now I'll just assume it is solveable.*

With these assumptions, we proceed to some mathematics relevant to the (continuous) real-world situation.

---

* Lerman (1) has in fact presented results which demonstrate that this type of pattern matching can be accomplished when applied to images generated by eyes focused on infinity, the only case he considered. See Part III.

II. Definitions, Co-ordinate Systems, and Consequences

Let us consider a fixed orthogonal reference co-ordinate system S (the 'table' system), defined so that $\bar{I}$ is generally 'up', $\bar{J}$ is generally 'right', and $\bar{K}$ is generally 'away'. Presume that in this system, the point midway between the eyes is located out in the general $-\bar{K}$ direction at $\bar{P}_s$, and that the eyes are focused (in the trigonometric sense) on $\bar{F}_s$. A group of objects to be viewed lies near the origin, and both $\bar{P}_s$ and $\bar{F}_s$ are known.

The S system all by itself is adequate for representing the location of points in space, but it will be useful here to define a few more for clarity. One alternative is the system J_o (see Fig. I), whose origin lies at $\bar{P}_s$, and whose orientation is such that

$\bar{I}_o$ lies along $\bar{J}_o * \bar{K}_o$ (almost 'up');
$\bar{J}_o$ lies along $\bar{K}_o * \bar{I}_s$ (horizontal; almost 'right');
$\bar{K}_o$ lies along $\bar{F}_s - \bar{P}_s$ (towards $\bar{F}_s$).

Transformation between these systems is easily made through the relation

$$\bar{X}_s = \overset{*}{T}_{so}\,\bar{X}_o + \bar{P}_s \qquad (2-1)$$

where

$$\overset{*}{T}_{so} \equiv \begin{bmatrix} \rightarrow \text{UNIT}\left(\left((\bar{F}_s - \bar{P}_s) * (1,0,0)\right) * (\bar{F}_s - \bar{P}_s)\right) \rightarrow \\ \rightarrow \text{UNIT}\left((\bar{F}_s - \bar{P}_s) * (1,0,0)\right) \longrightarrow \\ \rightarrow \text{UNIT}\left(\bar{F}_s - \bar{P}_s\right) \longrightarrow \end{bmatrix} \qquad (2-2)$$

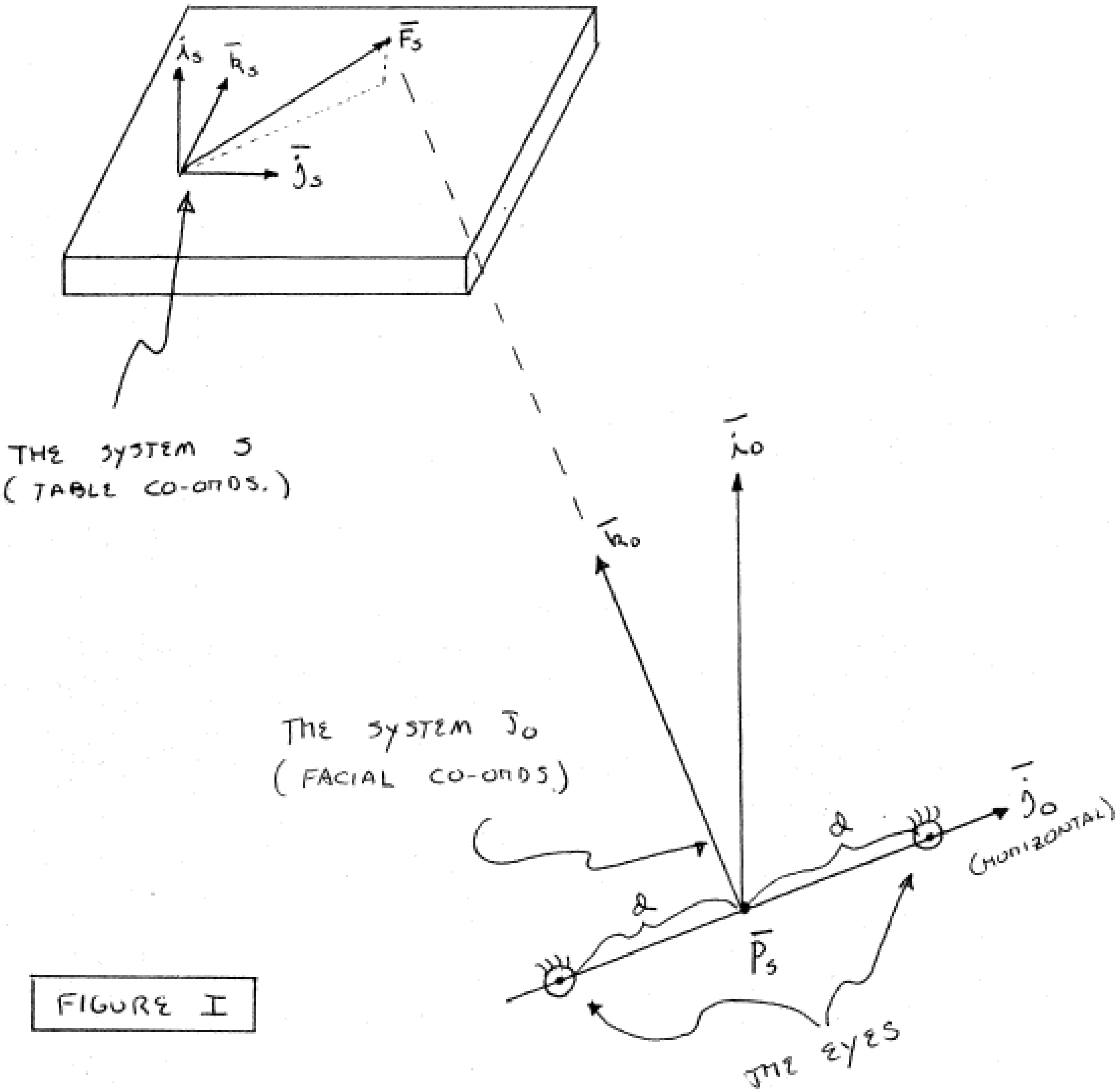

$\bar{i}_s$ $\bar{k}_s$ $\bar{F}_s$
$\bar{j}_s$
THE SYSTEM S
(TABLE CO-ORDS.)
$\bar{i}_o$
$\bar{k}_o$
THE SYSTEM J_o
(FACIAL CO-ORDS.)
d
$\bar{j}_o$
(HORIZONTAL)
d
$\bar{P}_s$
THE EYES

FIGURE I

It will be assumed that our eyes in the J_o system lie at $\pm\bar{D}$, where $\bar{D} \equiv d(\ 0,1,0\ )$, and it will be noted that $\bar{F}_s$ in this system appears to lies at $\bar{F} \equiv f(\ 0,0,1\ )$, where $f = |(\bar{F}_s - \ \bar{F}_s)|$. Thus, J_o in humans seems to be some sort of 'facial' system, with $\bar{k}$ out the nose, and $\bar{i}$ out the top of the forehead. Here we have made certain, though, that the eyes always focus on points only along the k-axis*, or 'straight-ahead".

We will be interested in finding from measured quantities the location of some point $\bar{X}_s$ in S, and it will simplify matters if we let $\bar{E}_s \equiv \bar{X}_s - \bar{F}_s$, and then find that vector instead. We note quickly that

$$\bar{E}_s \equiv \bar{X}_s - \bar{F}_s = \overset{*}{T}_{so}\, d\, \bar{\epsilon} \qquad (2\text{-}3a),$$

where

$$\bar{\epsilon} \equiv (\bar{X}_o - \bar{F})/d \qquad (2\text{-}3b),$$

and proceed to the problem of finding $\bar{\epsilon}$ (non-dimensional displacement from $\bar{F}$ in the 'facial' system) in terms of measureable quantities.

The quantities we will measure come from a comparison of the two images recorded by the eyes. I assume that these images are formed by the projection of distant points onto planes perpendicular to rays between the eyes themselves and

---

* It is possible to generalize and allow the eyes to focus on points other than straight-ahead, but the algebra becomes quite a bit more complicated. Since the 'head' can be moved, this doesn't seem a serious restriction.

the focal point, $\bar{F}$. Defining two new systems then, J_l and J_r, (see Fig. II), with origins located in J_o at $-\bar{D}$ and $+\bar{D}$ respectively, and oriented so that their k-axes point toward $\bar{F}$ and their i-axes remain aligned with $\bar{I}_o$, we can find the locations of any point $\bar{X}_o$ (in J_o) in the new systems as

$$\bar{X}_L = \overset{*}{T}_{LO}\left(\bar{X}_O + \bar{D}\right) = \overset{*}{T}_{LO}\left(d\bar{\epsilon} + \bar{F} + \bar{D}\right) \qquad (2\text{-}4a)$$

$$\bar{X}_R = \overset{*}{T}_{RO}\left(\bar{X}_O - \bar{D}\right) = \overset{*}{T}_{RO}\left(d\bar{\epsilon} + \bar{F} - \bar{D}\right) \qquad (2\text{-}4b)$$

where, (the upper signs applying to the left eye, etc.)

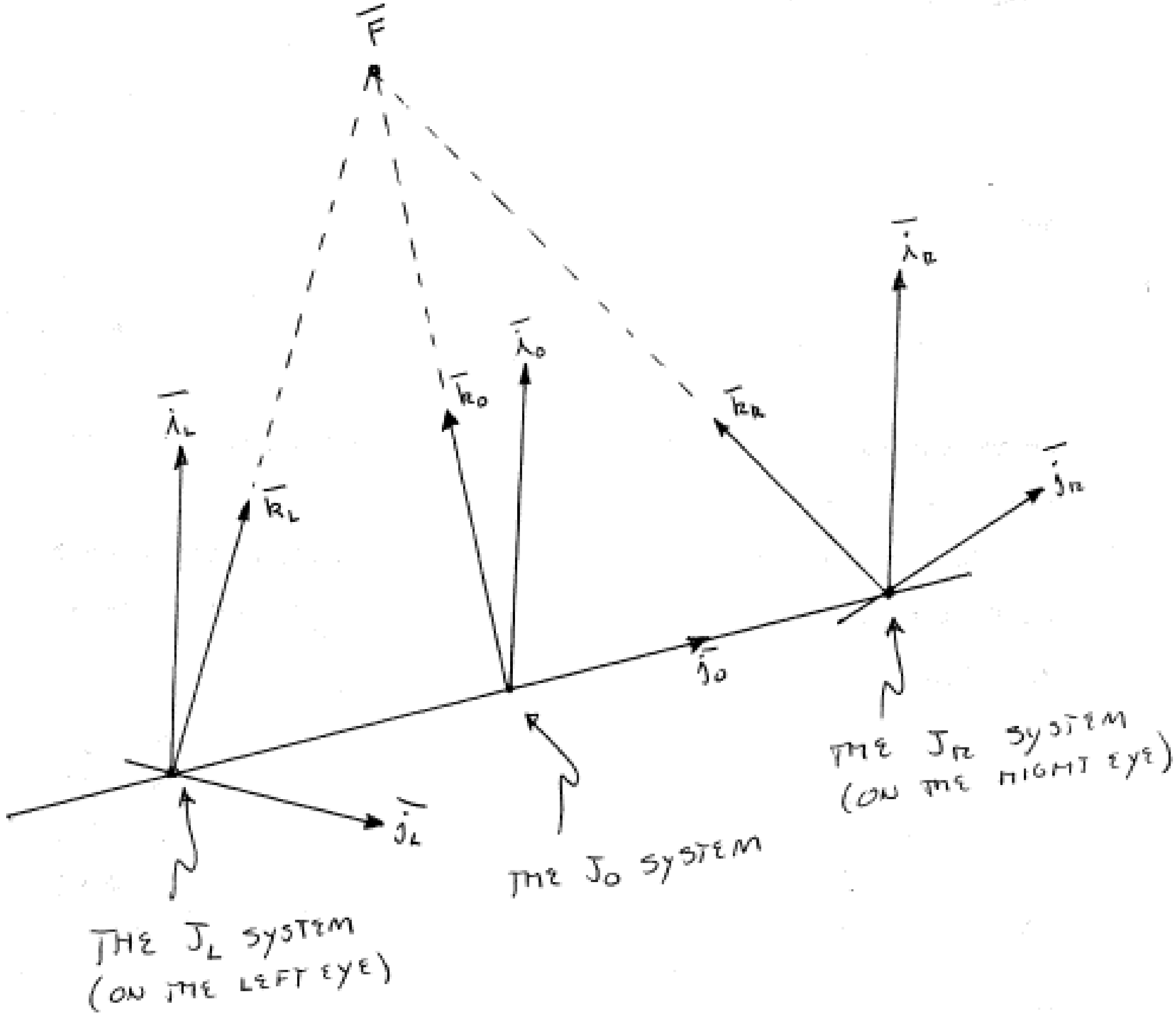
$$\overset{*}{T}_{\substack{LO\\RO}} \equiv \begin{bmatrix} \to \text{UNIT}\left(1,0,0\right) \longrightarrow \\ \to \text{UNIT}\left(\left(\bar{F} \pm \bar{D}\right) * (1,0,0)\right) \to \\ \to \text{UNIT}\left(\bar{F} \pm \bar{D}\right) \longrightarrow \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{f^2+d^2} & 0 & 0 \\ 0 & f & \mp d \\ 0 & \pm d & f \end{bmatrix} \frac{1}{\sqrt{f^2+d^2}} \qquad (2\text{-}5\,{}^{a}_{b})$$

If we define planes normal to $\bar{k}$ at (0,0,c) in both J_l and J_r, and then project a point $\bar{X}_o$ onto them, the intersections will occur at

$$c\left(\frac{x_{L,1}}{x_{L,3}}, \frac{x_{L,2}}{x_{L,3}}, 1\right) \equiv c\left(\alpha_L, \beta_L, 1\right) \qquad (2\text{-}6a)$$

in the 'left' system and plane,



$\bar{F}$
$\bar{i}_L$
$\bar{k}_L$
$\bar{j}_L$
$\bar{k}_0$
$\bar{i}_0$
$\bar{j}_0$
$\bar{k}_R$
$\bar{i}_R$
$\bar{j}_R$
THE J_L SYSTEM (ON THE LEFT EYE)
THE J_0 SYSTEM
THE J_R SYSTEM (ON THE RIGHT EYE)

FIGURE II

and at

$$\mathfrak{C}\left(\frac{x_{R,1}}{x_{R,3}}, \frac{x_{R,2}}{x_{R,3}}, 1\right) \equiv \mathfrak{C}\left(\alpha_R, \beta_R, 1\right) \qquad (2\text{-}6b)$$

in the 'right' system and plane.

$\mathfrak{C}$ is merely a scale factor determined by the optics alone. The key quantities, which define the location of the projections in the image planes of Part I, are, from (2-4) and (2-6),

$$\alpha_{\substack{L\\R}} = \frac{(\bar{X}_0 \pm \bar{\delta}) \cdot \text{UNIT}\,(1,0,0)}{(\bar{X}_0 \pm \bar{\delta}) \cdot \text{UNIT}\,(\bar{F} \pm \bar{\delta})} \qquad (2\text{-}7\,\substack{a\\c})$$

and

$$\beta_{\substack{L\\R}} = \frac{(\bar{X}_0 \pm \bar{\delta}) \cdot \text{UNIT}\,((\bar{F} \pm \bar{\delta}) * (1,0,0))}{(\bar{X}_0 \pm \bar{\delta}) \cdot \text{UNIT}\,(\bar{F} \pm \bar{\delta})} \qquad (2\text{-}7\,\substack{b\\d})$$

where again, and in what follows, the upper signs apply to the 'left' system.

It will be convenient to make several definitions, both to further non-dimensionalize the mathematics, and to save writing. We let

$$\alpha \equiv (\alpha_R + \alpha_L)/2 \qquad (2\text{-}8a)$$

$$\beta \equiv (\beta_R + \beta_L)/2 \qquad (2\text{-}8b)$$

$$\Delta \equiv (\beta_R - \beta_L)/2 \qquad (2\text{-}8c)$$

and

$$\phi \equiv f/d \qquad (2\text{-}8d)$$

and note that equations (2-5) now are

$$\overset{*}{T}_{\substack{LO\\RO}} \equiv \begin{bmatrix} \sqrt{\phi^2+1} & 0 & 0 \\ 0 & \phi & \mp 1 \\ 0 & \pm 1 & \phi \end{bmatrix} \frac{1}{\sqrt{\phi^2+1}}$$

and that equations (2-7), after some matrix algebra simplification, reduce to

$$\alpha_{\substack{L\\R}} = \frac{\epsilon_1 \sqrt{\phi^2+1}}{(\phi\epsilon_3+\phi^2+1) \pm \epsilon_2} \qquad (2\text{-}9\ \substack{a\\c})$$

$$\beta_{\substack{L\\R}} = \frac{\phi\epsilon_2 \mp \epsilon_3}{(\phi\epsilon_3+\phi^2+1) \pm \epsilon_2} \qquad (2\text{-}9\ \substack{b\\d})$$

Their usefulness now begins to become apparent, for with (2-8), we can solve for $\bar{\epsilon}$ in terms of $\alpha, \beta, \Delta, \phi$:

$$\epsilon_1 = \alpha\sqrt{\phi^2+1}\left[\frac{1 + \Delta/\phi - \beta^2/\phi(\phi+\Delta)}{1+\beta^2-\Delta^2-\Delta\phi+\Delta/\phi}\right] \qquad (2\text{-}10a)$$

$$\epsilon_2 = \beta\sqrt{\phi^2+1}\left[\frac{1 + (\sqrt{1+1/\phi^2}-1)}{1+\beta^2-\Delta^2-\Delta\phi+\Delta/\phi}\right] \qquad (2\text{-}10b)$$

$$\epsilon_3 = (\phi\Delta-\beta^2+\Delta^2)\sqrt{\phi^2+1}\left[\frac{1 + (\sqrt{1+1/\phi^2}-1)}{1+\beta^2-\Delta^2-\Delta\phi+\Delta/\phi}\right] (2\text{-}10c$$

It would be nice (as I've indicated by the weird forms of equations (2-10)) to simplify these with some approximations, since in most cases of interest,

$$\phi >> 1 \ ; \quad \beta^2 << 1 \ , \quad \Delta << 1 \qquad (2\text{-}11)$$

We can't do this just yet, though, because of the term $\Delta\phi$, whose magnitude is unclear. We can get a handle on it, though, from equation (2-3), where we noted that

$$\frac{\overline{X_0}}{d} \equiv \overline{\xi} = (\epsilon_1, \epsilon_2, \epsilon_3 + \phi)$$

$\overline{\xi}$ actually may be more useful to us than $\overline{\epsilon}$ in some cases, because it represents the actual (dimensionless) position of the point $\overline{X}_s$ in facial co-ordinates. We note that

$$\xi_1 = \epsilon_1$$

$$\xi_2 = \epsilon_2$$

and after some mathematics, we find that

$$\xi_3 = \epsilon_3 + \phi = \epsilon_3 \left[\frac{\phi^2 + 2\Delta\phi - \beta^2 + \Delta^2}{(\phi^2+1)(\Delta\phi - \beta^2 + \Delta^2)} \right]$$

This quantity, however, can be reduced with (2-11) to

$$\xi_3 \hat{=} \epsilon_3 \left[\frac{1}{\Delta\phi - \beta^2} \right]$$

and it is then clear that

$$\Delta\phi - \beta^2 \hat{=} \frac{\epsilon_3}{\xi_3} \qquad (2\text{-}12)$$

Several cases are possible and interesting:

A. $-\epsilon_3 >> \xi_3$ (focusing on infinity) -

In this case $(\phi\Delta - \beta^2)$ = 'a large negative number', and

$$\epsilon_1 = \xi_1 \hat{=} -\alpha/\Delta \qquad (2\text{-}13a)$$

$$\epsilon_2 = \xi_2 \hat{=} -\beta/\Delta \qquad (2\text{-}13b)$$

$$\epsilon_3 \hat{=} -\phi \qquad (2\text{-}13c)$$

$$\xi_3 \hat{=} -1/\Delta \qquad (2\text{-}13d)$$

Equation (2-13d) is the only interesting one. It allows us to find the depth of any point by just focusing on infinity and measuring Δ !

B. $|\epsilon_3| \ll \quad \xi_3 \doteq \phi$ (any point of nearly equal depth with the point of focus.)

In this case $(\Delta\phi - \beta^2)$ is a very small number $\ll 1$, and

$$\epsilon_1 = \xi_1 \mathrel{\hat{=}} \frac{\alpha\phi}{1-\Delta\phi} \mathrel{\hat{=}} \alpha\phi \tag{2-14a}$$

$$\epsilon_2 = \xi_2 \mathrel{\hat{=}} \frac{\beta\phi}{1-\Delta\phi} \mathrel{\hat{=}} \beta\phi \tag{2-14b}$$

$$\epsilon_3 \mathrel{\hat{=}} (\phi\Delta - \beta^2)\phi/(1-\Delta\phi) \mathrel{\hat{=}} (\phi\Delta-\beta^2)\phi \tag{2-14c}$$

$$\xi_3 \mathrel{\hat{=}} \phi \tag{2-14d}$$

Here it is the first three which are of interest. They allow via simple calculations for the deduction of position relative to a known focal point, ϕ.

C. $\epsilon_3 \geq \xi_3$ (see what follows)

In this case, $(\Delta\phi - \beta^2) \geq 1$, and a look at equations (2-10) may cause some mathematicians to worry about small denominators. Let them rest easily though. Physically this is impossible since

$$\epsilon_3 \geq \xi_3 \quad \text{implies} \quad \epsilon_3 \geq \phi + \epsilon_3 \quad \text{implies} \quad \phi \leq 0 \quad !$$

Negative focal lengths don't happen very often in practice.

Summarizing the results of this section, then, we have shown that in facial co-ordinates, whenever

$$\phi >> 1$$

$$\beta^2 << 1$$

$$\Delta << 1$$

it is approximately true that

$$\bar{\xi} = -(\alpha, \beta, 1)/\Delta \qquad \text{if} \quad \phi >> \xi_3 \qquad (2\text{-}15)$$

$$\bar{\epsilon} = (\alpha, \beta, (\phi\Delta - \beta^2)) \frac{\phi}{1-\Delta\phi} \qquad \text{if} \quad \xi_3 \doteqdot \phi \qquad (2\text{-}16)$$

The translation of these results from the facial system to the table system may be made through the use of

$$\bar{E}_s = \alpha \overset{*}{T}_{so} \bar{\epsilon} \qquad (2\text{-}17)$$

or

$$\bar{X}_s = \alpha \overset{*}{T}_{so} \bar{\xi} + \bar{P}_o \qquad (2\text{-}18)$$

where $\overset{*}{T}_{so}$ is given by equation (2-2).

III. Search Limitation in the Pattern Matching Problem

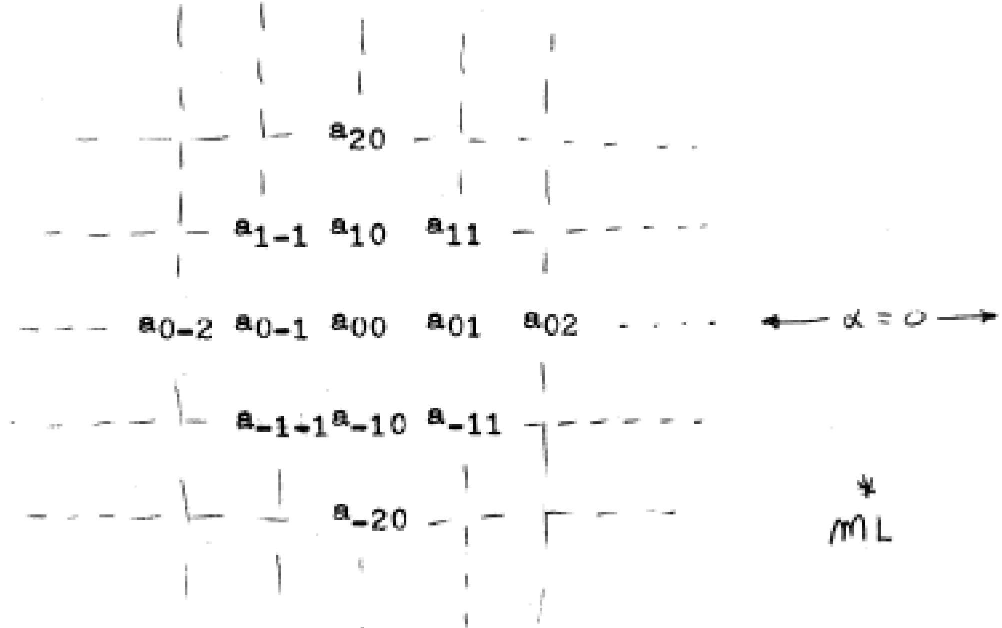
I really have no right to jump into this aspect of the problem too far, because I don't know enough of the hardware limitations and capabilities, but if my first assumption holds, then what follows should not be too far off the track, and since it's important, I should say something about it. If light-intensity measurements are indeed representable in a 2-D matrix for each eye, then informationally these matrices, $\overset{*}{ML}$ and $\overset{*}{MR}$, will look as is shown in Fig. III.

The matching by local evidence of elements within these matrices involves, I will assume, something procedurally akin to:

1. Plucking a local region out of one matrix.
2. Choosing an untested region of the other matrix.
3. Overlaying the local region on top of it.
4. Evaluating their local differences.
5. Iterating steps 2-4 until some cutoff occurs.
6. Choosing the match with the smallest differences.

I can't really be less vague here without knowing more about the hardware and noise aspects of the problem, but its easy to imagine something like a "minimize-the-sum-of-the-squares-of-the-differences-over-several-elements" approach, which would require that for some local region of N x N (N odd) elements,

$$f(\ell,k) = \sum_{m=-\frac{N-1}{2}}^{\frac{N-1}{2}} \sum_{\rho=-\frac{N-1}{2}}^{\frac{N-1}{2}} \left(b_{i+m+k,\, j+\rho+\ell} - a_{i+m-k,\, j+\rho-\ell} \right)^2$$



here each a_{ij}, b_{ij} represent average light intensity values over a small region $\delta_\alpha \times \delta_\beta$ and centered on $\alpha = i\delta_\alpha$, $\beta = i\delta_\beta$. δ_α and δ_β are the limits of resolution in the two directions,

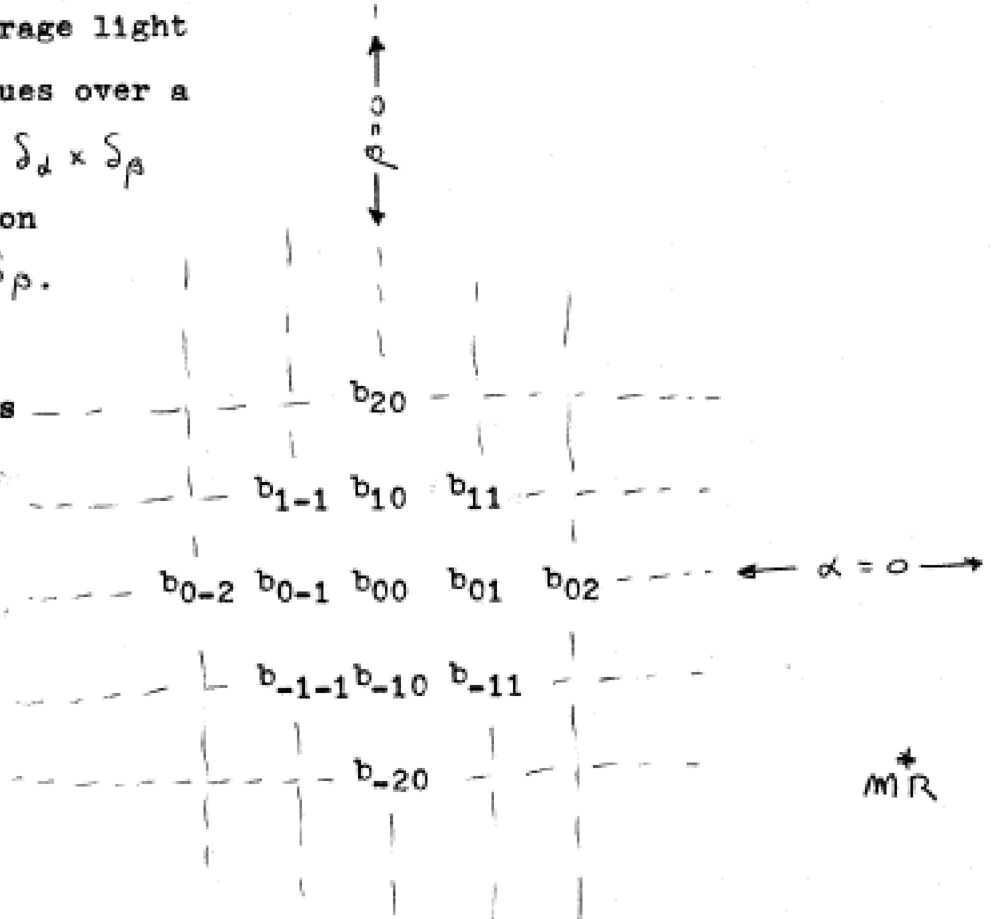


FIGURE III

be minimized by adjusting the k and l. In the minimum case, k and l would then represent half the quantized (integer) shifts,

$$k^* = \frac{\alpha_R - \alpha_L}{2\delta_\alpha} \quad , \quad l^* = \frac{\beta_R - \beta_L}{2\delta_\beta} \tag{3-1}$$

which existed in the 'combined' image matrix, M^* at

$$\alpha = i\,\delta_\alpha \qquad \text{and} \qquad \beta = j\,\delta_\beta$$

Regardless of the form taken by the 'evaluator' of step 4, however, other more basic questions remain;

1. What is an acceptable cut-off criterion;
2. In what order do we vary k and l;
3. How good is the answer we get?

Without getting too involved, it's easy to make some relevant observations using the results of Part II.

Equations (2-10), for instance, make use of the quantity

$$(\alpha_R + \alpha_L)/2 \equiv \alpha$$

but do not make any reference to the analogous quantiy

$$(\alpha_R - \alpha_L)/2$$

This quantity may be shown, however, to be redundant and, more importantly, quite small. From (2-9) and (2-11),

$$\frac{(\alpha_R - \alpha_L)}{2} = \alpha \frac{\epsilon_1}{\phi^2 + \phi\epsilon_3 + 1} \mathrel{\hat{=}} \frac{\alpha\beta}{\Phi} \ll \alpha$$

and it thus would be expected that

$$\frac{\alpha_R - \alpha_L}{2\delta_\alpha} = \frac{ij\,\delta_\beta}{\Phi}$$

To ensure

$$|k^*| = \left|\mathrm{round}\left(\frac{\alpha_R - \alpha_L}{2\delta_\alpha}\right)\right| < k'$$

then, it is only necessary to keep

$$\left|\frac{ij\,\delta_\beta}{\Phi}\right| < k' - 1$$

or

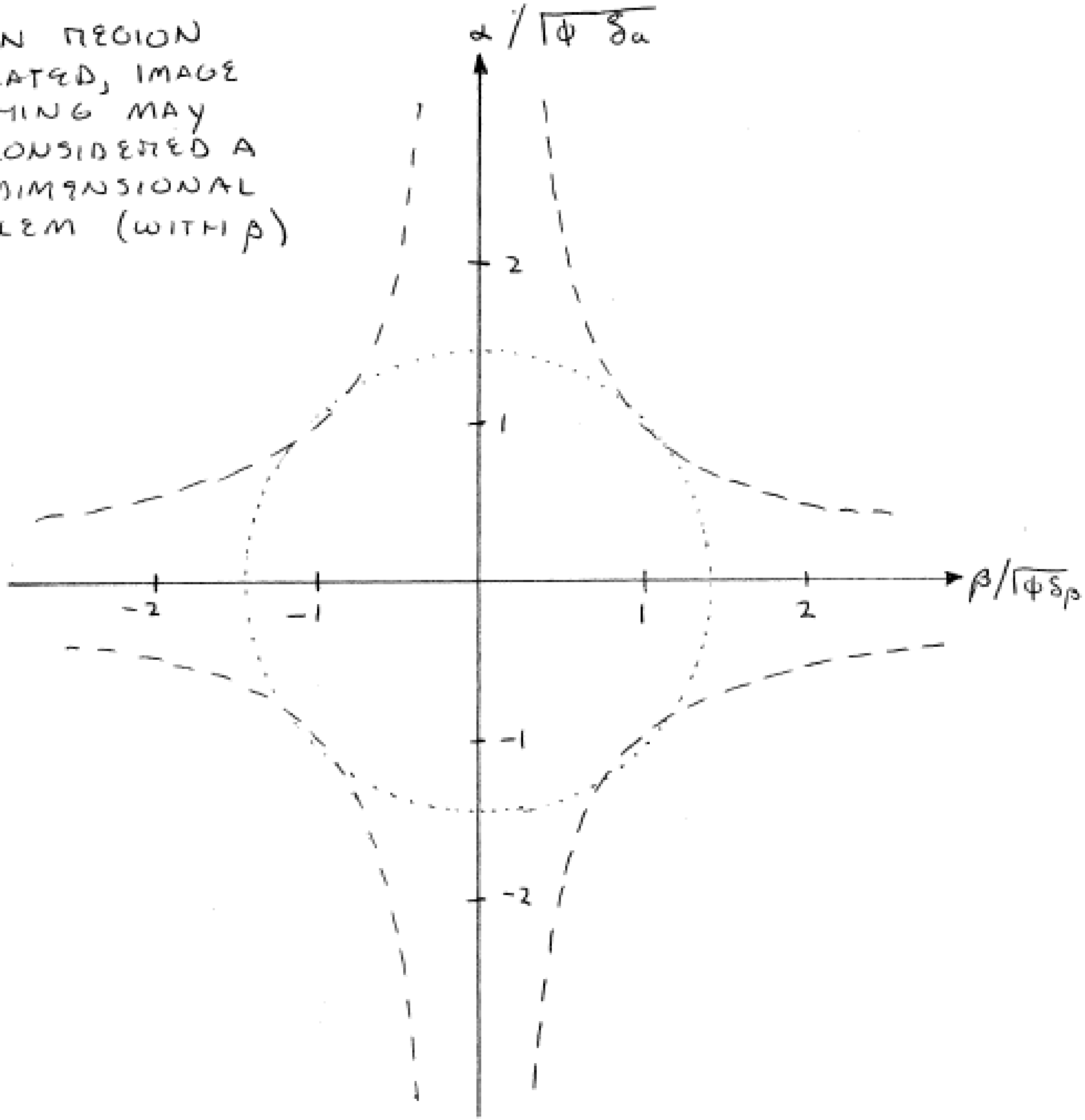
$$|ij| < \frac{\Phi}{\delta_\beta}(k'-1)$$

If we wish to let k'=2, this becomes

$$|\alpha\beta| < \Phi\delta_\alpha (k'-1) = \Phi\delta_\alpha \qquad (3\text{-}2)$$

Equation (3-2) defines a region (see Fig IV) in the image matrix, centered on the projection of the focal point at (0,0) and strictly bounded except along the axes. If we stay within this region, we will be assured that shifts in the α direction will be below the quantization level, and hence in our search we may neglect all k except the trivial case of k = 0.

WITHIN REGION INDICATED, IMAGE MATCHING MAY BE CONSIDERED A ONE-DIMENSIONAL PROBLEM (WITH β)



THEORETICAL REGION SHOWN DASHED —

POSSIBLE PRACTICAL ALTERNATE SHOWN DOTTED —

FIGURE IV

This limits our search within this region to one dimension only, along β, and states that if ever we wish to find Δ outside the region defined by (3-2)

$$|\alpha\beta| < \Phi\,\delta_\alpha$$

we must either go to a more complicated search or else shift the point of focus. Time tradeoffs based on speed would seem easy to develop.

We can also note that the approximate value of Δ (approximate value of the shift) should be predictable if nearby shifts are already known,* and if the region of 3-space corresponding to the local regions being compared contains no step discontinuities in Δ (depth). Δ should vary in a continuous and piecewise smooth manner along β, <u>except where 3-D object boundaries exist</u>! (see Part IV). Remaining "on" an object then, we would expect that 1 would be almost or exactly the same for adjacent points in the matrix, and we thus not only find a natural heuristic to help speed up search in these cases, but we have an immediate flag which signals object boundaries. Whether or not a high level of noise would trigger this flag too often, is something I haven't been able to realistically figure out.

---

*This will always be true if we seek Δ_{ij} in the following order: start next to the focal point, (where $\Delta \equiv 0$), and progress spirally outward. "Old" points will always be adjacent in the direction of the origin and "behind".

As far as cut-off is concerned, it is sort of hard to think of a heuristic that works equally well when the predictions are "good", and when they are not.

We might test predictions by looking exhaustively at some small number of points near the predicted one, and if the differences seem to be increasing as we move away in either direction, the prediction is probably "good", and we can use the best match found from this small set of data. If on the other hand, the "match" doesn"t seem particularly good anywhere along this line, then it is likely that we have crossed a discontinuity-of-depth in the scene viewed, and we have to do something strange. Perhaps it would be best to keep looking over greater and greater areas for a match, but perhaps not, for it is quite possible that a match cannot be found! One must remember, after all, that near regions of depth-discontinuity, one eye sees things that are hidden to the other eye.

What I would then propose for a Δ-finding algorithm would look, in a more refined form, something like:

1. Pick a new point of focus.

2. Plan an outward-spiralling path of examination beginning at the origin and remaning within the region given by (3-2).

3. Pick the next point on the path adjacent to a "good" point.

4. Predict $1'$ at this point based on nearby values.

5. Evaluate a small set of "overlays" shifted by about $21'$.

6. If a definite best fit exists near the center of this line it is probably "good". Record the shift, go back to step 3 again.

7. The shift is not good. Record this fact, and go back to step 3 anyway.

As soon as no new points can be found at step 3, a region will have been mapped out, and we can either stop or go back to step 1, depending on the information we need. If we go back, we record what we have been able to detect before moving on.*

I have one final comment to make on "noise". In line-drawing type programs, noise means extraneous variations in light intensity relative to 'average' over planar surfaces, and so the easiest objects to work with are smooth and uniformly colored. In the kind of pattern recognition program I've mentioned here, 'smooth and uniform' blocks are obviously terrible to work with, for locally the only variations in intensity are due to the inverse-square losses in light from a point source. What we really want for a local pattern-matcher is objects with lots of local detail (a light spray painting might be good). The kind of noise we can't tolerate is variations <u>between</u> the eyes when they look at the same small region of space. Any 'noise' picked up consistantly by <u>both</u> eyes will only make the pattern matching (and depth-perception) more efficient.

---

* It is interesting to compare this algoritm with the much more complete work of Lerman (1). Although derived independantly, and applicable to different eye configurations, both attempt to deal with similar effects, and the reader is encouraged to regard them as complementary. Basically, Lerman obtains a set of possible 'matches' by comparing intensity differences between the shifted image elements with a fixed cut-off, and then refines this 'possible set' to remove ambiguities and eliminate spurious points. His results are conceptually encouraging, but when applied to actual images take a great deal of time. If a shifting point of focus and a goal-oriented measurement scheme were to be incorporated, it is possible that a more widely applicable set of programs could be generated.

IV. General Remarks, and Figures

So far this has been pretty mathematical, and it may be interesting (and instructive) to see what these results look like when applied to human vision. People's eyes are about 2" apart, and so in what follows, d = 1", and distances may be interpreted either as dimensionless or in inches, since the numbers come out the same either way. (α , β , Δ , ϕ are always dimensionless, however.)

If you hold a pencil up at arms length (ξ_3 = 33") and focus on infinity, equation (2-15) predicts that

$$\Delta \doteq -\frac{1}{\xi_3} \doteq -.03$$

and since this is one half the difference between your right eye's image and your left eye's, you should see the "two pencil tips" shifted apart by about $\beta_R - \beta_L = -.06$. (The minus sign claims that your right eye is responsible for the left image. You can check this by blinking.) This gives you (or an uninitiated vision computer) a handle on the size of the β-scale, and I assume that the α-scale is the same. (So far the restriction to $\beta^2 \ll 1$ doesn't seem too limiting.)

Now try looking at something nearby and heavily textured, like a flower or a crumpled piece of paper. When I did this I found my eyes "jumped around" over the surface, making leaps of about $|(\alpha, \beta)| \doteq .03$ at $\phi = 10$. Equation (3-2) then fixes my approximate* limit of resolution such that

$$(.03)^2 = 10(k'-1)\delta\alpha$$

or,

$$\delta\alpha \cong \frac{(.03)^2}{10(k'-1)} \leq \frac{(.03)^2}{10} \cong 10^{-4}$$

This is about 40 seconds of arc (the thickness of a piece of newsprint at 20 feet), which sounds like the right ballpark at least. Computer 'eyes' won't be able to keep up with this kind of accuracy, and so we will have to expect some major differences in performance from depth-sensitive programs linked to any realistic hardware.

Pictures are also interesting, and I've included some in the following pages which I've taken the trouble to draw fairly accurately. One aside that strikes me as I look at them is that parallel lines don't come out looking very parallel, and yet I seem to remember the mention of some heuristics which made use of parallelism in line drawings. Perhaps their authors made different assumptions than I have.

These are line-drawing type pictures, even though a depth-sensitive program would be just as happy with curves, and wouldn't work at all without local detail in the planes themselves. I hope this doesn't bother anyone; curves and surfaces are hard to draw.

Figure V shows the scene from the top, as a perspectiveless

---

* I am assuming here that it is my 'depth-searcher' which is driving my point of focus around the object. Actually it might not be uniquely responsible, but on very irregular objects it seems likely it would be important. Also, I have assumed that my eyes 'jump' only when they reach the edge of the region defined by (3-2). If something more conservative was taking place, the limit of resolution would come out smaller.

blueprint included only for clarity. The other figures are self-explanatory. Things to look for include the sign and magnitude of Δ over the image, the basic scale of the two axes of the image, and the (very small) effects of β^2 in the prediction of ϵ_3. Whenever $\Delta-\beta^2/\phi$ is positive, the point indicated is farther away than is the focal point; When the reverse is true, it is closer. All 'right eye' images are shown dashed.

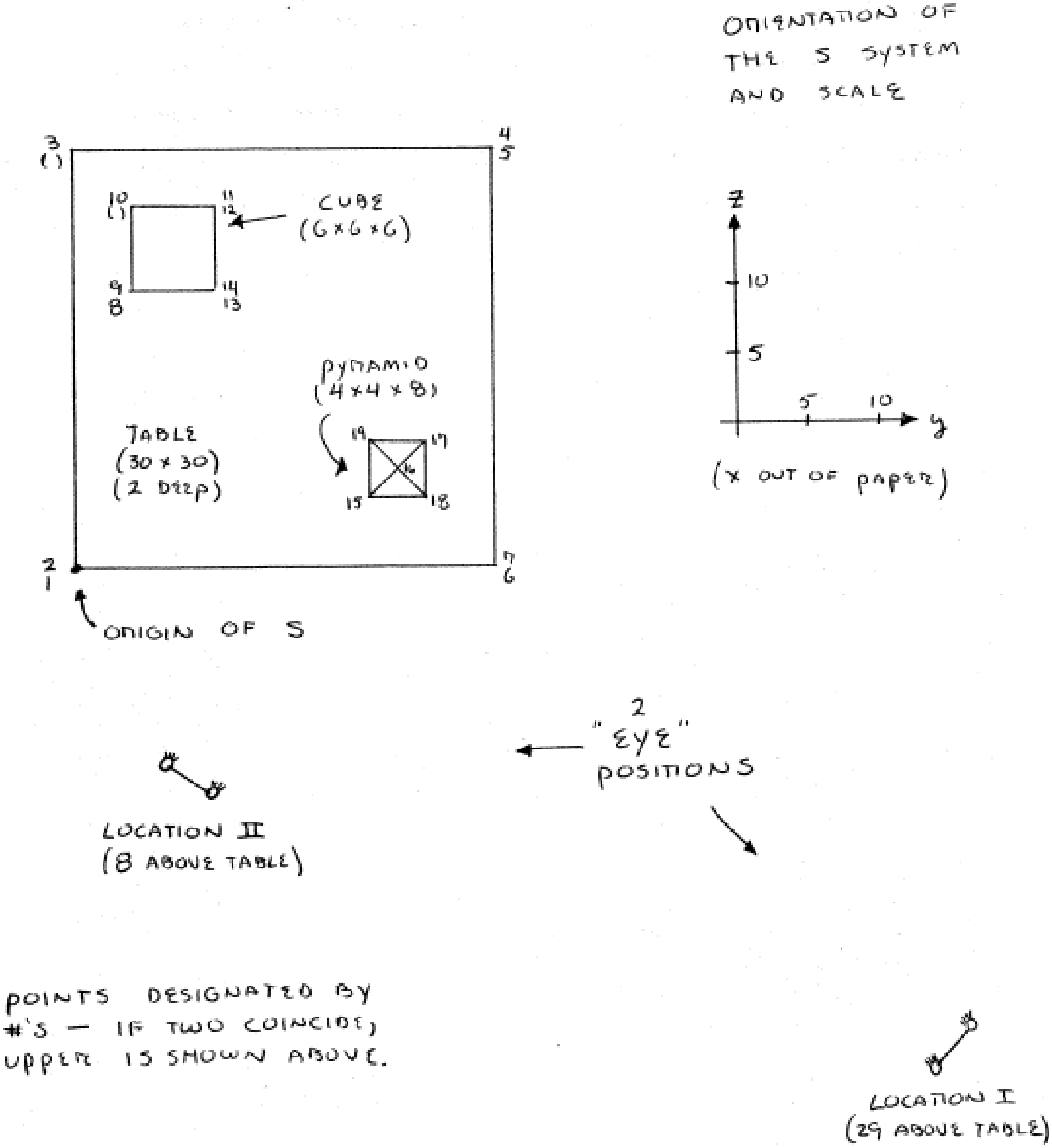
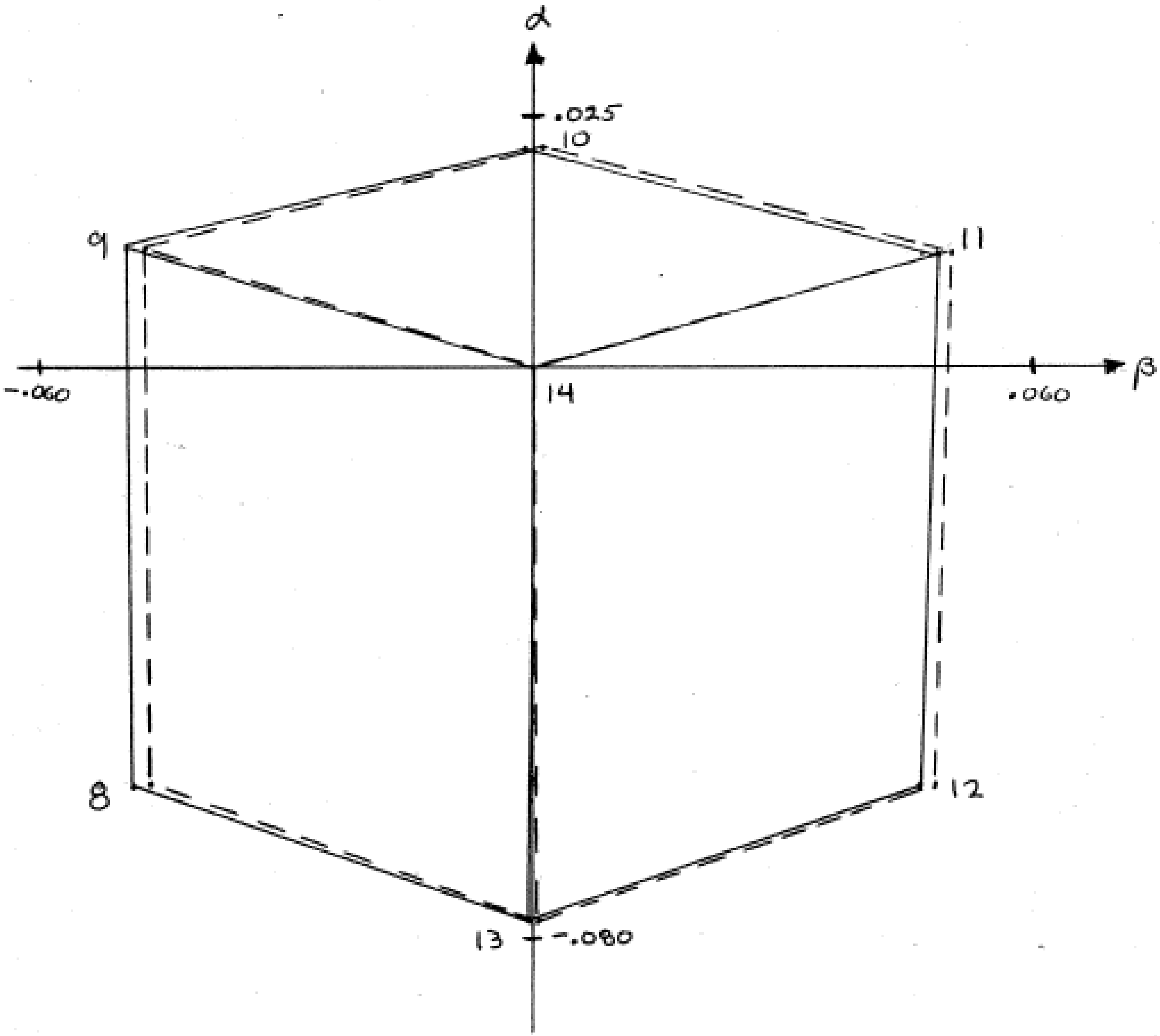

ORIENTATION OF THE S SYSTEM AND SCALE
CUBE (6×6×6)
PYRAMID (4×4×8)
TABLE (30×30) (2 DEEP)
ORIGIN OF S
z
y
(x OUT OF PAPER)
2 "EYE" POSITIONS
LOCATION II (8 ABOVE TABLE)
LOCATION I (29 ABOVE TABLE)
POINTS DESIGNATED BY #'S — IF TWO COINCIDE, UPPER IS SHOWN ABOVE.

FIGURE V

FIGURE VI



6" CUBE FROM 81" AWAY ($d = 1"$) (LOCATION I)

FOCUSING AT ORIGIN (PT # 14)

DASHED IMAGE SEEN BY RIGHT EYE

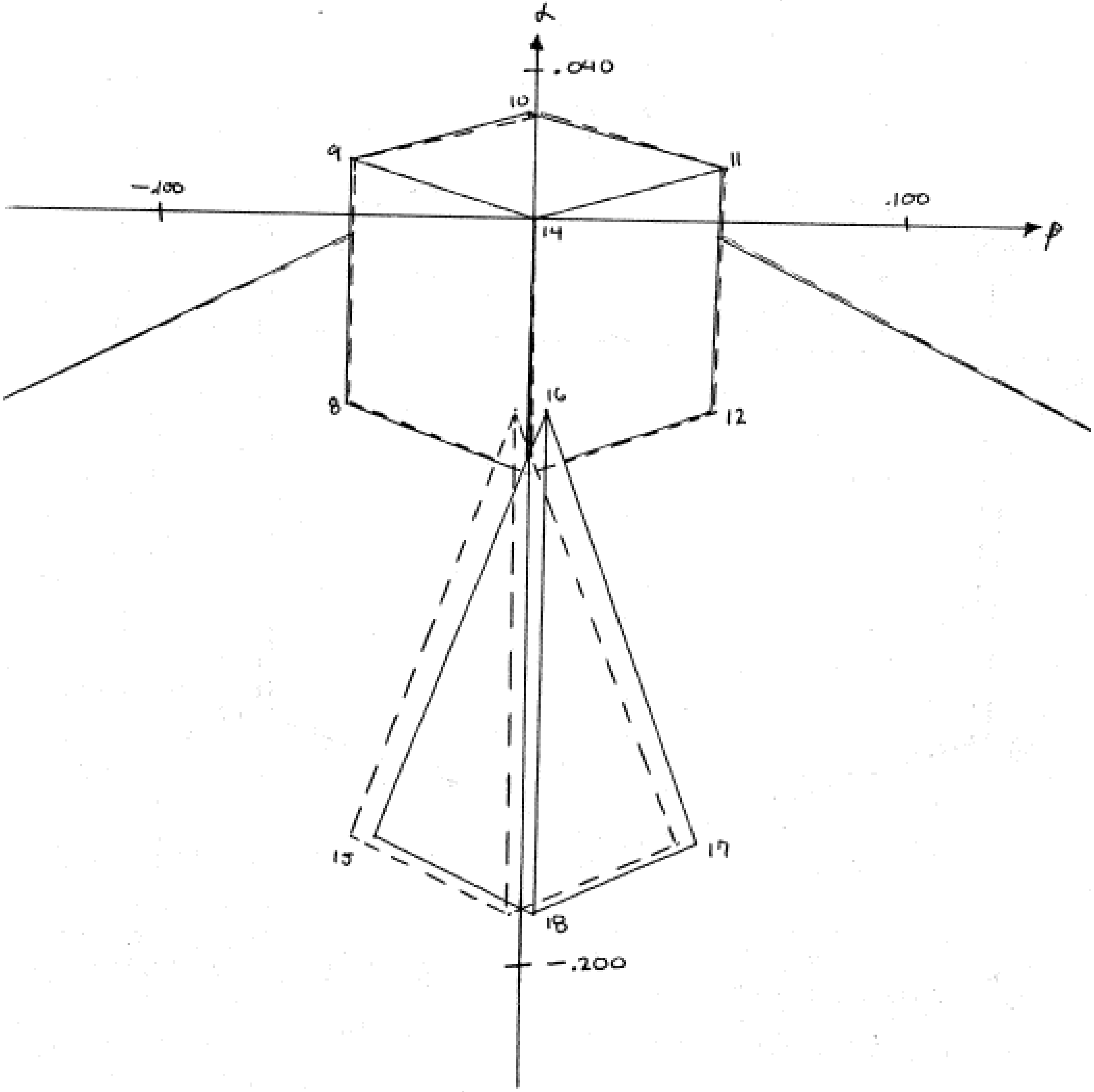
NOTE: MAGNITUDE OF α, β;

MAGNITUDE OF Δ

(RIGHT IMAGE "RIGHT" OF LEFT $\Rightarrow \epsilon_3 > 0$)

(HERE $\beta^2 \ll \phi\Delta$ EVERYWHERE)

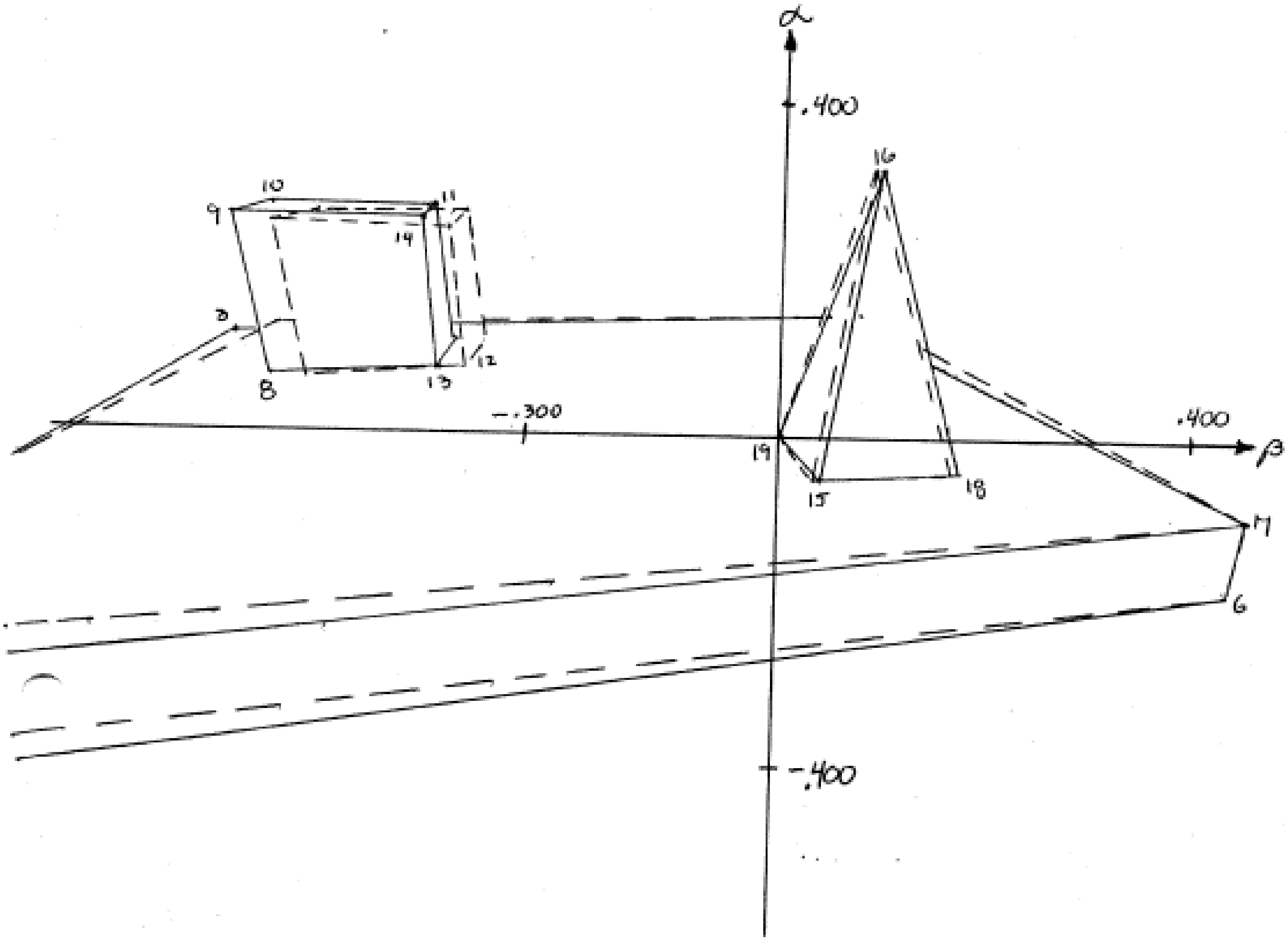
FIGURE VII



SAME PICTURE "BLOWN UP" AS FIG VI. —

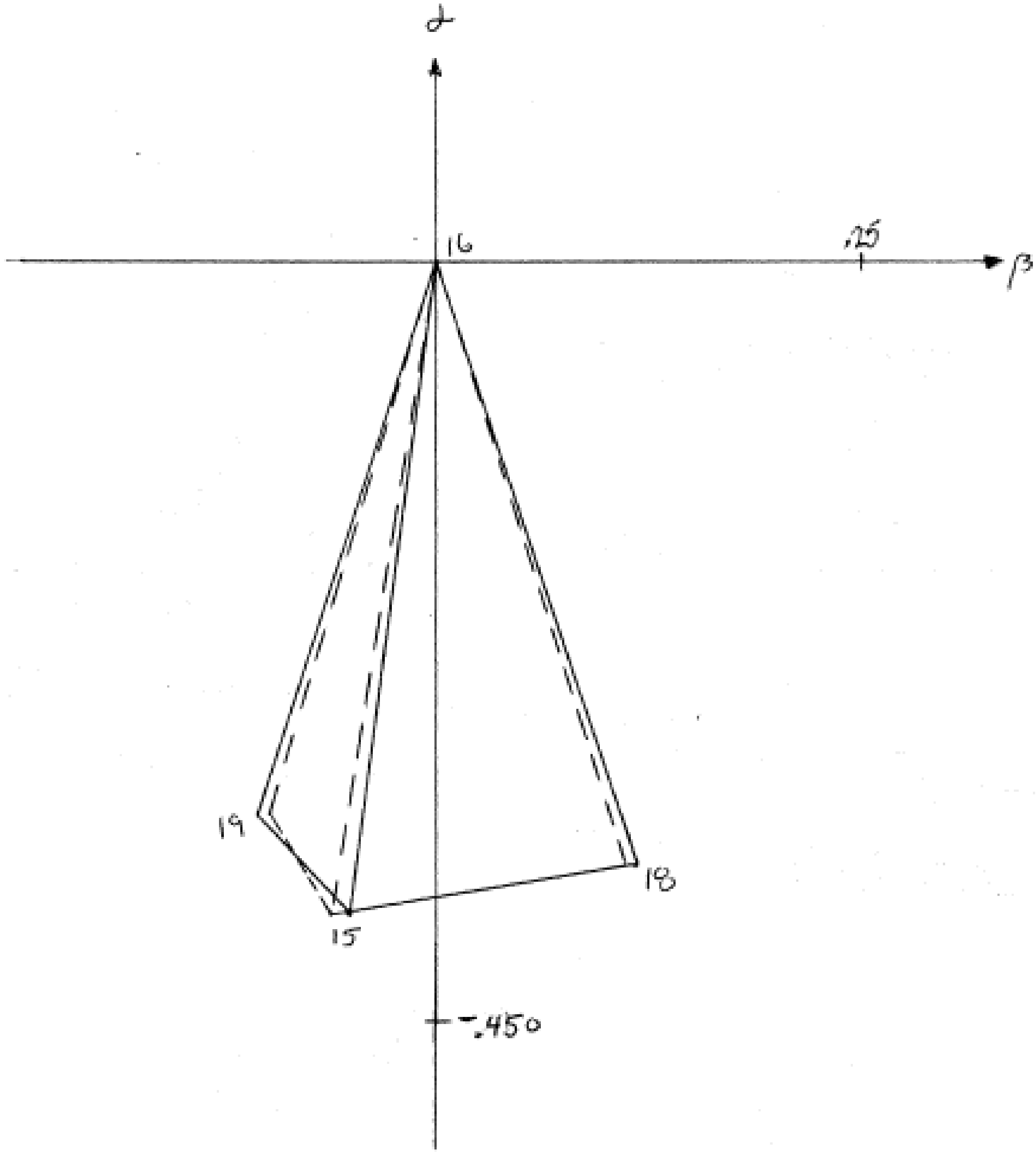
"β^2" DISTORTION BEGINS TO AFFECT $\Delta\phi - \beta^2$ NEAR EDGES —

NOTE SUBSTANTIAL SHIFT IN PYRAMID IMAGES —



STILL BIGGER SCALE DUE TO CLOSER POSITION (LOC. II) —
FOCUS LENGTH = 25 —
NOTE: BEGINNINGS OF α-SHIFT ON CUBE —
LARGE DISTORTION AWAY FROM FOCUS —

FIGURE VIII



8" PYRAMID FROM 24" — (LOCATION II)

NOTE CROSSOVER IN LINE 15-19 (≅ DEPTH WITH 16)

FIGURE IX

V. Error Analysis; Resolution Capability

Equations (2-10) represent solutions to relative 3-D displacement in terms of continuous and precise values of $\alpha, \beta, \Delta, \phi$. Approximations to these solutions have been given in equations (2-14), and are valid when the conditions (2-11) are satisfied. It remains to be seen, however, just how accurate these approximations are when fed the quantized data, $\alpha^*, \beta^*, \Delta^*, \phi^*,$ by a system with limits of resolution, δ. This section will examine such questions, and produce first order error and uncertainty estimates.

The errors inherent to equations (2-14) may be written down directly as the difference between the two sets of equations:

$$r_1 \equiv \epsilon_1^* - \epsilon_1 = \alpha^*\phi^* - \alpha\sqrt{\phi^2+1}\left[\frac{1+\frac{\Delta}{\phi}-\frac{\beta^2}{\phi(\phi+\Delta)}}{1+\beta^2-\Delta^2-\Delta\phi+\Delta/\phi}\right]$$

$$r_2 \equiv \epsilon_2^* - \epsilon_2 = \beta^*\phi^* - \beta\sqrt{\phi^2+1}\left[\frac{1+\left(\sqrt{1+1/\phi^2}-1\right)}{1+\beta^2-\Delta^2-\Delta\phi+\Delta/\phi}\right]$$

$$r_3 \equiv \epsilon_3^* - \epsilon_3 = \left(\phi^*\Delta^*-\beta^{*2}\right)\phi^* - \left(\phi\Delta-\beta^2+\Delta^2\right)\sqrt{\phi^2+1}\left[\frac{1+\left(\sqrt{1+1/\phi^2}-1\right)}{1+\beta^2-\Delta^2-\Delta\phi+\Delta/\phi}\right]$$

or,

$$r_1 = \left(\alpha+(\alpha^*-\alpha)\right)\left(\phi+(\phi^*-\phi)\right) - \alpha\sqrt{\phi^2+1}\left[\frac{1+\frac{\Delta}{\phi}-\frac{\beta^2}{\phi(\phi+\Delta)}}{1+\beta^2-\Delta^2-\Delta\phi+\Delta/\phi}\right]$$

$$r_2 = \left(\beta+(\beta^*-\beta)\right)\left(\phi+(\phi^*-\phi)\right) - \beta\sqrt{\phi^2+1}\left[\frac{\sqrt{1+1/\phi^2}}{1+\beta^2-\Delta^2-\Delta\phi+\Delta/\phi}\right] \qquad (5\text{-}1)$$

$$r_3 = \left[\left(\phi+(\phi^*-\phi)\right)\left(\Delta+(\Delta^*-\Delta)\right)-\left(\beta+(\beta^*-\beta)\right)^2\right]\left[\phi+(\phi^*-\phi)\right]$$
$$-\left(\phi\Delta-\beta^2+\Delta^2\right)\sqrt{\phi^2+1}\left[\frac{\sqrt{1+1/\phi^2}}{1+\beta^2-\Delta^2-\Delta\phi+\Delta/\phi}\right]$$

For any given values of $\alpha, \beta, \Delta, \phi,$ these may be regarded as functions of the variables

$$\alpha_\epsilon \equiv \alpha^* - \alpha \; ; \qquad \Delta_\epsilon \equiv \Delta^* - \Delta \; ;$$
$$\beta_\epsilon \equiv \beta^* - \beta \; ; \qquad \phi_\epsilon \equiv \phi^* - \phi \tag{5-2}$$

so that

$$\bar{r} = \bar{r}(\alpha_\epsilon, \beta_\epsilon, \Delta_\epsilon, \phi_\epsilon)$$
$$\hat{=} \; \bar{r}(0,0,0,0) + \frac{\partial \bar{r}}{\partial \alpha_\epsilon}\alpha_\epsilon + \frac{\partial \bar{r}}{\partial \beta_\epsilon}\beta_\epsilon + \frac{\partial \bar{r}}{\partial \Delta_\epsilon}\Delta_\epsilon + \frac{\partial \bar{r}}{\partial \phi_\epsilon}\phi_\epsilon + \cdots \tag{5-3}$$

when $\alpha_\epsilon, \beta_\epsilon, \Delta_\epsilon, \phi_\epsilon$ are small.

We can solve exactly for the partials in (5-3), and produce the results <u>in terms of measured quantities</u>:

$$\frac{\partial \bar{r}}{\partial \alpha_\epsilon} = [\phi^*, 0, 0] \; ; \quad \frac{\partial \bar{r}}{\partial \beta_\epsilon} = [0, \phi^*, -2\phi^*\beta^*]$$
$$\frac{\partial \bar{r}}{\partial \Delta_\epsilon} = [0, 0, \phi^{*2}] \; ; \quad \frac{\partial \bar{r}}{\partial \phi_\epsilon} = [\alpha^*, \beta^*, 2\phi^*\Delta^* - \beta^{*2}] \tag{5-4}$$

The constant term in (5-3) arises from the use of (2-11) and the assumption that $\epsilon_3 \ll \xi_3$. To first order terms, it may be written as

$$\bar{r}(0,0,0,0) \; \hat{=} \; -\epsilon_3\left[\alpha, \beta, \frac{\epsilon_3}{\phi - \epsilon_3}\right]\frac{\phi}{\phi - 2\epsilon_3} \tag{5-5}$$

Combining (5-4) and (5-5), and setting

$$\alpha_\epsilon, \beta_\epsilon = \frac{\delta}{2}, \quad \Delta_\epsilon = \delta$$

we obtain:

$$|r_1| < \left|\frac{\alpha\phi\epsilon_3}{\phi - 2\epsilon_3}\right| + \left|\alpha^*\phi_\epsilon\right| + \frac{\phi^*\delta}{2}$$

$$|r_2| < \left|\frac{\beta\phi\epsilon_3}{\phi - 2\epsilon_3}\right| + \left|\beta^*\phi_\epsilon\right| + \frac{\phi^*\delta}{2}$$

$$|r_3| < \left|\frac{\epsilon_3^2\phi}{(\phi-\epsilon_3)(\phi-2\epsilon_3)}\right| + \left|(2\phi^*\Delta^* - \beta^{*2})\phi_\epsilon\right| + \phi^{*2}\delta$$

or approximately, with

$$\epsilon_3 \ll \phi \doteq \phi^*, \quad \alpha \doteq \alpha^*, \quad \beta \doteq \beta^*$$

a still simpler form,

$$\left|\frac{r_1}{\phi^*}\right| < \left|\frac{\alpha^*\epsilon_3}{\phi^*}\right| + \left|\frac{\alpha^*\phi_\epsilon}{\phi^*}\right| + \frac{\delta}{2}$$

$$\left|\frac{r_2}{\phi^*}\right| < \left|\frac{\beta^*\epsilon_3}{\phi^*}\right| + \left|\frac{\beta^*\phi_\epsilon}{\phi^*}\right| + \frac{\delta}{2}$$

$$\left|\frac{r_3}{\phi^*}\right| < \left(\frac{\epsilon_3}{\phi^*}\right)^2 + 2\left|\frac{\epsilon_3\phi_\epsilon}{\phi^{*2}}\right| + \phi^{*2}\delta$$

With suitable restrictions on the use of the techniques of measurement, it would be expected that the first two terms in each of the above could be held arbitrarily small. The fundamental limitation on accuracy in position measurement, however, is fixed by the quantization level, and is represented by the third terms; approximately,

$$|\bar{r}| < \phi\delta\,[.5,\ .5,\ \phi] \qquad (5\text{-}6)$$

While this limitation is small in its effect on horizontal and vertical position measurements, its effect on range resolution is not, and for $\delta = .001$, the limits of range resolution near the focal point may be found as a function of ϕ to be

ϕ	r_3/ϕ
10	.01
20	.02
50	.05
100	.10
200	.20
500	.50
1000	1.0

References

(1) Lerman, Jerome; Computer Processing of Stereo Images for the Automatic Extraction of Range; MS/BS thesis, MIT EE Dept., May, 1970