Pattern Recognition Letters 13 (1992) 813-820
North-Holland

November 1992

Optimal discretization for stereo reconstruction

Anup Basu

Alberta Center for Machine Intelligence and Robotics, Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada T6G 2H1

Received 15 January 1992
Revised 1 April 1992

Abstract

Basu, A., Optimal discretization for stereo reconstruction, Pattern Recognition Letters 13 (1992) 813-820.

A significant amount of research has been done on designing sensors for digitizing visual images. Over the years vidicon and CCD technologies have been developed to improve sensor performance. However the problem of how the sensors should be distributed over a two-dimensional array has been largely overlooked. Sensor resolution is often determind by industry and international standards, and has little to do with the problem for which it is being used. In this work we investigate the optimal horizontal and vertical resolution (given the total resolution) for solving the problem of stereo-reconstruction. We show that the best arrangement of sensor elements depends on both the parameters of a stereo system and the assumptions on the 3-D scene.

Keywords. Pixel placement, stereo, sampling theorem.

1. Introduction

Design of video cameras is essentially based on two types of technologies: vacuum-tubes (used by vidicons) and semi-conductors (used by charge-coupled devices). Vidicon devices use a photosensitive layer of millions of cells, each representing a tiny capacitor whose charge depends on the incident light. The basic unit of a charge-coupled device (CCD) is an analog shift register consisting of capacitors placed close together. An analog representation of the incident light is obtained by the accumulation of charge in the capacitors. (For further detail see [8].) Presently, most solid-state cameras are designed using charge-coupled devices. CCD cameras reduce the lag significantly compared to vidicon devices, and can also detect light in a larger wavelength (ultra-violet to infra-red).

Even though a great deal of research has been done on designing sensor elements, there is no previous work addressing the problem of how these sensor elements should be arranged in a two-dimensional (2-D) array. The resolution and number of pixels (picture elements) along the X- and Y-axes (for a 2-D array of sensors), are usually arbitrarily determined according to industry and international standards. Such standards have little (if anything) to do with the underlying problem being solved. Often the only criterion used to obtain

Correspondence to: A. Basu, Alberta Center for Machine Intelligence and Robotics, Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada T6G 2H1.

This work is supported in part by the Canadian Natural Sciences and Engineering Research Council under Grant OGP0105739.

the minimum resolution along a scan line is the Shannon sampling theorem. This theorem determines how far apart pixels should be to allow digital to analog conversion of a processed image without the *aliasing* effect. (Detailed analysis on this topic can be found in [5].) In this paper we analyze the problem of pixel distribution (in the vertical 'Y' and horizontal 'X' directions) given the total resolution (number of pixels) when we are solving the problem of stereo reconstruction. The problem of obtaining the structure of the scene by corresponding features in images has received a great deal of attention in the past [1–4, 6, 7].

In stereo vision problems we are given a pair of cameras, and typically we need to obtain parameters in the 3-D scene based on corresponding features between the left and right images. The position of a 3-D structure (or point) projected onto an image is estimated by the location of the nearest pixel in the image. This results in an approximation that is commonly referred to as the 'discretization error'. The maximum possible magnitude of this error depends on the distance between two adjacent pixels, and may be different for the x and y components in the image. The remaining portion of this work will study the problem of minimizing the error in estimating the 3-D position of structures in a scene. This will be done by relating the discretization error in an image to the estimation error in the scene.

Section 2 introduces some notation used in this paper. Section 3 derives the best pixel placement when the object under consideration is simply a point. The results obtained in Section 3 are generalized for an arbitrary object (assuming constraints on its location) in Section 4. Some experimental results showing the resolution distribution in 2-D for various values of the parameters, are described in Section 5.

2. Notation

Before we proceed with the solution, the following terms and notation used in this paper need to be defined.

- f: focal length of camera.
- (X, Y, Z): 3-D point.
- $(\hat{X}, \hat{Y}, \hat{Z})$: estimated position of the 3-D point.

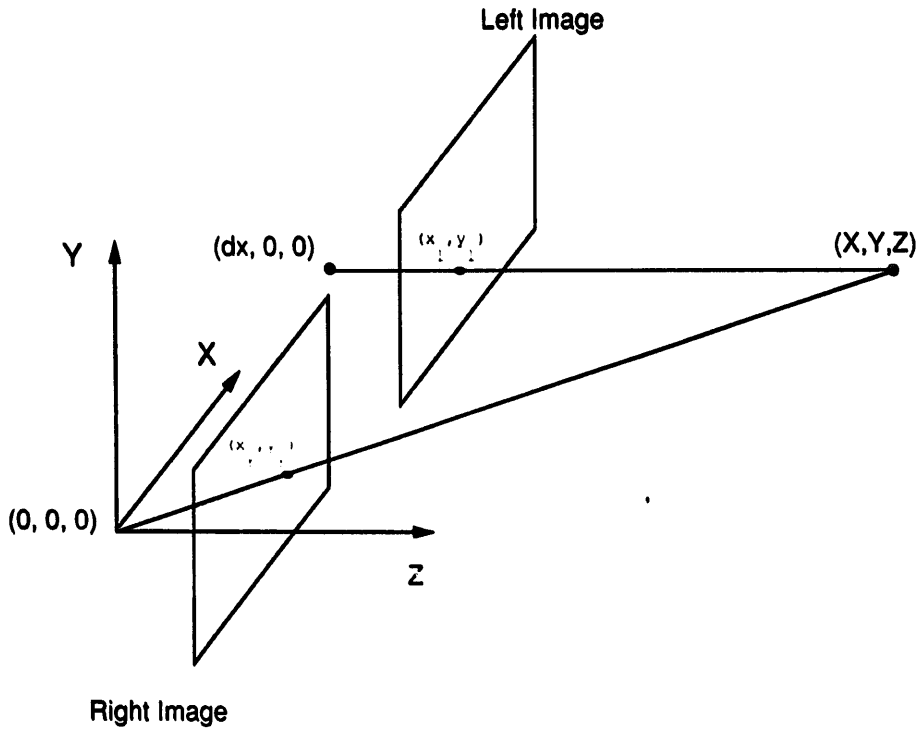

Figure 1. Imaging configuration.

- (x_l, y_l): projection of the 3-D point on the left image.
- (x_r, y_r): projection of the 3-D point on the right image.
- $(\hat{x}_l, \hat{y}_l)$: estimated projection of the 3-D point on the left image. This approximation occurs due to the discrete placement of pixels.
- $(\hat{x}_r, \hat{y}_r)$: estimated projection of the 3-D point on the right image.
- e_x: distance between two neighboring pixels along the x-direction.
- e_y: distance between two neighboring pixels along the y-direction.
- dx: distance between the nodal points of the left and right cameras.
- R: resolution, i.e., total number of pixels in the sensor array.

To better understand some of the above terms please see Figure 1. Also, Figure 2 describes how the discretization errors occur during image formation.

3. Optimal estimation of location of a 3-D point

The relation of a 3-D point with respect to its image projections is given by:

$$x_r = \frac{fX}{Z}, \qquad x_l = \frac{f(X-dx)}{Z}, \qquad y = y_l = y_r = \frac{fY}{Z}. \tag{1}$$

From (1) it follows that:

$$Z = \frac{f\,dx}{x_r - x_l}.$$

Using this relationship the depth is usually estimated as:

$$\hat{Z} = \frac{f\,dx}{\hat{x}_r - \hat{x}_l}. \tag{2}$$

Also, the X, Y values of a 3-D point are estimated by:

$$\hat{X} = \hat{Z}\frac{\hat{x}_r}{f}, \qquad \hat{Y} = \hat{Z}\frac{\hat{y}_r}{f}.$$

The discretized image points are at most within half a pixel of the actual projections. Thus, at worst:

$$\hat{x}_r = x_r \pm \frac{e_x}{2}, \qquad \hat{y}_r = y_r \pm \frac{e_y}{2}, \qquad \hat{x}_l = x_l \pm \frac{e_x}{2}, \qquad \hat{y}_l = y_l \pm \frac{e_y}{2}. \tag{3}$$

From now on we will obtain the worst case (maximum possible) error in the different components of the 3-D estimates. First let us consider

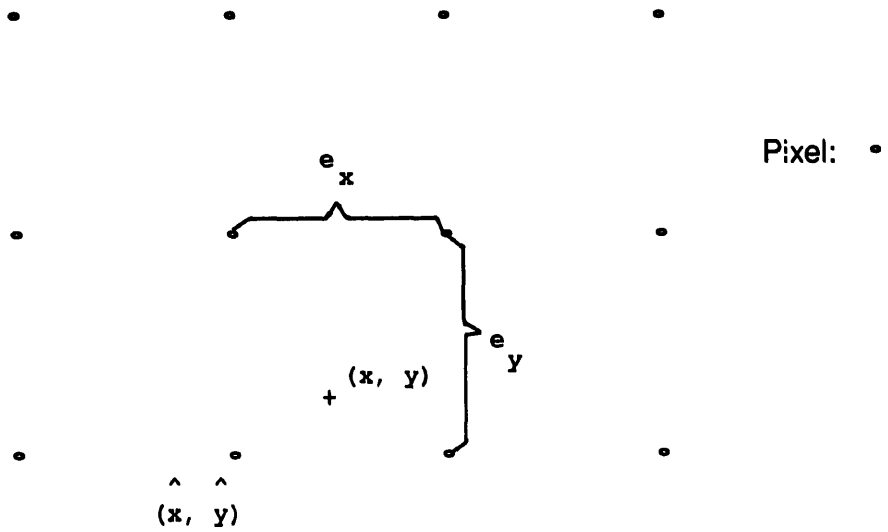

Figure 2. Discretization error.

the error in depth measurement. From (2) and (3) it follows that:

$$\hat{Z} = \frac{f\,dx}{(x_r - x_l) \pm e_x} = Z\left(1 \pm e_x \frac{Z}{f\,dx}\right)^{-1}.$$

Assuming $(e_x Z)/(f\,dx)$ is small, which is true if the depth is reasonable compared to the resolution of the image, we have:

$$\hat{Z} \cong Z\left(1 \pm \frac{e_x Z}{f\,dx}\right). \tag{4}$$

The error bounds on the estimate of X, can be obtained from:

$$\begin{aligned}\hat{X} &= \frac{\hat{Z}}{f} \cdot \hat{x}_r \\ &\cong \frac{Z}{f}\left(1 \pm \frac{e_x Z}{f\,dx}\right)\left(x_r \pm \frac{e_x}{Z}\right) \\ &= \frac{Z x_r}{f}\left(1 \pm \frac{e_x Z}{f\,dx}\right)\left(1 \pm \frac{e_x}{Z x_r}\right) \\ &= X\left(1 \pm \frac{e_x Z}{f\,dx} \pm \frac{e_x}{Z x_r} \pm \frac{e_x^2 Z}{2 f\,dx\, x_r}\right).\end{aligned}$$

Similarly,

$$\hat{Y} = Y\left(1 \pm \frac{e_x Z}{f\,dx} \pm \frac{e_y}{2y} \pm \frac{e_x e_y Z}{2 f\,dx\, y}\right).$$

Hence,

$$\begin{aligned}\left|\frac{\hat{Y} - Y}{Y}\right| &\leq \left\{\frac{e_x Z}{f\,dx} + \frac{e_y}{2|y|} + \frac{e_x e_y Z}{2f|y|\,dx}\right\} \\ &= \left\{\frac{e_x Z}{f\,dx} + \frac{1}{2Re_x|y|} + \frac{Z}{2fR\,dx\,|y|}\right\}.\end{aligned} \tag{5}$$

If R is the resolution (number of pixels) in a 1×1 region and e_x (e_y) is the horizontal (vertical) distance between pixels, then

$$\left(\frac{1}{e_x}\right)\left(\frac{1}{e_y}\right) = R$$

so

$$e_y = \frac{1}{Re_x}. \tag{6}$$

Minimizing the error in $\hat{X}$ or $\hat{Z}$ gives a solution which simply requires that the discretization size along the x-axis be as small as possible. However this is the worst possible solution for estimating Y. On the other hand, reducing error in $\hat{Y}$ gives a solution which is:

- optimal for estimating Y,
- better than uniform discretization for estimation of X and Z, in most practical situations.

Therefore, we will obtain the optimal discretization strategy of estimating Y.

Let

$$f(e_x) = \left|\frac{\hat{Y} - Y}{Y}\right|,$$

i.e.,

$$f(e_x) = \left\{\left(\frac{Z}{f\,dx}\right)e_x + \left(\frac{1}{2R|y|}\right)\frac{1}{e_x} + \frac{Z}{2fyR\,dx}\right\}.$$

Hence

$$f'(e_x) = \left\{\frac{Z}{f\,dx} - \frac{1}{e_x^2}\left(\frac{1}{2R|y|}\right)\right\}. \tag{7}$$

Equating the derivative to zero we have:

$$f'(e_x) = 0$$

so

$$\frac{1}{e_x^2}\left(\frac{1}{2R|y|}\right) = \frac{Z}{f\,dx}$$

thus

$$e_x = \frac{1}{\sqrt{R}}\sqrt{\frac{f\,dx}{2|y|Z}}. \tag{8}$$

By (6),

$$e_y = \frac{1}{\sqrt{R}}\sqrt{\frac{2|y|Z}{f\,dx}}. \tag{9}$$

[Note that if $e_x > 1$ in the above formula then we set $e_x = 1$ and $e_y = 1/R$. Similarly, if $e_y > 1$ we set $e_y = 1$ and $e_x = 1/R$.]

The formulae derived above are for specific values of Y and Z, i.e., they describe how the pixels should be distributed if our only goal is to estimate the 3-D position of a fixed point. In order to make the analysis meaningful, the solution needs to be applicable in some general environment. We will discuss this issue in the next section.

4. Optimal discretization with depth and range constraints

We will now generalize the above results under some assumptions on where an object of interest is located. The following two constraints are imposed on the 3-D scene:

A. Depth Constraint. The objects of interest are located in the depth range (Z_{min}, Z_{max}).

B. Range Constraint. The values of y in the image lie in a fixed range $(-y_{max}, +y_{max})$.

That is, the sensor array used to obtain a digital image is of a fixed known size.

A particular value of e_x will not be optimum for all points satisfying the above two constraints. Thus instead of solving (equation (7))

$$\left\{\frac{Z}{f\,dx}-\frac{1}{2Re_x^2}\cdot\frac{1}{|y|}\right\}=0 \quad \text{or} \quad \frac{f\,dx}{Z}=|y|2Re_x^2$$

we will instead minimize:

$$\left(\frac{f\,dx}{Z}-|y|2Re_x^2\right)^2$$

subject to the restrictions (A) and (B) above. That is, we want to find e_x minimizing

$$F(e_x)=\int_{Z_{min}}^{Z_{max}}\int_{-y_{max}}^{y_{max}}\left(\frac{f\,dx}{Z}-|y|2Re_x^2\right)^2 dy\,dZ.$$

The above equation can be rewritten as:

$$F(e_x)=(2Rfe_x^2)^2\iint y^2\,dy\,dZ$$

$$-(2Re_x^2)^{-1}f\,dx\iint\frac{|y|}{Z}\,dy\,dZ+I_3$$

where I_3 is a term independent of e_x and thus does not have any effect on the choice of e_x.

Let

$$I_1=\iint y^2\,dy\,dZ=(Z_{max}-Z_{min})\frac{2y_{max}^3}{3},$$

$$I_2=\int\frac{|y|}{Z}\,dy\,dZ=(\ln Z_{max}-\ln Z_{min})y_{max}^2.$$

Then

$$F(e_x)=4R^2e_x^4I_1-(2Rf\,dx)e_x^2I_2+I_3.$$

Hence

$$F'(e_x)=16R^2I_1e_x^3-4Rf\,dx\,I_2e_x.$$

Equating the derivative to zero, $F'(e_x)=0$, gives

$$16R^2I_1e_x^3=4Rf\,dx\,I_2e_x,$$

i.e.,

$$e_x^2=\left(\frac{f\,dx}{4R}\right)\left(\frac{I_2}{I_1}\right)$$

so

$$e_x^2=\frac{3}{8}\frac{f\,dx}{R}\frac{\ln(Z_{max}/Z_{min})}{y_{max}(Z_{max}-Z_{min})}$$

finally

$$e_x=\left\{\frac{3}{8}\frac{f\,dx}{R}\frac{\ln(Z_{max}/Z_{min})}{y_{max}(Z_{max}-Z_{min})}\right\}^{1/2}.$$

The formula above can be used to determine the best pixel distribution (given the total resolution R) for estimating the Y component of 3-D features under constraints (A) and (B) above. However, the experimental results in the next section demonstrate that the optimum discretization for estimating Y also happens to be better (in most cases) for obtaining $\hat{X}$ and $\hat{Z}$, than the case where pixels are distributed uniformly along both axes.

5. Experimental results

For all the experiments we will fix some of the parameters to values relating to actual stereo systems, and vary the depth range to study its effect on the pixel placement. We selected the focal length $f=2$ mm (or 0.2 cm), range of y values lying in -0.5 cm to 0.5 cm (i.e., a 1×1 cm imaging area is assumed), total resolution $R=400$ (20×20 resolution if pixels were distributed uniformly), and distance between the stereo pair of cameras to be 50 cm (half a meter distance between stereo cameras). The resulting pixel placement for depth range between 1 to 100 cm is shown in Figure 3. In this example the discretization error along the y-axis is about 3 times larger than the error along the x-axis. Therefore the errors in the estimates of X and Z components are also very much reduced compared to the case where pixels are placed uniformly along both axes.

Figure 4 shows the resolution distribution when the depth range is 1 cm to 2 meters (200 cm). In this

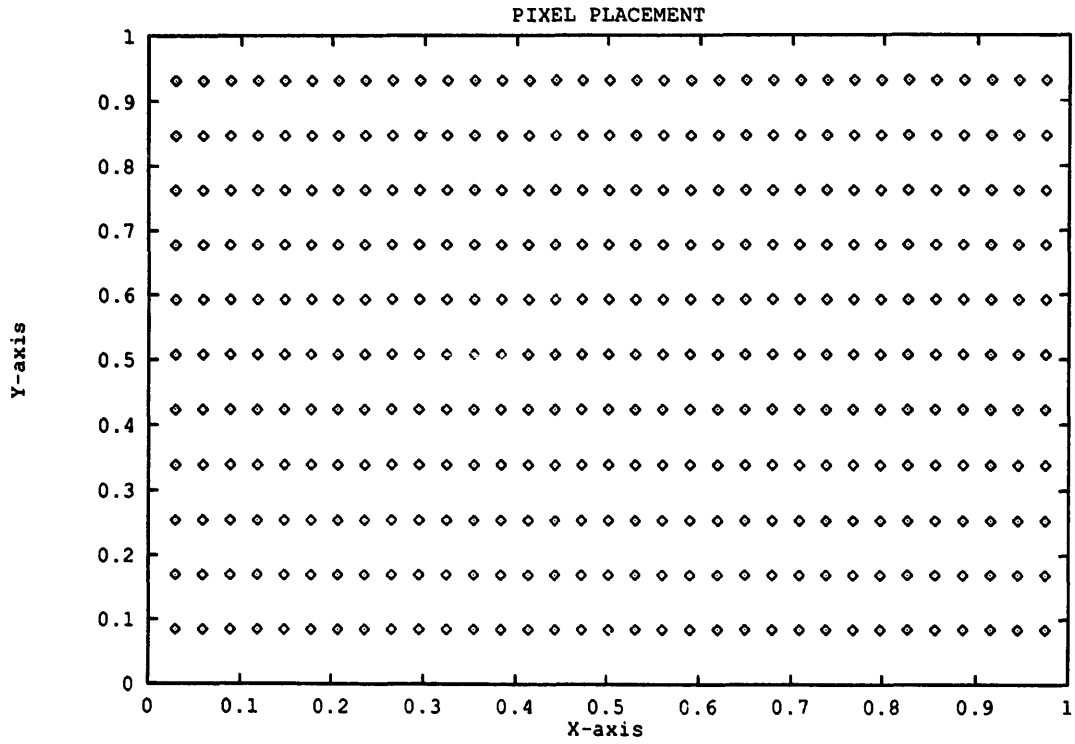

Figure 3. Depth range 1-100 cm.

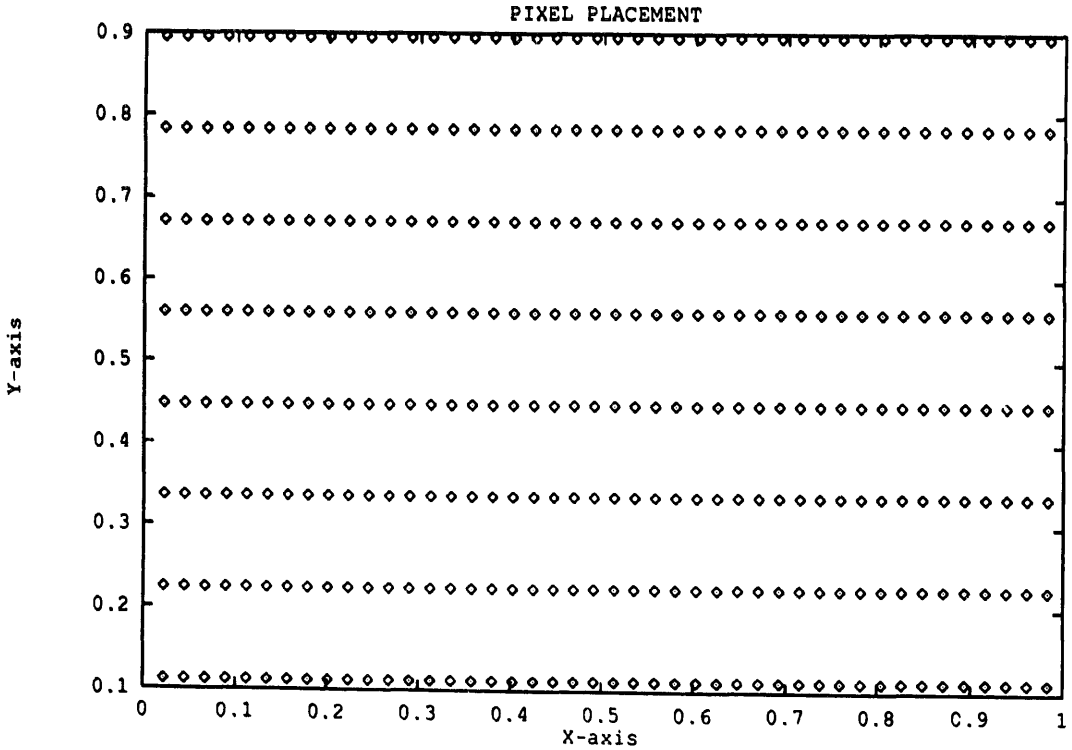

Figure 4. Depth range 1-200 cm.

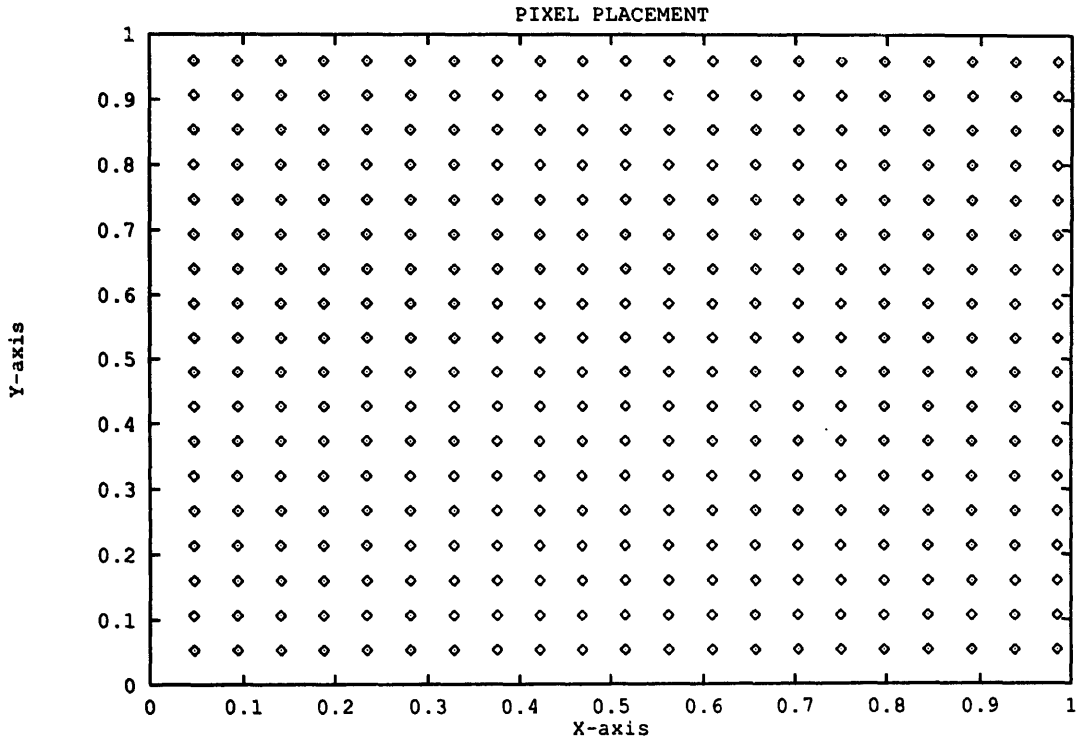

Figure 5. Depth range 1-30 cm.

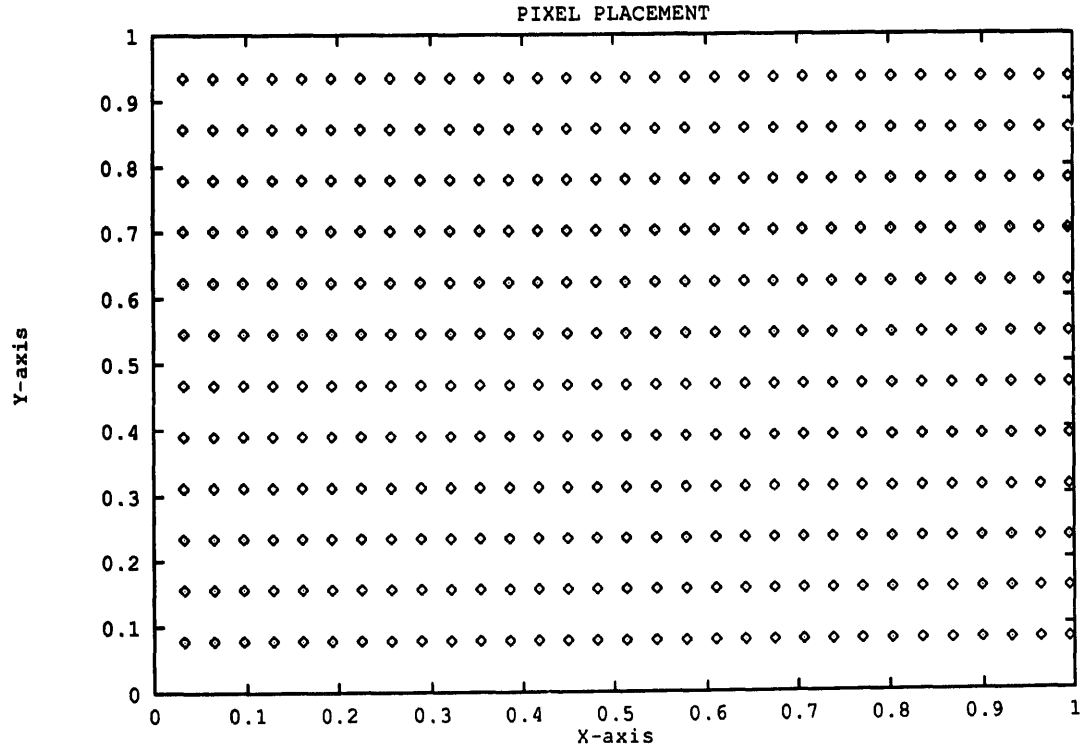

Figure 6. Depth range 10-30 cm.

case the distance between pixels along the Y-axis is more than 5 times that along the X-axis. There is a simple explanation for these results. As the depth of a point increases, to obtain a better estimate for Y we need a better estimate for Z, since $Y = yZ/f$. Estimate of the Z-component (also X-component) is improved if the discretization error along the x-axis is reduced. To further verify this argument, consider the next experiment: in Figure 5 the depth is between 1 to 30 cm, and the resulting discretization gives e_x approximately equal to e_y. The last example is unrealistic except for problems such as character recognition, where text at a close range needs to be interpreted.

Finally, the effect of increasing the minimum depth is shown in Figure 6. The depth range in Figure 6 is 10-30 cm, and the discretization size in the y direction is almost 2.5 times that in the x direction. This shows that e_x decreases rapidly as the minimum depth is increased.

6. Conclusion

We addressed the problem of determining the optimal method of distributing pixels (in a two-dimensional array) for the problem of stereo reconstruction. This type of analysis can be used, along with the Shannon sampling theorem, to obtain the best resolution for solving a class of vision problems. The derivations in this paper were based on the error in estimating only the Y-component of 3-D structures. The results can be generalized by minimizing the average error of all the three components (X, Y, Z) instead. This would make the resulting equations much more complicated, and is thus left for future research.

References

[1] Jenkin, M. and J.K. Tsotsos (1986). Applying temporal constraints to the dynamic stereo problem. *Computer Vision, Graphics and Image Processing* 33, 16-32.

[2] Kim, Y.C. and J.K. Aggarwal (1987). Positioning three-dimensional objects using stereo images. *IEEE Trans. Robot. Automat.* 3(4), 361-373.

[3] Marr, D. and T. Poggio (1979). A computational theory of human stereo vision. *Proc. Roy. Soc. London* B204, 301-328.

[4] Mayhew, J.E.W. and H.C. Longuet-Higgins (1984). A computational model for binocular depth perception. In: S. Ullman and W. Richards, Eds., *Image Understanding 1984*.

[5] Rosenfeld, A. and A.C. Kak (1982). *Digital Image Processing*. Academic Press, New York.

[6] Spetsakis, M.E. and J. Aloimonos (1990). Structure from motion from line correspondences. *Int. J. Computer Vision* 4, 171-183.

[7] Terzopoulos, D. (1983). Multiresolution computational process for visual surface reconstruction. *Computer Vision, Graphics and Image Processing* 24, 52-96.

[8] Vernon, D. (1991). *Machine Vision*. Prentice-Hall, London, UK.