Key words: machine vision, range estimation, image processing

J.W. BRUCE, Peter STUBBERUD and Ashok IYER[1]

RANGE ESTIMATION AND OBJECT IDENTIFICATION WITH A SINGLE CAMERA MACHINE VISION SYSTEM

Typically machine vision systems estimate ranges using stereoscopic vision systems or specialized range detectors. Because these methods are complex, they are not cost-effective. In this paper, a single camera range estimation algorithm is developed for a low-cost machine vision system that uses off-the-shelf components.

1. INTRODUCTION

Typically machine vision systems estimate ranges using stereoscopic vision systems or specialized range detectors. Although these methods are effective, they are expensive. In this paper, a single camera range estimation algorithm is developed for a low-cost machine vision system that uses off-the-shelf components. This algorithm uses the perspective transformation and *a priori* knowledge of scene's objects.

The perspective or imaging transformation maps a point, (X, Y, Z), from the three-dimensional world coordinate system into an image point, (u, v, w), in a two-dimensional image plane. Assuming that the single camera's optical axis lies along the w and Z axes as shown in Figure 1, the image plane lies at $(u, v, 0)$. In the Cartesian coordinate system, the perspective transformation is nonlinear. However, the perspective transformation can be linearized by mapping Cartesian coordinates into homogeneous coordinates. A point, $\mathbf{w}$, with Cartesian coordinates, (X, Y, Z), has the homogeneous coordinates

$$\mathbf{w}_h = \begin{bmatrix} kX & kY & kZ & k \end{bmatrix}^{\mathrm{T}} \tag{1}$$

where k is an arbitrary nonzero constant [1]. The perspective transformation matrix, $\mathbf{P}$, that maps a point in the four-dimensional world homogeneous coordinate system into an image point in a four-dimensional image homogeneous coordinate system can be written as

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -1/\lambda & 1 \end{bmatrix} \tag{2}$$

where λ is the focal length of the single camera's optics [1]. Therefore, given a world homogeneous point, $\mathbf{w}_h = \begin{bmatrix} k_w X & k_w Y & k_w Z & k_w \end{bmatrix}^{\mathrm{T}}$, an image homogeneous point, $\mathbf{c}_h$, can be calculated by

$$\mathbf{c}_h = \mathbf{P}\mathbf{w}_h . \tag{3}$$

1. The authors are with the Department of Electrical & Computer Engineering at the University of Nevada Las Vegas, 4505 Maryland Parkway, Las Vegas, NV 89154-4026 USA

The perspective transformation in (3) is valid as long as $Z > \lambda$.

The inverse perspective transformation matrix, $\mathbf{P}^{-1}$, can be written as

$$\mathbf{P}^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1/\lambda & 1 \end{bmatrix}. \tag{4}$$

Therefore, given an image homogeneous point, $\mathbf{c}_h = \begin{bmatrix} k_c u & k_c v & k_c w & k_c \end{bmatrix}^{\mathrm{T}}$, a world homogeneous location, $\mathbf{w}_h$, can be calculated by

$$\mathbf{w}_h = \mathbf{P}^{-1}\mathbf{c}_h. \tag{5}$$

To map a point in the four-dimensional homogeneous coordinate system into a point in the three-dimensional Cartesian coordinate system, divide the first three elements of vector describing the homogeneous coordinate by the fourth element in the same vector, and discard the fourth element. For example, consider the world point, $\mathbf{w}=(X, Y, Z)$, which has homogeneous coordinates $\mathbf{w}_h = \begin{bmatrix} kX & kY & kZ & k \end{bmatrix}^{\mathrm{T}}$. Using (3), the image homogeneous coordinates are

$$\mathbf{c}_h = \begin{bmatrix} kX & kY & kZ & k - \dfrac{kZ}{\lambda} \end{bmatrix}^{\mathrm{T}} \tag{6}$$

and the Cartesian point, $\mathbf{c}$, is

$$\mathbf{c} = \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} \dfrac{\lambda X}{\lambda - Z} \\ \dfrac{\lambda Y}{\lambda - Z} \\ \dfrac{\lambda Z}{\lambda - Z} \end{bmatrix}. \tag{7}$$

The transformation in (7) shows that mapping from world coordinates into image coordinates is a many-to-one mapping. Therefore, it is not possible to recover a point's world location from its image location without *a priori* information of its Z coordinate. To further illustrate, consider the image point, $\mathbf{c}=(u, v, 0)$, which has homogeneous coordinates $\mathbf{c}_h = \begin{bmatrix} ku & kv & 0 & k \end{bmatrix}^{\mathrm{T}}$. The zero in the third element of $\mathbf{c}$ simply indicates that the image plane is located at $w=0$. Using (5), the world homogeneous coordinates are

$$\mathbf{w}_h = \begin{bmatrix} ku & kv & 0 & k \end{bmatrix}^{\mathrm{T}} \tag{8}$$

and the Cartesian point, $\mathbf{w}$, is

$$\mathbf{w} = \begin{bmatrix} X & Y & Z \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} u & v & 0 \end{bmatrix}^{\mathrm{T}}. \tag{9}$$

The image point (u, v) corresponds to an infinite number of colinear world points lying on the line that passes through $(u, v, 0)$ and $(0, 0, \lambda)$.

The imaging transformation in (3) assumes that the optical axis is coincident with the w and Z axis as shown in Figure 1. If the camera is positioned with some angle of tilt as shown in Figure 2, the optical axis is not coincident with the w and Z axis. To rotate the homogeneous coordinate system so that the w and Z axes are coincident with the optical axis, premultiply by the rotation matrix, $\mathbf{R}$, where $\mathbf{R}$ is given by

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & \sin\theta & 0 \\ 0 & -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{10}$$

and θ is the angle of rotation about the *X* axis. Combining the rotation transformation in (10) and the perspective transformation in (3), the image homogeneous coordinates of the world homogeneous location when viewed by a tilted camera is found by

$$\mathbf{c}_h = \mathbf{PRw}_h. \tag{11}$$

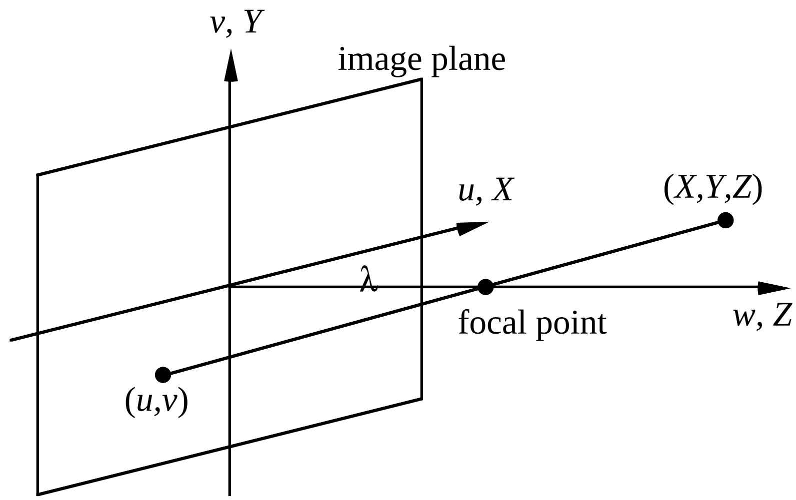

Figure 1: Basic imaging geometry

2. MONOCULAR RANGE ESTIMATION

As illustrated by (9), a machine vision system with a single camera requires *a priori* information to calculate range. In many practical applications, *a priori* knowledge of an object's size is known. By exploiting this knowledge, the world location of an object is calculated from its image location.

In Figure 3(a), an object with a known geometry is shown in world Cartesian coordinates. The object has two distinguishable points that vary only in the *X* direction. One point is located at (*X*, *Y*, *Z*) and the other is located at (*X*+Δ, *Y*, *Z*). The image plane view of this scene would be similar to the one shown in Figure 3(b). The image plane would have points (*u*, *v*) and (*u*+δ, *v*) corresponding to world points (*X*, *Y*, *Z*) and (*X*+Δ, *Y*, *Z*), respectively. After application of the perspective transformation, the image Cartesian coordinates of the first object point is given in (7). Applying (3) to the second object world location, converting to Cartesian coordinates and equating to the image Cartesian location gives

$$\begin{bmatrix} u+\delta \\ v \\ w \end{bmatrix} = \begin{bmatrix} \dfrac{\lambda(X+\Delta)}{\lambda - Z} \\ \dfrac{\lambda Y}{\lambda - Z} \\ \dfrac{\lambda Z}{\lambda - Z} \end{bmatrix}. \tag{12}$$

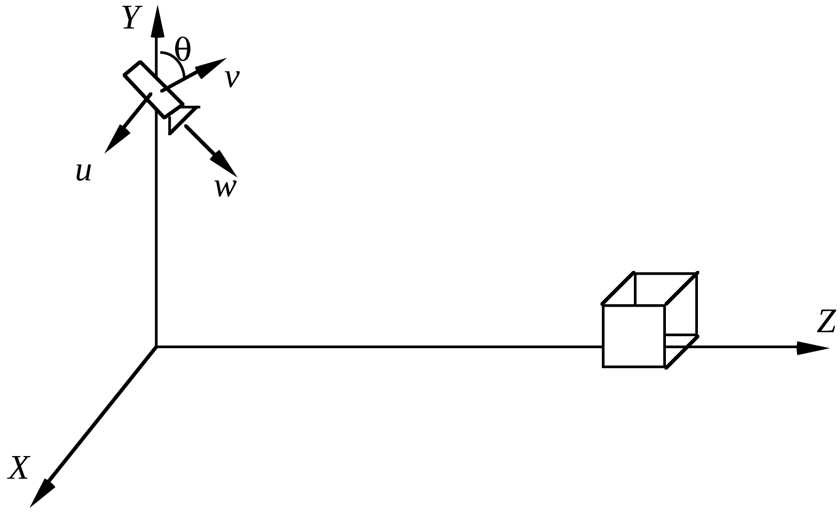

Figure 2: Imaging geometry with tilted camera

Using (7) and (12) gives

$$Z = -\lambda \frac{X + \Delta - u - u\Delta}{u + \delta} \tag{13}$$

Substituting (13) into (12) and solving for *X* and *Y* yields

$$X = \frac{u\Delta}{\delta} \tag{14}$$

$$Y = \frac{v\Delta}{u+\delta}\left(1 + \frac{u}{\delta}\right) \tag{15}$$

Using (13)-(15), the world location of an object can be determined from its image plane location using two points with a known relationship that vary in the *X* direction. Similar relationships exists for two object points which vary only in the *Y* and *Z* directions.

When a world location $\mathbf{w}_1=(X, Y, Z)$ is viewed by a camera tilted by an angle, θ, with respect to the *X* axis, its image Cartesian location can be calculated using (11). Converting the resulting homogeneous coordinates into the Cartesian coordinate system gives

$$\mathbf{c}_1 = \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} \lambda \dfrac{X}{Y\sin\theta - Z\cos\theta + \lambda} \\ \lambda \dfrac{Y\cos\theta + Z\sin\theta}{Y\sin\theta - Z\cos\theta + \lambda} \\ \lambda \dfrac{Z\cos\theta - Y\sin\theta}{Y\sin\theta - Z\cos\theta + \lambda} \end{bmatrix}. \tag{16}$$

If another known location, $(X+\Delta, Y, Z)$, is similarly mapped into the image location $(u+\delta, v)$ by the rotation and perspective transformation in (11), the image Cartesian location is

$$\mathbf{c}_2 = \begin{bmatrix} u+\delta \\ v \\ w \end{bmatrix} = \begin{bmatrix} \lambda \dfrac{X+\Delta}{Y\sin\theta - Z\cos\theta + \lambda} \\ \lambda \dfrac{Y\cos\theta + Z\sin\theta}{Y\sin\theta - Z\cos\theta + \lambda} \\ \lambda \dfrac{Z\cos\theta - Y\sin\theta}{Y\sin\theta - Z\cos\theta + \lambda} \end{bmatrix}. \tag{17}$$

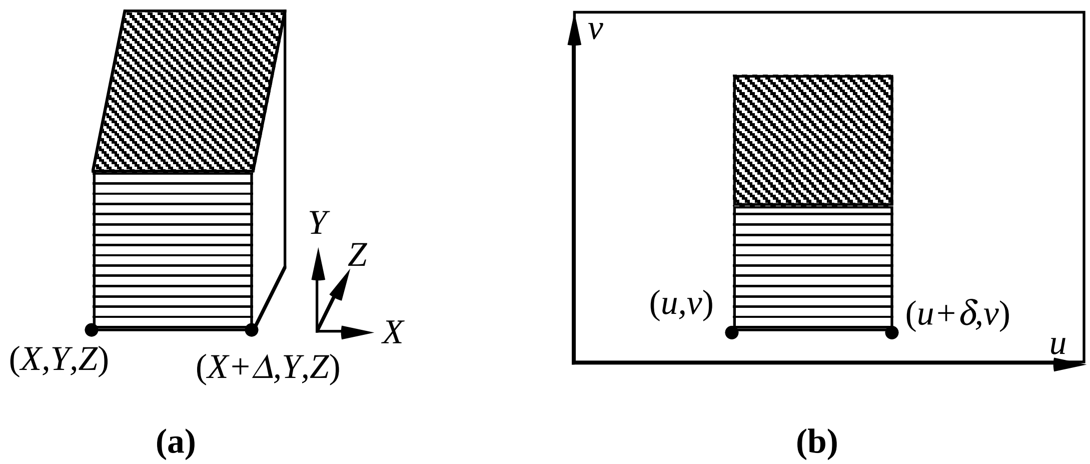

Figure 3: (a) Arbitrary object in world coordinates and (b) corresponding image view

Using the first two equations in (16) and the first equation in (17) and solving for (X, Y, Z) yields the desired world Cartesian location of the object. The solution is

$$\mathbf{w}_1 = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \Delta \dfrac{u}{\delta} \\ \dfrac{\Delta v\cos\theta - \lambda\delta\sin\theta + \lambda\Delta\sin\theta}{\delta} \\ \dfrac{\Delta v\sin\theta + \lambda\delta\cos\theta - \lambda\Delta\cos\theta}{\delta} \end{bmatrix} \tag{18}$$

The solution in (18) assumes that the object has two points which vary only in the X direction. A similar solution exists for two points which vary in the X, Y, and Z directions. Of course, (16) still holds for the object world location $\mathbf{w}_1=(X, Y, Z)$. Examine another point on the object which is located some fixed and known distance away in the X,Y and Z directions. The additional point would be located at $\mathbf{w}_3=(X+\Delta_x, Y+\Delta_y, Z+\Delta_z)$. The corresponding image location of $\mathbf{w}_3$ is $\mathbf{c}_3=(u+\delta_u, v+\delta_v)$. The image homogeneous location is found by application of (11). After conversion to the Cartesian coordinate system, the image Cartesian location is

$$\mathbf{c}_3 = \begin{bmatrix} u+\delta_u \\ v+\delta_v \\ w+\delta_w \end{bmatrix} = \begin{bmatrix} \lambda \dfrac{X+\Delta_x}{Y\sin\theta+\Delta_y\sin\theta-Z\cos\theta-\Delta_z\cos\theta+\lambda} \\ \lambda \dfrac{Y\cos\theta+\Delta_y\cos\theta+Z\sin\theta+\Delta_z\sin\theta}{Y\sin\theta+\Delta_y\sin\theta-Z\cos\theta-\Delta_z\cos\theta+\lambda} \\ \lambda \dfrac{Z\cos\theta+\Delta_z\cos\theta-Y\sin\theta-\Delta_z\sin\theta}{Y\sin\theta+\Delta_y\sin\theta-Z\cos\theta-\Delta_z\cos\theta+\lambda} \end{bmatrix} \tag{19}$$

Using the first equation of (16) and the first two equations of (19) and solving the system for the object's world Cartesian location, (X,Y,Z), gives

$$X = \frac{1}{\lambda\delta_u}\left(\Delta_z u^2\cos\theta-\Delta_y u^2\sin\theta-\delta_u\Delta_y u\sin\theta+\delta_u\Delta_z u\cos\theta+\lambda\Delta_x u\right), \tag{20}$$

$$Y = \frac{1}{\lambda\delta_u}\begin{pmatrix} \delta_v\Delta_z u\cos^2\theta-\Delta_y uv\sin\theta\cos\theta+\Delta_z uv\cos^2\theta- \\ \delta_v\Delta_y u\cos\theta\sin\theta+\lambda\delta_v\Delta_x\cos\theta-\lambda\Delta_y u-\lambda\delta_u\Delta_y+ \\ \lambda\Delta_y u\cos^2\theta+\lambda\Delta_z u\sin\theta\cos\theta+ \\ \lambda\Delta_x v\cos\theta-\lambda^2\delta_u\sin\theta+\lambda^2\Delta_x\sin\theta \end{pmatrix}, \tag{21}$$

and

$$Z = \frac{1}{\lambda\delta_u}\begin{pmatrix} \delta_v\Delta_y u\cos^2\theta+\Delta_z uv\sin\theta\cos\theta+\Delta_y uv\cos^2\theta+ \\ \delta_v\Delta_z u\cos\theta\sin\theta-\delta_v\Delta_y u+\lambda\delta_v\Delta_x\sin\theta-\lambda\Delta_z u\cos^2\theta+ \\ \lambda\Delta_y u\sin\theta\cos\theta+\lambda\Delta_x v\sin\theta-\lambda\delta_u\Delta_z+ \\ \lambda^2\delta_u\cos\theta-\lambda^2\Delta_x\cos\theta \end{pmatrix}. \tag{22}$$

An alternate solution may be found by using the second equation of (16) and the first two equations of (19) and solving for (X, Y, Z) yielding

$$X = \frac{1}{\lambda\delta_v}\begin{pmatrix} \Delta_z uv\cos\theta+\lambda\Delta_y u\cos\theta+\delta_u\Delta_z v\cos\theta+ \\ \lambda\delta_u\Delta_y\cos\theta-\Delta_y uv\sin\theta-\delta_u\Delta_y v\sin\theta- \\ \lambda\delta_v\Delta_x+\lambda\Delta_z u\sin\theta+\lambda\delta_u\Delta_z\sin\theta \end{pmatrix}, \tag{23}$$

$$Y = \frac{1}{\lambda\delta_v}\begin{pmatrix} \Delta_z v^2\cos^2\theta-\Delta_y v^2\sin\theta\cos\theta+2\lambda\Delta_y v\cos^2\theta+ \\ 2\lambda\Delta_z v\cos\theta\sin\theta-\delta_v\Delta_y v\cos\theta\sin\theta+\Delta_z\delta_v v\cos^2\theta- \\ \lambda^2\Delta_z\cos^2\theta-\lambda\delta_v\Delta_y+\lambda^2\Delta_z+\lambda\delta_v\Delta_y\cos^2\theta- \\ \lambda^2\delta_v\sin\theta-\lambda\Delta_y v+\lambda\delta_v\Delta_z\cos\theta\sin\theta+\lambda^2\Delta_y\cos\theta\sin\theta \end{pmatrix}, \tag{24}$$

and

$$Z = \frac{1}{\lambda\delta_v}\begin{pmatrix} \Delta_y v^2\cos^2\theta+\Delta_z v^2\sin\theta\cos\theta-\lambda^2\Delta_y\cos^2\theta+\lambda\Delta_z v+ \\ \delta_v\Delta_y v+\lambda^2\delta_v\cos\theta-\Delta_y v^2+2\lambda\Delta_y v\cos\theta\sin\theta- \\ 2\lambda\Delta_z v\cos^2\theta+\Delta_y\delta_v v\cos^2\theta+\lambda\delta_v\Delta_y\cos\theta\sin\theta+ \\ \delta_v\Delta_z v\cos\theta\sin\theta-\lambda\delta_v\Delta_z\cos^2\theta-\lambda^2\Delta_z\cos\theta\sin\theta \end{pmatrix}. \tag{25}$$

Use of (20)-(25) will allow a machine vision system to obtain the world location of an object with known geometric characteristics from a single camera view. The solution given in (20)-(22) has the term δ_u in the denominator which is not applicable if the optical geometry renders a zero image pixel differential in the *u* direction. However, the solution given in (23)-(25) may be used instead. Conversely, if δ_u=0, the solution in (20)-(22) may be used. If both solutions are defined, then the results may be averaged to obtain a more accurate range estimate.

4. IMPLEMENTATION

An machine vision system using off-the-shelf components was constructed. Scenes were captured from a black-and-white CCD camera using a high-speed frame grabber in a desktop computer which performed all subsequent image processing operations.

Because of the intense lighting in our application environment,the scene is initially processed with a localized nonlinear glare reduction algorithm [2]. After glare reduction, a Sobel edge detection algorithm is applied using ambient background levels to determine algorithm thresholds. After an area of interest is selected by the system operator, binary morphological operations are used to remove outliers and spurious artifacts of thresholding. When the object edge map is obtained, points with known geometric relationships are identified via a Hough transform [3]. Pairs of points on the object are used in (20)-(25) to obtain estimates of the object location with respect to the camera. The average of the estimates is taken. Typical results are accurate with less than 5% error depending on size of the object and distance from the camera to the object.

5. CONCLUSIONS

In this paper, a single camera range estimation algorithm is developed for a low-cost machine vision system that uses off-the-shelf components. This algorithm uses the perspective transformation and *a priori* knowledge of scene’s objects to compute the three-dimensional location of objects.

REFERENCES

[1] GONZALES, R. and WOOD, R., *Digital image processing*, Addison-Wesley, New York 1992.

[2] LIM, J., *Two-dimensional signal and image processing*, Prentice-Hall, Englewood Cliffs, NJ 1990.

[3] BALLARD, D. and BROWN, C., *Computer vision*, Prentice-Hall, Englewood Cliffs, NJ 1982.