農業工程學報　第41卷第4期
中華民國84年12月出版

Journal of Chinese Agricultural Engineering
Vol. 41. No. 4, December 1995

樹果自動搜尋與定位之研究

Automatic Search and Location of Fruit on the Tree

國立中興大學農業機械工程學系副教授
李芳繁
Fang-Fan Lee

國立中興大學農業機械工程學研究所碩士
施清德
Ching-Der Shy

摘　要

本研究使用三個步進馬達及減速齒輪組，製造由作者自行設計之攝影機定向機構。本定向機構在與立體機器視覺系統結合後，在實驗室內控制照度的環境下，進行樹果自動搜尋與3-D定位之研究。在樹果搜尋方面，使用了影像分割、侵蝕、膨脹及標記等影像處理方法。在影像分割時係使用紅色和色相影像。在3-D定位方面，使用了攝影機像距，水平轉軸、俯仰轉軸及聚焦轉軸之角度，以及攝影機的相對位置等參數。樣板比對法被用來找尋右影像與左影像間之對應點。樹果搜尋成功率約爲93％，而X、Y與Z軸之平均定位誤差均小於5％。定位一粒樹果約需時1分41秒。

關鍵詞：立體視覺，影像處理，自動搜尋，樹果。

ABSTRACT

A camera-pointing device was designed and fabricated using three stepping motors and reduction gears. This device was integrated with the stereo machine vision system and was employed to perform automatic tree fruit searching and 3-D locating tasks in the laboratory under controlled illumination. In fruit searching aspect, image segmentation, erosion, dilation, and labeling operations were used. The red and hue images were utilized for image segmentation. In 3-D locating aspect, parameters such as the image distances of the cameras, the angles of pan, tilt, and vergence, the relative positions of the cameras were needed. The template matching method was employed to find the corresponding points in the right and left images. The success rate of fruit searching was about 93 %. The average 3-D locating errors were all less than 5 % for the X, Y, and Z axes. It took approximately 1 minute and 41 seconds to locate a fruit.

Keywords : Stereo vision, Image processing, Automatic search, Tree fruit.

INTRODUCTION

Because the labor cost is rising and the labors available are decreasing, the labor-intensive agricultural industry has to be automated in order to make reasonable profits or even to survive. For this reason, agricultural automation is currently one of the most important agricultural policies in Taiwan. In the automatic harvesting of citrus fruit, the fruits have to be identified and located first before they can be picked by a robotic manipulator. Due to the rapid advance in machine vision and computer technologies, the machine vision can now be employed for fruits detection and 3-D locating at relatively fast speed. The robotic arm then will be guided by using the 3-D information to reach the target fruit.

The objectives of this research were: (a) to design and build a camera-pointing device; (b) to develop an automatic fruit searching algorithm; and (c) to develop an automatic 3-D fruit locating algorithm.

LITERATURE REVIEW

A tree fruit has to be detected and segmented in an image of the tree before calculating its 3-D position. Parrish and Goksel (1977) employed smoothed binary images of the apple tree scene to identify apple fruit and used the centroid of the fruit to guide a robotic arm for fruit-picking tests. Slaughter and Harrell (1987) identified oranges in natural orange grove scenes using color segmentation by thresholding color images at appropriate levels of hue and saturation. Chung and Lee (1993) also used the hue and saturation images to segment citrus fruit from the background in their tree fruit locating experiments. Slaughter and Harrell (1989) developed a classification model to discriminate oranges from the natural background of an orange grove using color information in a digital color image. In the AGROBOT project, Buemi et al. (1994) utilized hue and saturation to segment ripe and green tomatoes from the background. Molto et al. (1992) developed three algorithms for segmentation of citrus furit. These three algorithms were based on thresholding of the red image, the red/green images relation, and the curvature of the intensity image, respectively. The algorithm which was based on curvature could detect the ripe fruit as well as green fruit. Rabatel(1988) employed three CCD microcameras, which were fitted with 950 nm, 650 nm, and 550 nm filters respectively, to grab apple tree scene. The ratios of pixel values of these three images acquired at different wavelengths were used to classify pixels as either fruit pixels or background pixels.

In addition to color information, shape information can also be used to identify fruits. In order to enhance the contrast between fruit and background, Sites and Delwiche (1988) employed optical filters to acquire images of peach and apple trees. To identify fruit within the images obtained, these images were global thresholded, smoothed and segmented. Perimeter, area, and compactness were used to distinguish multiple fruits, single fruit, and noise. Whittaker et al. (1987) used circular Hough transform to locate tomatoes based on shape and not on the color difference between the tomato and plant foliage.

Binocular stereo vision, based on the principle of parallax, is a common method to compute the 3-D position of an object. Baylou et al. (1984) employed stereo vision to locate asparagus for automatic picking using a robot. Alden et al. (1988) developed a stereo imaging system capable of determining the size, shape, and location of a given plant which could be used to control a robotic trimming device. Jia and Krutz (1992) used machine vision to find the coordinates of the maize plant by using parallax equations. Unlike the industrial environment, the grove or field scene is more complex and the correspondence problem is more difficult to solve. To solve this problem, Buemi et al. (1994) divided the right and left images of tomato field in horizontal strips containing the possible corresponding points and then matched the regions according to their order into the strips. Chung and Lee (1993) utilized the projected laser light point on the target fruit as the corresponding points in the location of citrus furits.

Approaches other than stereo vision were also used

in range determination.Weckler and Kranzler (1989) utilized image focus of a calibrated camera-lens system to calculate the range values from the video camera to target objects using only one camera. Optimum focus of the image was determined by maximizing the high-frequency content of the Fourier transform of the image. Harrell et al. (1990) employed a sensing package consisting of an ultrasonic ranging transducer and a color CCD video camera to locate the targeted fruit with respect to the end-effector.

EQUIPMENT

1. Vision System

Two CCD RGB cameras (XC-711, SONY) were used as the image input devices.They can output both the NTSC and RGB video signals. Each camera is equipped with a 2/3" interline transfer CCD with 768(H)x493(V) pick-up elements. Two identical lenses (B2514D, COSMICAR) with 25 mm focal length were employed. The video signal outputted from the camera was inputted to the color frame grabber (CFG, IMAGING TECHNOLOGY INC.) for digitization. This frame grabber has 1024x512x28 bits on-board image memory and is capable of storing two true color images. It was also built with 3 input look-up tables and 32 output look-up tables. Acolor video monitor (PVM-1342Q, SONY) was used for image display. Two 60W incandescent light bulbs were placed at both sides in front of the sample fruit tree (a bonsai tree) to provide proper illumination. The host microcomputer was an ECS 80486-DX50 IBM compatible PC. Microsoft C7.0 compiler was used.

2. Camera-Pointing Device (CPD)

The camera-pointing device was designed and built in the laboratory. It consists of three stepping motors with 6 kg-cm, 12 kg-cm, and 22 kg-cm magnetic torques, respectively. One revolution is divided into 800 steps for each motor. The accuracies are the same for these three motors and are equal to 0.45 degrees. The CPD has three degrees of freedom: pan, tilt, and vergence. The pan and tilt angles were the same for both cameras. While vergence was achieved only by the rotation of the left camera. The ranges of pan is $0° \sim 233°$, of tilt is $-38° \sim 53°$, and of vergence is $-20° \sim 49°$. Because the pan, tilt, and vergence shafts have loads of different magnitude, different ratios of the reduction gears were employed. This, in turn, made the angle resolution values distinct for each shaft. The ratios of the reduction gears are 9, 6, and 4 for the pan, tilt, and vergence shafts, respectively. The position of the right camera can be adjusted in X, Y, and Z directions.The distance between the two cameras can be changed by moving the left camera along a slot in X direction. A pulse command motion control board was used to control the rotation of the motors. The three motors can rotate simultaneously. Table 1 lists the characteristics of the CPD. Figure 1 shows the picture of the CPD with cameras mounted on it. Figure 2 is the block diagram of electric circuits of the equipment of this study And Figure 3 shows the experimental setup.

Table 1. The CPD characteristics

	Pan(Mt_1)	Tilt(Mt_2)	Vergence(Mt_3)
Searching range	$0° \sim +233°$	$-38° \sim +53°$	$-20° \sim +49°$
Angle resolution	$0.0510°$/step	$0.0766°$/step	$0.1123°$/step
Maximum angular velocity	1.164 rad/s	1.746 rad/s	2.618 rad/s
Acceleration	1.455 rad/s^2	2.18 rad/s^2	3.275 rad/s^2
Weight	40.5 kg		
Dimensions	450mm(L)×450mm(W)×660mm(H)		

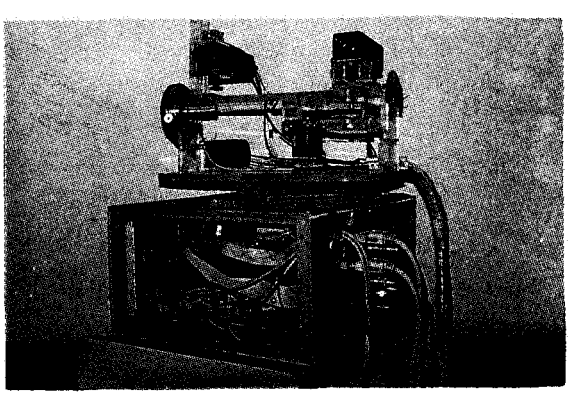

Fig. 1. The camera-pointing device

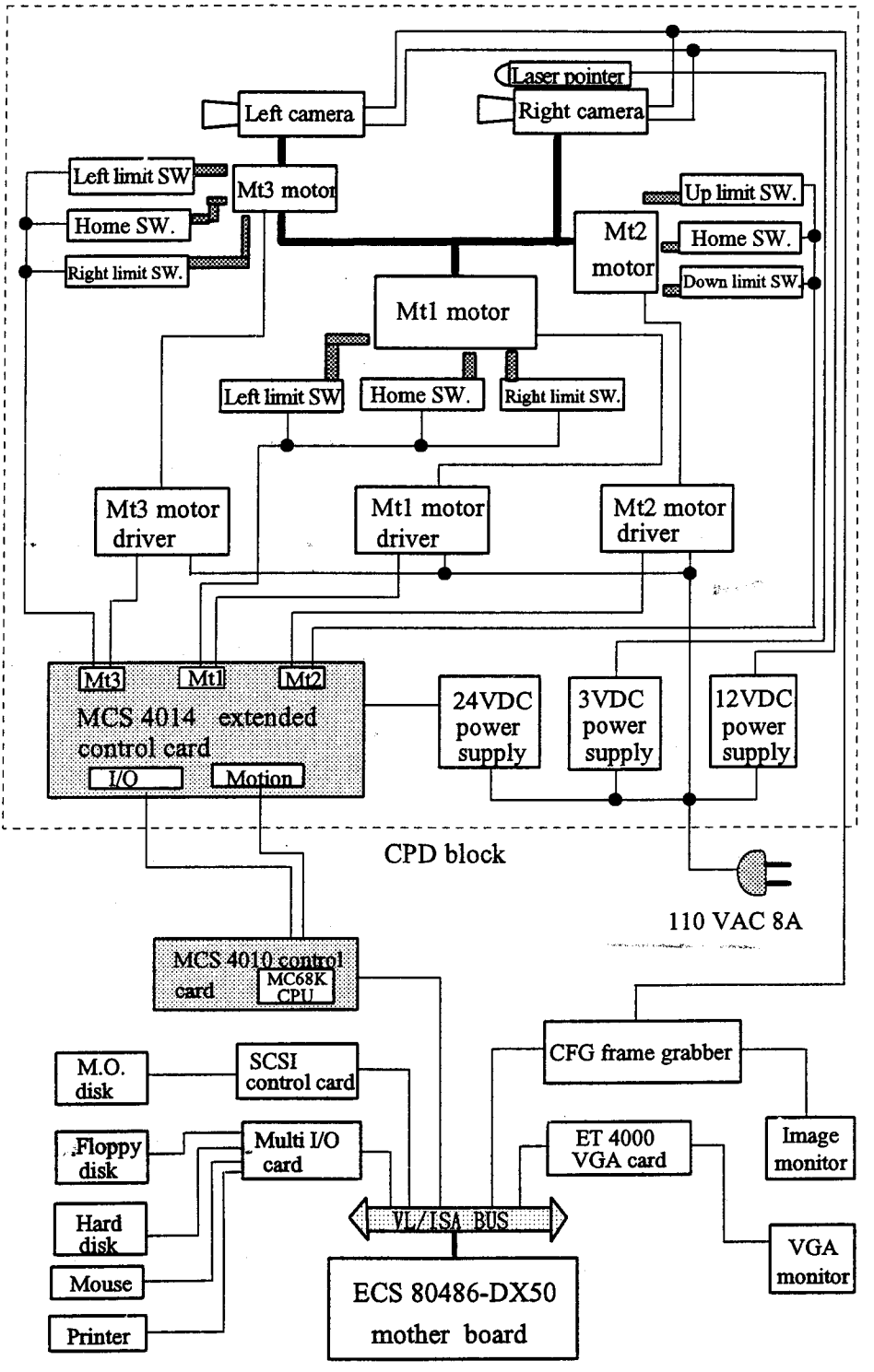

Fig. 2. Block diagram of electric circuits

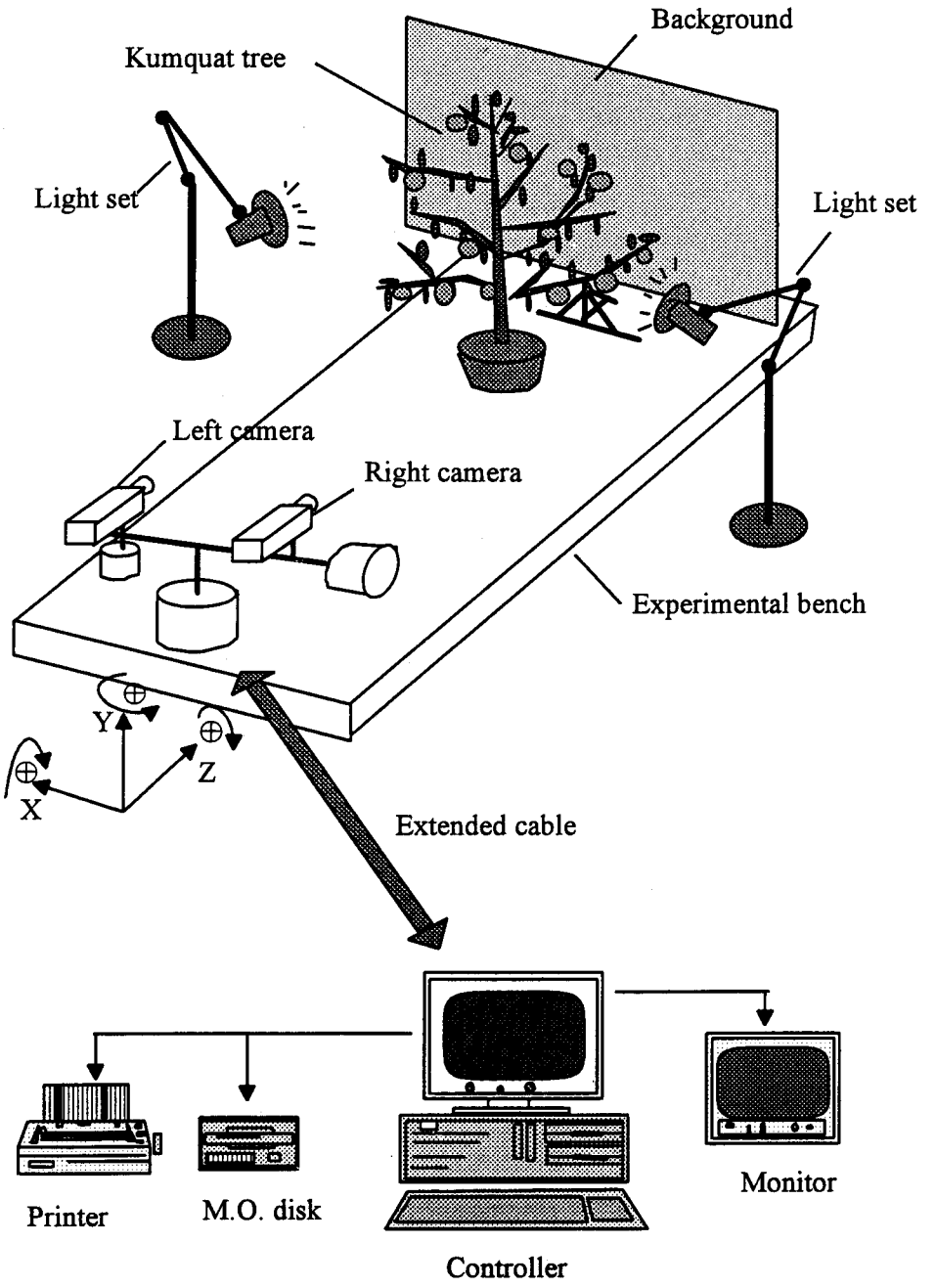

Fig. 3. The experimental setup

DERIVATION OF 3-D LOCATING EQUATIONS

Figure 4 is a common stereo imaging model, the parallel model (Gonzalez and Woods, 1992). In this model, two identical cameras are used and the coordinate systems of both cameras are aligned. Also, the two optical axes are parallel. The distance between the two lens centers is called the baseline.The world point P (X, Y, Z) has image points $P_1(x_1,y_1)$ and $P_2(x_2,y_2)$ on the image planes as shown. By using similar triangles relationship, the following equations:

$$X_1 = \frac{x_1}{\lambda}(\lambda - Z_1) \quad (1)$$

$$Y_1 = \frac{y_1}{\lambda}(\lambda - Z_1) \quad (2)$$

$$X_2 = \frac{x_2}{\lambda}(\lambda - Z_2) \quad (3)$$

$$Y_2 = \frac{y_2}{\lambda}(\lambda - Z_2) \quad (4)$$

can be obtained; where λ is the image distance, and the subscripts indicate the camera which was moved to the origin of the world coordinate system.

The camera model used in this study (Fig. 5) is the general stereo model(Alvertos et al., 1989) since during the fruit searching process both camera coordinate systems are not aligned. Before deriving the 3-D locating equations, the symbols that will be used are defined below:

I_1 : the image plane center of the right camera,

I_2 : the image plane center of the left camera,

C_0 : the intersection of the center line of pan shaft and the upper surface of the camera-supporting plate,

C_1 : a point located on the center line of the tilt shaft with zero X_1 value,

C_2 : the intersection of the center line of vergence shaft and the center line of tilt shaft,

(X_w,Y_w,Z_w) : the world coordinate system; its origin 0_w is located at a corner of the CPD top cover,

0(X, Y, Z); the 3-D position of an object with respect to the world coordinate system (X_w,Y_w,Z_w),

(X_1,Y_1,Z_1) : the coordinate system of the right camera,

(X_2,Y_2,Z_2) : the coordinate system of the left camera,

T_1 : the position vector of point I_1 with respect to point C_1 and its components are (t_{1x},t_{1y},t_{1z}),

T_2 : the position vector of point I_2 with respect to point C_2 and its components are (t_{2x},t_{2y},t_{2z}),

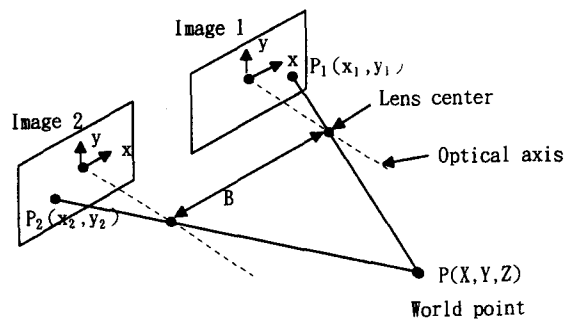

Fig. 4. The parallel model

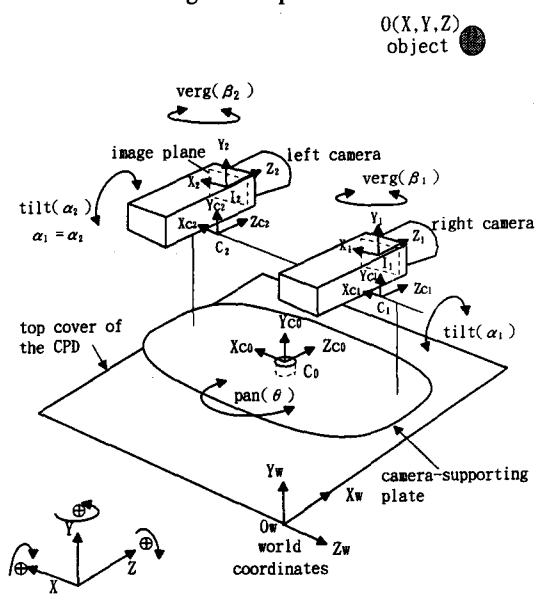

Fig. 5. The camera model of this study

T_3 : the position vector of point C_2 with respect to point C_1 and its components are (t_{3x}, t_{3y}, t_{3z}),

T_4 : the position vector of point C_1 with respect to point C_0 and its components are (t_{4x}, t_{4y}, t_{4z}),

T_5 : the position vector of point C_0 with respect to point 0_w and its components are (t_{5x}, t_{5y}, t_{5z}),

pan (θ): the rotation matrix of the pan shaft,

tilt (α_1) : the rotation matrix of the tilt shaft for the right camera,

tilt(α_2): the rotation matrix of the tilt shaft for the left camera; because there is only one tilt shaft, α_1 is equal to α_2 ,

verg (β_1) : Since there is no vergence motor, β_1 is equal to zero,

verg (β_2) : the rotation matrix of the vergence shaft for the left camera.

To obtain the 3-D position (X, Y, Z) of the object with respect to the world coordinate system, the following 3 steps have to be computed.

Step 1: $I_2(X_2, Y_2, Z_2) \xrightarrow{T_2^{-1} \times tilt^{-1}(\alpha_2) \times verg^{-1}(\beta_2)} C_2(X_{c2}, Y_{c2}, Z_{c2}) \xrightarrow{T_3^{-1}}$

$C_1(X_{c1}, Y_{c1}, Z_{c1}) \xrightarrow{verg(\beta_1) \times tilt(\alpha_1) \times T_1} I_1(X_1, Y_1, Z_1)$

Step 2: $I_1 \xrightarrow{T_1^{-1} \times tilt^{-1}(\alpha_1) \times verg^{-1}(\beta_1)} C_1(X_{c1}, Y_{c1}, Z_{c1})$

Step 3: $C_1(X_{c1}, Y_{c1}, Z_{c1}) \xrightarrow{T_4^{-1}} C_0(X_{co}, Y_{co}, Z_{co}) \xrightarrow{pan^{-1}(\theta) \times T_5^{-1}} O(X, Y, Z)$

The operations of Step 1 can be expressed as

$$\begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \\ 1 \end{bmatrix} = A \begin{bmatrix} X_2 \\ Y_2 \\ Z_2 \\ 1 \end{bmatrix} \tag{5}$$

where $A = T_1 \times tilt(\alpha_1) \times verg(\beta_1) \times T_3^{-1} \times verg^{-1}(\beta_2) \times tilt^{-1}(\alpha_2) \times T_2^{-1}$, and

$$T_1 = \begin{bmatrix} 1 & 0 & 0 & -t_{1x} \\ 0 & 1 & 0 & -t_{1y} \\ 0 & 0 & 1 & -t_{1z} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{6}$$

$$tilt(\alpha_1) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\alpha_1 & \sin\alpha_1 & 0 \\ 0 & -\sin\alpha_1 & \cos\alpha_1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{7}$$

$$verg(\beta_1) = \begin{bmatrix} \cos\beta_1 & 0 & -\sin\beta_1 & 0 \\ 0 & 1 & 0 & 0 \\ \sin\beta_1 & 0 & \cos\beta_1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{8}$$

$$T_3 = \begin{bmatrix} 1 & 0 & 0 & -t_{3x} \\ 0 & 1 & 0 & -t_{3y} \\ 0 & 0 & 1 & -t_{3z} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{9}$$

$$verg(\beta_2) = \begin{bmatrix} \cos\beta_2 & 0 & -\sin\beta_2 & 0 \\ 0 & 1 & 0 & 0 \\ \sin\beta_2 & 0 & \cos\beta_2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{10}$$

$$tilt(\alpha_2) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\alpha_2 & \sin\alpha_2 & 0 \\ 0 & -\sin\alpha_2 & \cos\alpha_2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{11}$$

$$T_2 = \begin{bmatrix} 1 & 0 & 0 & -t_{2x} \\ 0 & 1 & 0 & -t_{2y} \\ 0 & 0 & 1 & -t_{2z} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{12}$$

Multiplication of the above seven transforming matrices yields the 4 × 4 matrix A. The elements of matrix A are as follows:

$$A_{11} = \cos(\beta_1 - \beta_2) \tag{13-1}$$

$$A_{12} = \sin\alpha_2 \sin(\beta_2 - \beta_1) \tag{13-2}$$

$$A_{13} = \cos\alpha_2 \sin(\beta_2 - \beta_1) \tag{13-3}$$

$$A_{14} = t_{2x}\cos(\beta_1 - \beta_2) + \sin(\beta_2 - \beta_1)(t_{2y}\sin\alpha_2 + t_{2z}\cos\alpha_2) + t_{3x}\cos\beta_1 - t_{3z}\sin\beta_1 - t_{1x} \tag{13-4}$$

$$A_{21} = \sin\alpha_1 \sin(\beta_1 - \beta_2) \tag{13-5}$$

$$A_{22} = \sin\alpha_1 \sin\alpha_2 \cos(\beta_1 - \beta_2) + \cos\alpha_2 \cos\alpha_1 \tag{13-6}$$

$$A_{23} = \cos\alpha_2 \sin\alpha_1 \cos(\beta_1 - \beta_2) - \sin\alpha_2 \cos\alpha_1 \tag{13-7}$$

$$A_{24} = \sin\alpha_1(t_{2x}\sin(\beta_1 - \beta_2) + t_{3x}\sin\beta_1 + t_{3z}\cos\beta_1) + (t_{2y}\sin\alpha_2 + t_{2z}\cos\alpha_2)\sin\alpha_1\cos(\beta_2 - \beta_1) + \cos\alpha_1(t_{2y}\cos\alpha_2 - t_{2z}\sin\alpha_2 + t_{3y}) - t_{1y} \tag{13-8}$$

$$A_{31} = \cos\alpha_1 \sin(\beta_1 - \beta_2) \tag{13-9}$$

$$A_{32} = \sin\alpha_2 \cos\alpha_1 \cos(\beta_1 - \beta_2) - \cos\alpha_2 \sin\alpha_1 \tag{13-10}$$

$$A_{33}=\cos\alpha_2\cos\alpha_1\cos(\beta_1-\beta_2)+\sin\alpha_2\sin\alpha_1 \quad (13\text{-}11)$$

$$A_{34}=\cos\alpha_1(t_{2x}\sin(\beta_1-\beta_2)+t_{3x}\sin\beta_1+t_{3z}\cos\beta_1)+(t_{2z}\sin\alpha_2-t_{2y}\cos\alpha_2-t_{3y})\sin\alpha_1+\cos(\beta_2-\beta_1)(t_{2y}\cos\alpha_1\sin\alpha_2+t_{2z}\cos\alpha_1\cos\alpha_2)-t_{1z} \quad (13\text{-}12)$$

$$A_{41}=0 \quad (13\text{-}13)$$

$$A_{42}=0 \quad (13\text{-}14)$$

$$A_{43}=0 \quad (13\text{-}15)$$

$$A_{44}=1 \quad (13\text{-}16)$$

Substituting Eqs. (13-1) ~ (13-16) into Eq. (5), the following three simultaneous equations are obtained:

$$X_1=A_{11}X_2+A_{12}Y_2+A_{13}Z_2+A_{14} \quad (14)$$

$$Y_1=A_{21}X_2+A_{22}Y_2+A_{23}Z_2+A_{24} \quad (15)$$

$$Z_1=A_{31}X_2+A_{32}Y_2+A_{33}Z_2+A_{34} \quad (16)$$

Using Eqs. (1) ~ (4) and Eqs. (14) ~ (16), Z_1 can be solved as:

$$Z_1=\frac{M_1M_3-M_2M_4}{M_1\lambda+x_1M_4} \quad (17)$$

where $M_1=-A_{11}x_2-A_{12}y_2+A_{13}\lambda$

$M_2=(A_{11}x_2+A_{12}y_2+A_{14}-x_1)\lambda$

$M_3=(A_{31}x_2+A_{32}y_2+A_{34})\lambda$

$M_4=-A_{31}x_2-A_{32}y_2+A_{33}\lambda$

λ: the image distance,

x_1 : the x coordinate of the object image point in the right image plane,

x_2 : the x coordinate of the object image point in the left image plane,

y_2 : the y coordinate of the object image point in the left image plane.

Solving Eqs. (17), (1), and (2), the values of X_1, Y_1 , and Z_1 can be determined.

Step 2 can be written in the form:

$$\begin{bmatrix} X_{c1} \\ Y_{c1} \\ Z_{c1} \\ 1 \end{bmatrix} = D \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \\ 1 \end{bmatrix} \quad (18)$$

where $D=verg^{-1}(\beta_1)\times tilt^{-1}(\alpha_1)\times C_1^{-1}$ .The elements of matrix D are listed below:

$$D_{11}=\cos\beta_1 \quad (19\text{-}1)$$

$$D_{12}=\sin\alpha_1\sin\beta_1 \quad (19\text{-}2)$$

$$D_{13}=\sin\beta_1\cos\alpha_1 \quad (19\text{-}3)$$

$$D_{14}=t_{1x}\cos\beta_1+\sin\beta_1(t_{1y}\sin\alpha_1+t_{1z}\cos\alpha_1) \quad (19\text{-}4)$$

$$D_{21}=0 \quad (19\text{-}5)$$

$$D_{22}=\cos\alpha_1 \quad (19\text{-}6)$$

$$D_{23}=-\sin\alpha_1 \quad (19\text{-}7)$$

$$D_{24}=t_{1y}\cos\alpha_1-t_{1z}\sin\alpha_1 \quad (19\text{-}8)$$

$$D_{31}=-\sin\beta_1 \quad (19\text{-}9)$$

$$D_{32}=\sin\alpha_1\cos\beta_1 \quad (19\text{-}10)$$

$$D_{33}=\cos\alpha_1\cos\beta_1 \quad (19\text{-}11)$$

$$D_{34}=-t_{1x}\sin\beta_1+\cos\beta_1(t_{1y}\sin\alpha_1+t_{1z}\cos\alpha_1) \quad (19\text{-}12)$$

$$D_{41}=0 \quad (19\text{-}13)$$

$$D_{42}=0 \quad (19\text{-}14)$$

$$D_{43}=0 \quad (19\text{-}15)$$

$$D_{44}=1 \quad (19\text{-}16)$$

Substituting Eqs. (19-1) ~ (19-16) into Eq. (18) and expanding the matrix product yidlds

$$X_{c1}=D_{11}X_1+D_{12}Y_1+D_{13}Z_1+D_{14} \quad (20)$$

$$Y_{c1}=D_{21}X_1+D_{22}Y_1+D_{23}Z_1+D_{24} \quad (21)$$

$$Z_{c1}=D_{31}X_1+D_{32}Y_1+D_{33}Z_1+D_{34} \quad (22)$$

As before, Step 3 is expressed as

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = W \begin{bmatrix} X_{c1} \\ Y_{c1} \\ Z_{c1} \\ 1 \end{bmatrix} \quad (23)$$

where $W=T_5^{-1}\times pan^{-1}(\theta)\times T_4^{-1}$, and

$$T_5^{-1}=\begin{bmatrix} 1 & 0 & 0 & -t_{5x} \\ 0 & 1 & 0 & -t_{5y} \\ 0 & 0 & 1 & -t_{5z} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (24)$$

$$pan^{-1}(\theta)=\begin{bmatrix} \cos(-\theta) & 0 & -\sin(-\theta) & 0 \\ 0 & 1 & 0 & 0 \\ \sin(-\theta) & 0 & \cos(-\theta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (25)$$

$$T_4^{-1}=\begin{bmatrix} 1 & 0 & 0 & -C_{4x} \\ 0 & 1 & 0 & -C_{4y} \\ 0 & 0 & 1 & -C_{4z} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (26)$$

The product of the above transforming matrices yields

$$W_{11}=\cos\theta \quad (27\text{-}1)$$

$$W_{12}=0 \quad (27\text{-}2)$$

$$W_{13}=\sin\theta \quad (27\text{-}3)$$

$$W_{14}=t_{4x}\cos\theta+t_{4z}\sin\theta+t_{5x} \quad (27\text{-}4)$$

$$W_{21}=0 \quad (27\text{-}5)$$

$$W_{22}=1 \quad (27\text{-}6)$$

$$W_{23}=0 \quad (27\text{-}7)$$

$$W_{24}=T_{4y}+T_{5y} \quad (27\text{-}8)$$

$$W_{31}=-\sin\theta \quad (27\text{-}9)$$

$$W_{32}=0 \quad (27\text{-}10)$$

$$W_{33}=\cos\theta \quad (27\text{-}11)$$

$$W_{34}=-t_{4x}\sin\theta+t_{4z}\cos\theta+t_{5z} \quad (27\text{-}12)$$

$$W_{41}=0 \quad (27\text{-}13)$$

$$W_{42}=0 \quad (27\text{-}14)$$

$$W_{43}=0 \quad (27\text{-}15)$$

$$W_{44}=1 \quad (27\text{-}16)$$

Substituting Eqs. (27-1) ~ (27-16) into Eq. (23) and expanding the matrix product yields

$$X=W_{11}X_{C1}+W_{12}Y_{C1}+W_{13}Z_{C1}+W_{14} \quad (28)$$

$$Y=W_{21}X_{C1}+W_{22}Y_{C1}+W_{23}Z_{C1}+W_{24} \quad (29)$$

$$Z=W_{31}X_{C1}+W_{32}Y_{C1}+W_{33}Z_{C1}+W_{34} \quad (30)$$

where X, Y, and Z are the 3-D positions of the object relative to the world coordinate system.

EXPERIMENTAL METHODS AND PROCEDURES

1. Camera Calibration

In order to determine the three-dimensional position of the target fruit, the position of the image plane, the image distance, the angles of pan, tilt, and vergence, the positions of the cameras with respect to both the rotating centers and the world coordinate system are needed. The position of the image plane of camera is available from the user's manual. Because in this study both the target fruits range and the cameras employed were the same as those used by Chung and Lee (1993), the image distance value utilized by Chung and Lee was used. The angles of pan, tilt, and vergence were calculated by the computer using the angle resolutions. While the relative position vectors are constant and are given in Table 2.

Table 2. The X, Y, and Z components of position vectors

Position vectors	X components	Y components	Z components
T_1	0	103.3	12.4
T_2	0	103.3	12.4
T_3	297	0	0
T_4	-105.4	127.1	13.5
T_5	226.3	58.7	-223.5

unit: mm

2. Angle Resolution Calibration

The angle resolutions of the pan, tilt, and vergence motors have to be calibrated before starting the 3-D locating tests. It is not possible to obtain accurate 3-D position if the resolution angles used are not correct. Figure 6 shows the schematic diagram of angle

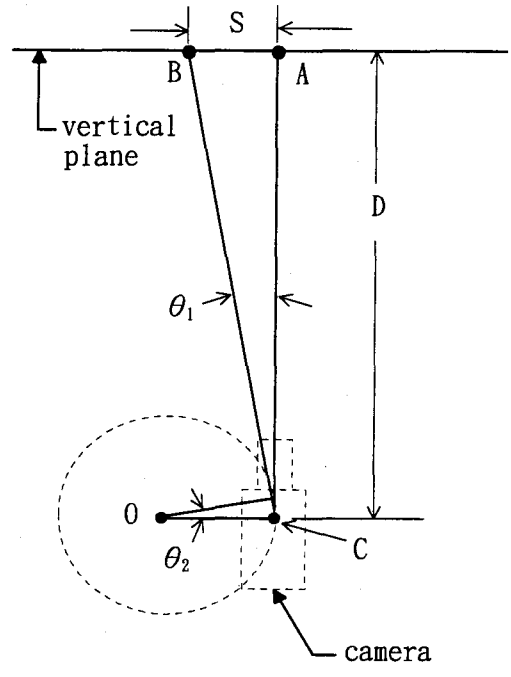

Fig. 6. Schematic diagram for angle resolution calibration

resolution calibration for the pan motor. Point 0 is the center of the shaft of camera-supporting plate, CA is the laser ray, and point C is the intersection of the laser ray and a line passing through point 0 and perpendicular to line CA. Rotating the pan motor for 10 steps, the laser light point projected on the vertical plane will move from point A to point B. The distance D used is equal to

2245mm. Sinceθ_1 is equal to θ_2, the angle resolution is equal to S/10D radians. The angle resolution calibration methods for the tilt and the vergence motors are the same as that for the pan motor.

3. Planning of Fruit-Searching Path

The fruit-searching path is shown in Figure 7. Searching started from the top right corner relative to the cameras of the whole searching area and followed the path shown in the figure. The tilt searching angle is 12.45 degrees. Tilt searching angle is the angle that the tilt shaft should rotate after completing a horizontal searching path. The searching ranges are $0° \sim 233°$ for pan, $-38° \sim 53°$ for tilt.

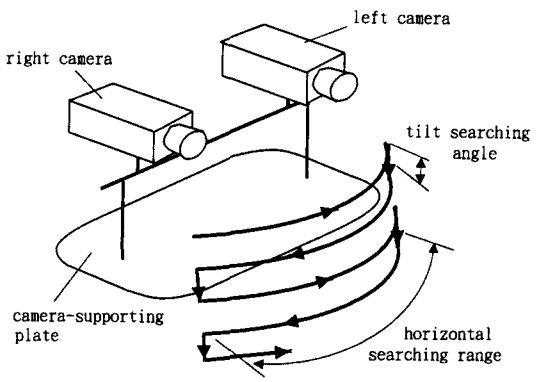

Fig. 7. The fruit-searching path

4. Fruit Detection with the Right Camera

(1) Image Segmentation

Each RGB image grabbed by the right camera during the searching time was transformed into an HSI image. Both the red and hue images were thresholded to segment out the ripe fruits from the background. If only the red image was thresholded, it was difficult to differentiate the ripe fruits from the unripe ones because both of them have similar red intensity. Conversely, if only the hue image was thresholded, too much noise was produced. The threshold used for segmenting the red image was 180. Any pixel whose gray-level value being less than 180 was classified as background pixel and was set to 0; all other pixels were classified as ripe fruit pixels and were set to 255. Double thresholds, 35 and 10, were employed to segment the hue image. In this case, any pixel with gray-level value between 35 and 10 was classified as pixel of ripe fruits and was set to 255, all other pixels were classified as background pixel and was set to 0.

(2) Logic "AND" Operation

The segmented red and hue images were then processed using the logic "AND" operation. The unripe fruits and some noise would disappear from the resulting image of this logic operation.

(3) Erosion and Dilation

Erosion has the effect of shrinking an image and dilation has the effect of expanding an image (Giardina and Dougherty, 1988). Thus, they can be used to remove image noise. The resulting image of logic operation was eroded and dilated again to remove the remaining noise. In the erosion operation, if anyone of the 9 pixels of a 3x3 subarea of an image was zero, then the center pixel of this subarea was set to zero, otherwise the center pixel was not changed. For the dilation operation, if anyone of these 9 pixels was 255, then the center pixel was set to 255, otherwise it was not changed. The 3x3 subarea was moving on the image, just like a mask moving on an image, from left to right and from top to bottom. In this study two consecutive erosion operations were executed first followed by two consecutive dilation operations on the resulting image of (2). This morphological operations were implemented by using the spatial convolution method.

(4) Labeling of Fruit Images

In order to identify the individual fruit in an image, the labeling algorithm was used. Four-connectivity was used in performing this algorithm.This algorithm is as follows: The resulting image of (3) is scanned from left to right and from top to bottom. Let p denote the pixel at any step in the scanning process. If the value of p is 0, move on to the next pixel, If the value of p is not 0, check the values of its upper and left-hand neighbors. If both neighbors are all zero, assign a new label (i. e. new gray level value) to p. If only one of the two neighbors is not zero, assign its value to p. If both neighbors have

same value other than zero, assign this value to p. If both neighbors have different non-zero values, assign anyone of these two value to p and record that these two non-zero values are equivalent. After completing this scanning process, set pixels having equivalent values to a single value. Thus, the individual fruit is identified by different gray level values and all pixels belonging to each individual fruit are 4-connected. The connected region having the largest area is defined as the target fruit and is used in 3-D locating tests. Figure 8 is the flow chart of target fruit identification.

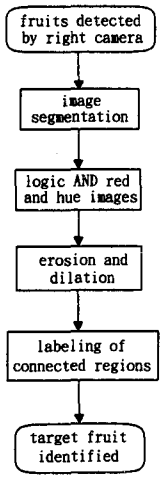

Fig. 8. Flow chart of target fruit identification

5. Fruit Centroid Determination

The centroid (x, y) of an image region R is defined as

$$x = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{31}$$

and

$$y = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{32}$$

where N is the total number of pixels of region R. The centroid of the target fruit region is determined and used in template matching.

6. Vergence of Cameras

For a binocular stereo vision system to be able to determine the 3-D position of an object placed in 3-D space, the vergence of both cameras is necessary. In other words, the target object has to be within the fields of view of both cameras.

The epipolar line constraint and the template matching were used to check camera vergence. An epipolar line on one stereo image corresponding to a given point in another stereo image is the perspective projection on the first stereo image of the three-dimensional ray that is the inverse perspective projection of the given point from the other stereo image. The epipolar line constraint is the strongest constraint in image matching (Haralick and Shapiro, 1993). In this study, a rectangular subimage, 90 pixels wide and 4 pixels high, centered at the centroid of the target fruit detected by the right camera was used as the template for template matching.

After the right camera had identified the target fruit, it would temporarily stop its fruit-searching task; i. e., its position was fixed. At the same time, the left camera started to search the targeted fruit using template matching.

The correlation measure between the left image and the template is used to evaluate template matching and is defined as

$$CR(s) = \sum_x \sum_y [f(x,y) - tem(x-s,y)]^2 \tag{33}$$

where s=0, 1, 2, ⋯, 511, f(x, y) is the left image, and tem (x, y) is the template. In words, the squared differences between the template pixels and their corresponding pixels of the left image were summed. For an exact match, CR(s) is equal to zero. In the template matching process, the template was moving along the epipolar line on the left image from left to right. If no CR (s) value less than the predefined threshold was obtained for an image, the left camera would rotate 6.75 ° clockwise to acquire a new image and the matching task was repeated. If more than one CR(s) values were less than the threshold, the one having the smallest value had the best match and the corresponding points of the right and left images were found. Finally, the 3-D position of the target fruit could be calculated using the equations derived earlier. Figure 9 shows the flow chart for 3-D position determination.

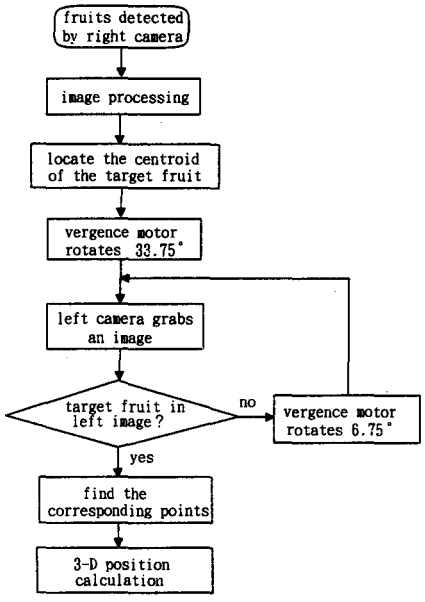

Fig. 9. Flow chart of 3-D position determination

Table 3. Calibrated and calculated angle resolutions

	pan(Mt_1)	tilt(Mt_2)	vergence(Mt_3)
S*	20mm	30mm	44mm
D*	2245mm	2245mm	2245mm
Calibrated value	0.0510°	0.0766°	0.1123°
Calculated value	0.0500°	0.0750°	0.1125°
Percent difference	1.96%	2.09%	0.18%

* For S and D, see Fig. 6

RESULTS AND DISCUSSION

The calibration results of the angle resolutions of the pan, tilt, and vergence shafts are shown in Table 3. The angle resolution of the pan shaft is the angle that the pan shaft rotates when the pan motor rotates one step. The same definition applies to the tilt and vergence shafts. The calculated angle resolutions are also listed in Table 3. These angles are equal to the step angles of the pan, tilt, and vergence motors divided by the reduction gear ratio of each motor. The percent differences between the calibrated and calculated angle resolutions are listed in Table 3 as well. The percent differences were computed based on the calibrated angle resolutions. The calibrated angle resolutions were employed in 3-D locating tests because more accurate 3-D positions could be obtained than using the calculated angle resolutions.

In order to find the target fruit, the image processing operations shown in Figure 8 were performed. Figure 11 is the segmented red image of the image shown in figure 10 using a threshold of 180. And Figure 12 is the segmented hue image of the same image using two threshold values, 35 and 10. The right one of the four overlapped fruits in Figure 10 was unripe and had yellowgreen peel. Although it appears in Figure 11, however, it does not show in Figure 12. After the logic " AND" operation on Figures 11 and 12, the unripe fruit does not appear in the resulting image (Figure 13) and most noises were removed too. It is noted that the left fruit becomes a small region; since it was located in the shadow and, therefore, had low gray level values. Attempting to remove the small noise existing in the resulting image after the "AND" operation, the morphological operations, erosion and dilation, were performed and Figure 14 is the resulting image. As can be seen in Figure 14, there are seven connected regions in the image. In order to choose the largest one for 3-D positioning, the 4-connected labeling algorithm was implemented to label each connected region using different intensities (Figure 15). Figure 16 shows the largest region which was kept for 3-D position determination. Figure 17 is an image grabbed by the right camera with the matching template indicated on the detected target fruit. While Figure 18 is an image grabbed by the left camera having the best match region marked on the same target fruit of Figure 17. The success rate of automatic fruit searching was about 93 % because some indoor scenes were misjudged as fruits. For practical fruit-harvesting operations, since the illumination in the citrus grove is not constant, using fixed threshold values will not be able to segment fruits correctly from the background all the time. This problem might be solved by using variable threshold values which are determined by the statistical method or using cameras equipped with auto-iris lens.

Fig. 10. Original image grabbed by the right camera

Fig. 13. Resulting image after logic AND Figs. 11 and 12

Fig. 11. The segmented red image of Fig. 10

Fig. 14. Resulting image of Fig. 13 after erosion and dilation

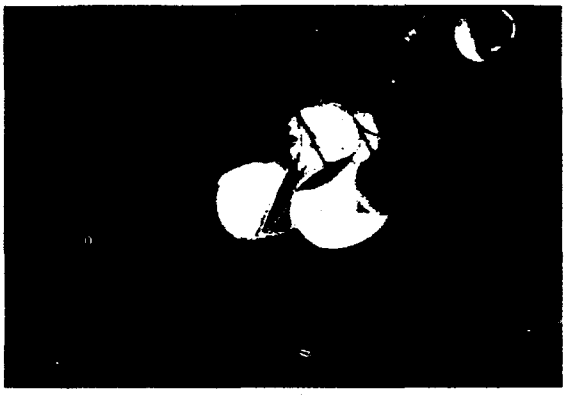

Fig. 12. The segmented hue image of Fig. 10

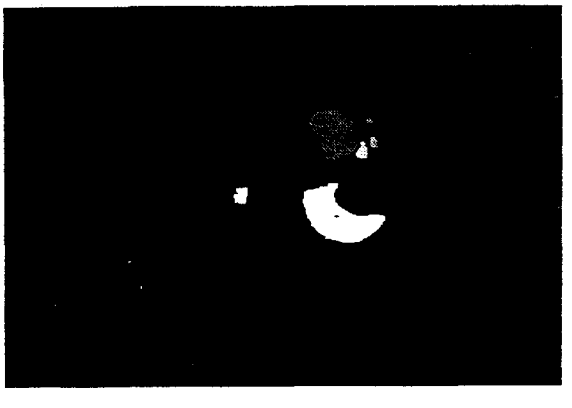

Fig. 15. The labeled image of Fig. 14

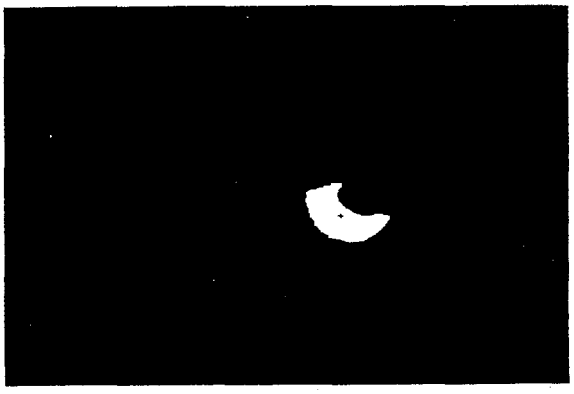

Fig. 16. The target fruit of Fig. 15

Fig. 17. A right image with the template marked on the target fruit

Fig. 18. The left image with the best-match area marked on the same target fruit of Fig. 17

In the 3-D locating test, fruits were placed at approximately 727mm, 869mm, 1012mm, and 1145mm from the origin of the world coordinate system in the Z-axis direction. At each distance, fruits were placed at four different locations. At each location, the 3-D locating test was repeated for five times. Table 4 through Table 7 list the testing results; while Figure 19 shows the percent errors between the measured values and the calculated values. The above tables show that the maximum and minimum errors in the X direction are 11. 74 % and 0.47 % , respectively; in the Y direction are 2.

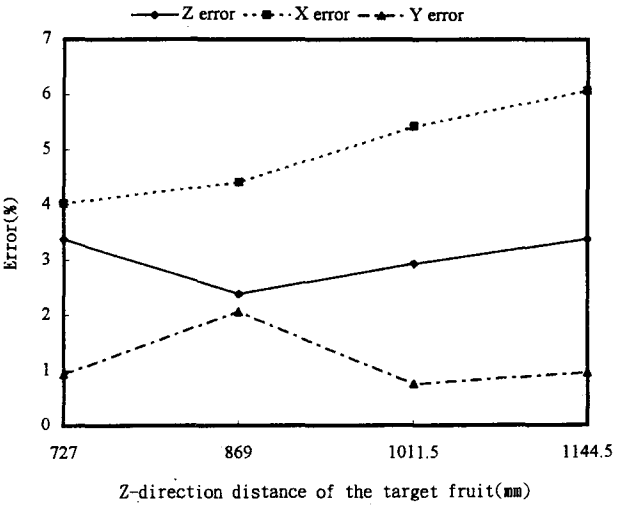

Fig. 19. 3-D location errors for different Z-direction distances of the target fruit

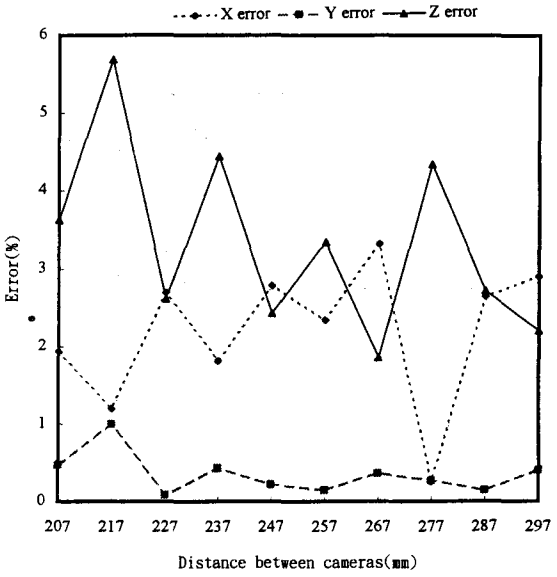

Fig. 20. 3-D location errors for different distances between cameras

Table 4. 3-D location results for Z values around 727 mm

Test no.	Measured values(mm)			Calculated values(mm)		
	Z	X	Y	Z	X	Y
1	729.00	209.00	418.00	752.87	201.87	419.71
2				758.97	202.98	420.55
3				758.87	202.38	420.53
4				758.97	202.98	421.03
5				758.97	202.98	421.03
Average values				757.73	202.64	420.57
Average errors(%)				3.94	3.04	0.61
i	718.00	312.00	419.00	739.16	336.60	423.89
2				741.19	343.00	428.15
3				741.20	343.60	428.17
4				739.28	343.15	427.88
5				741.21	344.20	428.66
Average values				740.41	342.11	427.35
Average errors(%)				3.12	9.65	1.99
1	727.00	358.00	419.00	758.74	361.61	421.86
2				760.71	362.72	422.15
3				760.71	362.72	422.15
4				752.84	360.74	421.53
5				754.80	361.23	421.32
Average values				757.56	361.80	421.80
Average errors(%)				4.20	1.06	0.67
1	734.00	114.00	418.00	751.17	116.68	420.46
2				751.17	116.67	419.97
3				751.17	116.67	419.97
4				747.25	116.67	418.96
5				751.17	116.67	419.97
Average values				750.39	116.67	419.87
Average errors(%)				2.23	2.34	0.45
Average Z value	727.00					
Total average error(%)				3.37	4.02	0.93

Table 5. 3-D location results for Z values around 869 mm

Test no.	Measured(mm)			Calculated values(mm)		
	Z	X	Y	Z	X	Y
1	872.00	109.00	419.00	891.09	110.54	427.90
2				890.89	109.21	428.45
3				880.79	109.29	427.73
4				880.79	109.29	427.73
5				890.89	109.21	428.45
Average values				886.89	109.51	428.05
Average errors(%)				1.71	0.47	2.16
1	877.00	365.00	419.00	897.83	398.62	430.21
2				895.23	398.66	430.44
3				895.23	398.66	430.44
4				900.44	398.58	430.53
5				905.60	399.89	431.20
Average values				898.87	398.88	430.56
Average errors(%)				2.49	9.28	2.76
1	863.00	248.00	419.00	893.75	259.43	428.65
2				886.26	259.14	427.70
3				886.26	259.14	427.70
4				886.29	260.46	426.62
5				896.31	261.77	427.88
Average values				889.77	259.99	427.71
Average errors(%)				3.10	4.83	2.08
1	864.00	322.00	419.00	876.02	330.13	423.35
2				893.89	333.58	425.60
3				881.10	331.78	424.00
4				881.10	331.78	424.00
5				883.67	332.27	424.32
Average values				883.15	331.91	424.25
Average errors(%)				2.22	3.08	1.25
Average Z value	869.00					
Total average error(%)				2.38	4.41	2.06

Table 6. 3-D location results for Z values around 1012 mm

Test no.	Measured values			Calculated values		
	Z	X	Y	Z	X	Y
1	1011.00	226.00	419.00	1024.80	231.73	423.55
2				1034.45	233.37	425.23
3				1037.64	232.91	425.58
4				1037.64	232.91	425.58
5				1037.64	232.91	425.58
Average values				1034.44	232.77	425.10
Average errors(%)				2.32	2.99	1.46
1	1011.00	343.00	419.00	1033.23	357.34	422.49
2				1043.10	360.00	423.57
3				1043.10	360.00	423.57
4				1043.10	360.00	423.57
5				1039.80	359.37	423.22
Average values				1040.47	359.34	423.29
Average errors(%)				2.91	4.76	1.02
1	1011.00	178.00	419.00	1047.22	183.31	417.10
2				1053.94	193.70	417.78
3				1043.99	183.88	417.40
4				1043.99	183.88	417.40
5				1043.98	183.88	416.77
Average values				1046.63	185.73	417.29
Average errors(%)				3.52	4.34	0.41
1	1013.00	31.00	419.00	1048.42	35.26	421.56
2				1040.96	33.84	417.47
3				1040.96	33.84	417.47
4				1037.00	31.66	415.16
5				1048.42	35.26	421.56
Average values				1043.15	33.97	418.65
Average errors(%)				2.98	9.59	0.08
Average Z value	1011.50					
Total average error(%)				2.93	5.42	0.74

Table 7. 3-D location results for Z values around 1145 mm

Test no.	Measured values(mm)			Calculated values		
	Z	X	Y	Z	X	Y
1	1143.00	182.00	401.00	1177.65	192.35	398.06
2				1181.93	192.55	398.40
3				1177.65	192.35	398.06
4				1177.65	192.35	398.06
5				1190.42	194.75	400.93
Average values				1181.06	192.87	398.71
Average errors(%)				3.33	5.97	0.57
1	1143.00	291.00	401.00	1185.81	298.12	398.01
2				1190.12	299.51	398.36
3				1173.46	297.42	396.37
4				1190.12	299.51	398.36
5				1177.60	297.94	397.38
Average values				1183.42	298.50	397.70
Average errors(%)				3.54	2.58	0.82
1	1144.00	380.00	401.00	1190.77	397.44	400.29
2				1190.77	397.44	400.29
3				1170.92	392.63	398.69
4				1178.80	394.26	399.01
5				1178.80	394.26	399.01
Average values				1182.01	395.20	399.46
Average errors(%)				3.32	4.00	0.38
1	1148.00	29.00	401.00	1198.88	33.53	393.81
2				1198.85	33.53	393.12
3				1173.69	31.26	392.24
4				1173.69	31.26	392.24
5				1182.86	32.45	392.76
Average values				1185.59	32.40	392.83
Average errors(%)				3.27	11.74	2.04
Average Z value	1144.50					
Total average error(%)				3.37	6.07	0.95

76 % and 0.08 % , respectively; and they are 4.20 % and 1.71 % , respectively, in the Z direction. The total average errors of Tables 4 ~ 7 were plotted on Figure 19. This figure reveals that the X-direction errors were proportional to fruit distances, while the errors in the Y and Z directions were more constant. The results of 3-D locating test for different distance between cameras are shown in Figure 20. Throughout this test, the sample fruit was placed at (252mm, 428mm, 848mm) with respect to the world coordinate system. Figure 20 shows that all errors were less than 6 % . Any inaccuracies of the camera parameters, angle resolutions, and template matching would cause the 3-D locating errors.

Table 8 displays the execution time for each operation performed in the searching and locating processes. The total computing time required to detect and locate a fruit was approximately 1 minute and 41 seconds. All algorithms were implemented in software. If all the operations could be perfomed in hardware, the total execution time needed would be greatly reduced.

Table 8. Execution time for each operation

Operation	Execution time(seconds)
Conversion from RGB to HSI	12
Image segmentation	3
Logic "AND"	8
Erosion and dilation	21
Labeling	15
Rotation of vergence shaft	25
Template matching	15
3-D position calculation	2
Total execution time	1 min.41s

SUMMARY AND CONCLUSIONS

The automatic tree fruit searching and locating machine of this study consisted of a camera-pointing device and a stereo machine vision system. Three stepping motors along with reduction gears were used to control the pan, tilt, and vergence shafts. A kumquat bonsai was employed as the sample fruit tree. Experiments were conducted in the laboratory under constant controlled illumination of 750 luxes. Image captured by the right camera was transformed from the RGB model to the HSI model prior to image processing. The thresholded red image and hue image were used to check the presence of fruit.The threshold value for the red image was 180. For the hue image, double thresholds, 35 and 10, were utilized in the segmentation operation. The reason for using hue image is to segment out the unripe fruit; only ripe fruit was located in the test. Other image processing techniques such as logic "AND", erosion, dilation, and labeling were employed as well. A template, which was 90 pixels wide and 4 pixels high, centered at the centroid of the targeted fruit of the right image was used in template matching process. The correlation between the right image and left image was computed along the epipolar line in the left image to find the corresponding pixels necessary for 3-D position determination. The pan, tilt, and vergence angles computed by the computer, the image distance, the positions of the corresponding pixels with respect to the camera coordinates along with the positions of the cameras were used in 3-D location of the target fruit. The success rate of fruit searching was about 93 % . In the 3-D locating test, fruits were placed at 4 different distances from the cameras and the average locating errors were 4.98 % , 1.17 % , and 3.01 % , respectively, in the X, Y, and Z directions. Tests were also conducted with the fruit distance fixed but varying the baseline and the resulting errors were all less than 6 % . The performance of this fruit searching and locating system is considered to be satisfactory and will be used in future research in automatic fruit harvesting.

ACKNOWLEDGMENT

This work was supported by National Science Council, Republic of China under Grant NSC 84-2321-B-005-007.

REFERENCES

1. Alden C., M. J. Lichtensteiger and R. G. Holmes. 1988. Nursery plant detection and assessment using image analysis. ASAE Paper No. 883055.

2. Alvertos, N., D. Brzakovic and R. C. Gonzalez. 1989. Camera geometries for image matching in 3-D

(文轉第 19 頁)