Scientia Iranica B (2015) 22(6), 2188–2197

Sharif University of Technology
Scientia Iranica
Transactions B: Mechanical Engineering
www.scientiairanica.com

A novel computational model of stereo depth estimation for robotic vision systems

K.-Y. Chen*, Ch.-H. Chen and Ch.-Ch. Chien

Department of Mechanical Engineering, Chung Yuan Christian University, 200, Chungpei Rd., Chungli District, Taoyuan City, 32023, Taiwan, R.O.C.

Received 22 May 2014; received in revised form 4 May 2015; accepted 15 September 2015

KEYWORDS
Stereovision;
Variable focal length;
Robotic vision;
Affine deformation;
Non-uniform spacing.

Abstract. This paper presents a novel computational model of stereovision for improving the accuracy of three-dimensional data extracted from a stereo-pair image with no effect of changes in focal length. For decades, most previous studies on stereovision have focused on the establishment of stereo matching, and have made conclusions on the premise of a fixed focus. In general, error in the depth estimate becomes bigger when the focus and aperture are unknown or not fixed. For that reason, a three-stage framework is proposed in this paper to modify the conventional stereovision model for improving the accuracy of depth estimation. The first stage is to modify the computational model of conventional stereovision for varifocal cameras. Then, the spacing of depth intervals in the non-uniform spacing of discrete depth levels can be altered, in particular, to be unaffected by changes in focal length. Finally, by considering the affine transformation, we add the deformation coefficient into the modified stereovision model for correcting three-dimensional affine deformations. Experimental results demonstrated that the depth estimation from stereo images using the proposed scheme was more accurate than conventional methods. The percentage error of most estimates fell between 0.06%-0.82%, and the error value increased from 0.02 cm to 2.21 cm within 6 m.

1. Introduction

Extracting 3D data from a pair of stereo images is called binocular vision or stereovision, and has been studied by researchers for decades. The common binocular vision model uses two cameras placed in parallel to imitate the distance and depth-perception abilities of the human visual system. In general, depth information can be used to reproduce an object's 3D structure in a scene, track moving objects for visual navigation, or measure distance information for optical inspection systems [1,2]. The basic principle of stereovision is to determine the distance between object and camera using the positions of the object in the left and right images, simultaneously, and their geometrical relationship. There are two major problems of a stereovision model: 3D reconstruction and stereo matching. Over the past few decades, a number of different approaches have been proposed in the literature for solving these problems. Marr and Poggio [3] firstly analyzed the computational model of the stereo-disparity problem and proposed a cooperative algorithm to implement the computation of disparity information from a pair of stereo images. Barnard and Fischler [4] reviewed various computational stereo techniques and provided a representative sampling of computational stereo research. After reviewing advances in computational stereo from roughly the early 1990s to the early 2000s,

*. *Corresponding author. Tel.: +886-3-2654322; Fax: +886-3-2654399*
E-mail addresses: gychen@cycu.edu.tw (K.-Y. Chen); bpatrick771017@yahoo.com.tw (Ch.-H. Chen); ccash_chien@gss.com.tw (Ch.-Ch. Chien)

Brown et al. [5] focused primarily on three topics: correspondence techniques, methods for occlusion, and real-time implementations. However, many real-time stereovision applications, such as visual serving, visual navigation, and obstacle avoidance, do not necessarily need to extract accurate depth information. Therefore, several studies developed some active vision schemes for computing the relative depth of points in a pair of stereo images with no camera calibration or prior knowledge of the parameters of the stereovision system required [6,7]. Besides, because color images can provide more information than grey-level images, considerable research work has been done to offer various methods for matching color images [8].

As mentioned above, most studies in stereovision have focused on the establishment of stereo matching and all have assumed that the focal length of the lens is constant. However, varifocal lens systems have become more commonly available in recent years for mobile phones, digital cameras, and machine vision applications. Sengupta [9] analyzed the effects of unequal focal lengths, and derived the equations of modified epipolar lines. Furthermore, several studies have reported dealing with the design problem of the operation mechanism and effects of image acquisition with an extended depth of field for variable-focus liquid lenses [10].

There is an important issue in stereovision using varifocal cameras that should be noticed. If the focal length of the lens is not constant or exactly known, errors of depth estimation based on the conventional stereovision model will increase. More recently, a modified stereovision model without a focus factor has been proposed to replace conventional stereovision [11]. This modified stereovision model can decrease the error of depth estimation about one half compared with the conventional stereovision model, but only for the range of 2 m. In this paper, we propose further amendments to improve the depth estimation accuracy by expanding the scope of the application in two steps. First, due to inadequate camera resolution, object baselines with smaller depth difference vanish, and this seriously restricts the depth resolution. In this case, non-uniform spacing of the discrete depth level increases sharply with an increase in depth and distance. Second, when non-Euclidean geometric space, consisting of digital images, is used to inversely estimate scaling and affine space warping in affine space along the direction of the Z axis, the magnified effect of the non-uniform discrete depth level is not the uniform magnification of the square pixel in a scale of 1:1. We use the affine space warping coefficient to correct the above drawback.

The remainder of this paper is organized as follows. Section 2 briefly reviews the fundamentals of conventional and modified stereovision models. Section 3 presents the proposed correction equations of the modified stereovision without the focus factor. Experimental results of the proposed scheme are presented in Section 4. Finally, conclusions are given in Section 5.

2. Stereovision model

2.1. The conventional stereovision model

The conventional stereovision grabs a pair of images simultaneously using two horizontally placed cameras, as shown in Figure 1. Here, we assume that the

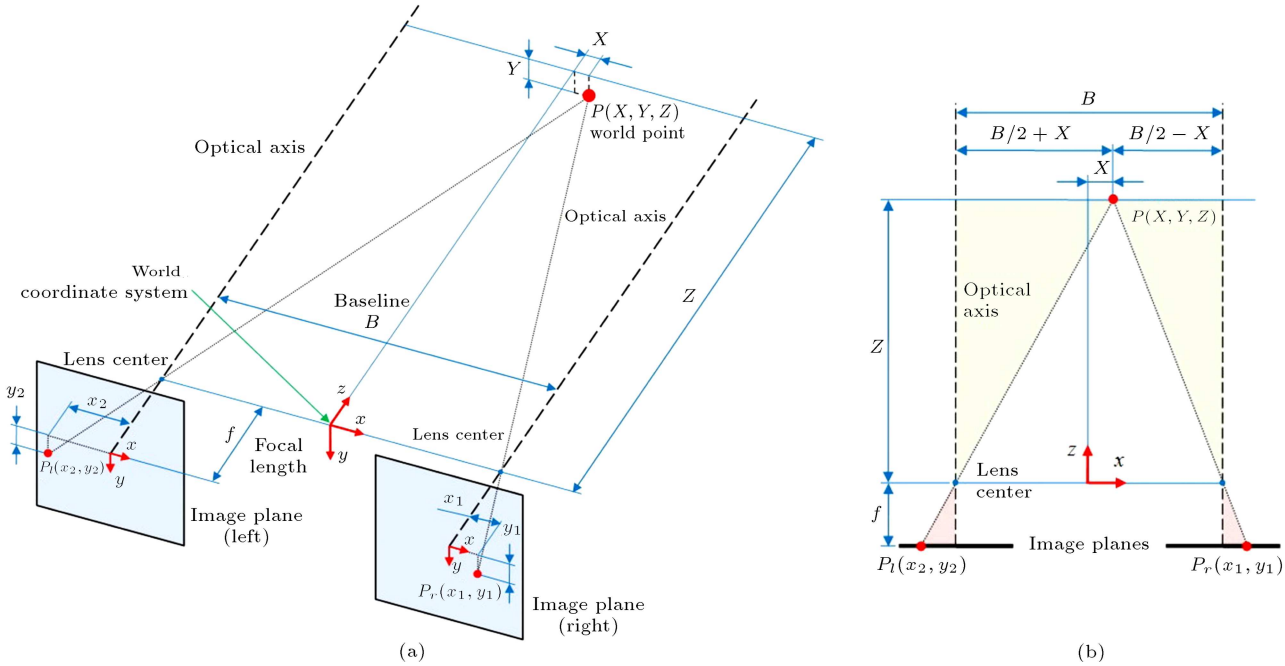


Figure 1. Parallel axis geometry of conventional stereo vision: (a) Three-dimensional view; and (b) top view.

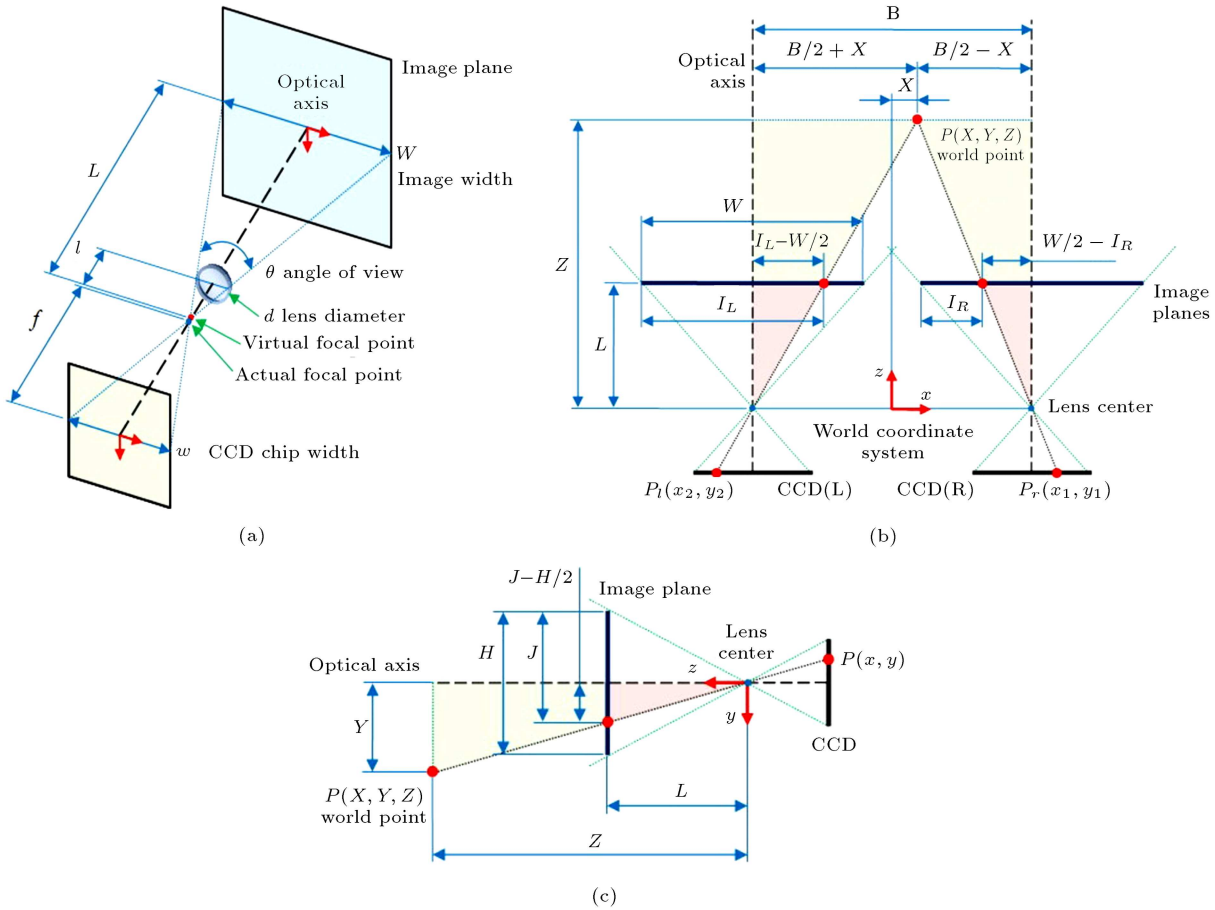


Figure 2. The modification of stereovision model: (a) CCD imaging; (b) top view; and (c) right view.

corresponding points in two images are identified, and the parameters of two cameras are known. Then, the geometry of stereovision can be simplified as a triangulation. In Figure 1, the objective of stereovision is to find the coordinates (X, Y, Z) of the world point, P, having corresponding points (x_1, y_1) and (x_2, y_2) in left and right images, respectively [12]. This is easily done using similar triangles, that is:

$$\frac{\frac{B}{2} - X}{x_1} = \frac{\frac{B}{2} + X}{-x_2} = \frac{Y}{y_1} = \frac{Y}{y_2} = \frac{Z}{f}, \tag{1}$$

where f is the focal length, and B, called the baseline, is the distance between the centers of the two lenses. The world coordinates (X, Y, Z) may be computed as follows:

$$X = \frac{B}{2}\left(\frac{x_1 + x_2}{x_2 - x_1}\right), \quad Y = \frac{By_1}{x_1 - x_2} = \frac{By_2}{x_1 - x_2},$$

$$\text{and} \qquad Z = \frac{Bf}{x_1 - x_2}. \tag{2}$$

Eq. (2) indicates that if baseline B and focal length f are known, and the corresponding image coordinates (x_1, y_1) and (x_2, y_2) can be determined, computing the coordinates (X, Y, Z) of P is a simple matter.

2.2. The modified stereovision model

In general, an important problem in variable focus stereovision is that the focal length of the lens may not be actually constant. For this reason, using Eq. (2) will reduce the accuracy of depth dimension [13,14]. Chen et al. [11] proposed a modified stereovision model for improving the accuracy of depth estimation from a pair of stereo images, especially suitable for varifocal cameras, as shown in Figure 2. Two cameras are separated by a distance, B, in the x-direction, and both optical axes are parallel. For convenience, the coordinate system centered between two cameras is called the world coordinate system.

As shown in Figure 2(a), distance L (in centimeters) between the image plane and the virtual focal point can be easily measured using a simple camera calibration procedure as follows. First, by measuring diameter, d, and the angle of view, θ, of the lens, the distance, l, between the lens and the virtual focal point can be obtained. Second, if an arbitrary image plane is in front of the camera and its location is known, the width of the captured image can be measured. From the similar triangles of imaging, as shown in Figure 2(b) and (c), Eq. (1) becomes as follows:

$$\frac{\frac{B}{2}-X}{\frac{W}{2}-I_R}=\frac{\frac{B}{2}+X}{I_L-\frac{W}{2}}=\frac{Y}{J-\frac{H}{2}}=\frac{Z}{L}, \tag{3}$$

where I_L and I_R are the horizontal positions, and J is the vertical position of the object in the two images (in pixels), W and H are the image width and height (in centimeters), respectively. In Eq. (3), only the units of I_L, I_R, and J are in pixels. Here, assume that the image resolution is $M \times N$ (in pixels). After converting pixels to centimeters, Eq. (3) can be expressed as:

$$\frac{\frac{B}{2}-X}{\frac{W}{2}-I_R\frac{W}{M}}=\frac{\frac{B}{2}+X}{I_L\frac{W}{M}-\frac{W}{2}}=\frac{Y}{J\frac{H}{N}-\frac{H}{2}}=\frac{Z}{L}. \tag{4}$$

Then, the world coordinates (X, Y, Z) may be computed as follows:

$$X=\frac{B(I_L+I_R-M)}{2(I_L-I_R)}, \qquad Y=\frac{BMH(2J-N)}{2NW(I_L-I_R)},$$

$$Z=\frac{BLM}{W(I_L-I_R)}. \tag{5}$$

Compared with Eq. (2), the term focal length, f, does not appear in Eq. (5). Note that all coordinates obtained with Eq. (5) are with respect to the world coordinate system centered between two cameras.

3. Correction equation of modified stereovision without focus

3.1. Depth error correction of non-uniform discrete depth level

Real space is a description based on Euclidean space, and cannot describe perspective projection. In Euclidean space, parallel lines and standard circles are used, but Euclidean space is a tactile, not visual, geometry [15]. As shown in Figure 3, visual geometry has no parallel lines, and all lines near or far may intersect on the vanishing point in space. From visual geometry, the sensing depth of the similar spaces can be estimated. The particle size in the actual image panel is not infinitely small, and the pixel is the smallest unit. As for object depth, smaller baselines may seriously restrict achievement of depth resolution. The granularity of pixels caused by the limited resolution of digital images may result in a discrete discernible depth level, and the affected interval may sharply increase with depth and distance.

Based on the non-uniform spacing of the discrete depth level of two parallel cameras, the depth and distance, Z and ΔZ, (spacing of depth interval) can be deduced, based on the relationship between similar triangles [16].

$$\Delta Z=\frac{\delta}{fB}Z(Z+\Delta Z), \tag{6}$$

Figure 3. All lines in visual geometry, either near or far, may intersect on vanishing point in the space.

where Z is the sensing depth (mm) from the baseline center point to some point in space; ΔZ is the spacing of depth interval (mm); δ is the smallest unit of the image (pixel); B is the baseline, namely distance (mm) between the two lens; f is the focal length of the lens, i.e. distance (mm) from pin hole aperture to image plane.

In Eq. (6), when ΔZ is smaller than Z, it can be reduced into:

$$\Delta Z \approx \frac{\delta}{fB}Z^2. \tag{7}$$

In the background environment of the variable focus lens system, the geometric structures of the non-uniform spacing of the discrete depth level of the two parallel cameras with two lenses are shown in Figure 4. In varifocal stereovision, the focal length of the lens can be unfixed. Thus, we modified Eqs. (6) and (7) into the correction equations of non-uniform spacing of the discrete depth level without focus. The baseline length is B and the distance between the image plane and the virtual focal point is L,. Thus, the depth and distance Z and ΔZ (spacing of depth interval) can be deduced

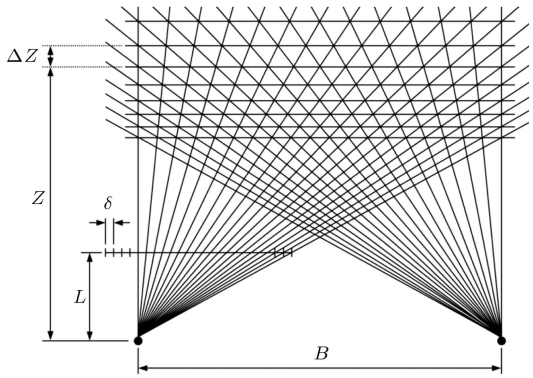


Figure 4. Non-uniform spacing of discrete depth level of variable focus lens system.

based on the relationship between similar triangles.

$$\Delta Z = \frac{\delta \frac{W}{M}}{LB} Z(Z + \Delta Z) = \frac{\delta W}{LBM} Z(Z + \Delta Z), \qquad (8)$$

where W and H are width and height (mm) of images, respectively; $M \times N$ represents image pixel.

When ΔZ is smaller than Z, Eq. (8) can be reduced into:

$$\Delta Z \approx \frac{\delta W}{LBM} Z^2. \qquad (9)$$

The above ΔZ is called the correction equation of the non-uniform spacing of the discrete depth level without focus, and $Z + \Delta Z$ is the result corrected by the modified stereovision algorithm. The corrected algorithm can be expressed as:

$$X = \frac{B(I_L + I_R - M)}{2(I_L - I_R)},$$

$$Y = \frac{BMH(2J - H)}{2NW(I_L - I_R)}, \quad \text{and}$$

$$Z = \frac{BLM}{W(I_L - I_R)} + \frac{\delta W}{LBM} \left[\frac{BLM}{W(I_L - I_R)}\right]^2. \qquad (10)$$

The error correction equations of the non-uniform spacing of the discrete depth level without focus are an ideal error correction method. The theoretical basis is a normal view perspective and a single-point perspective. In cases of other perspectives, or two-point or more perspectives, other methods should be used for correction. Besides, the digital image warping characteristics of space geometry changes are not considered. Thus, the second correction is proposed to increase accuracy.

3.2. Depth error correction of affine geometric space distortion

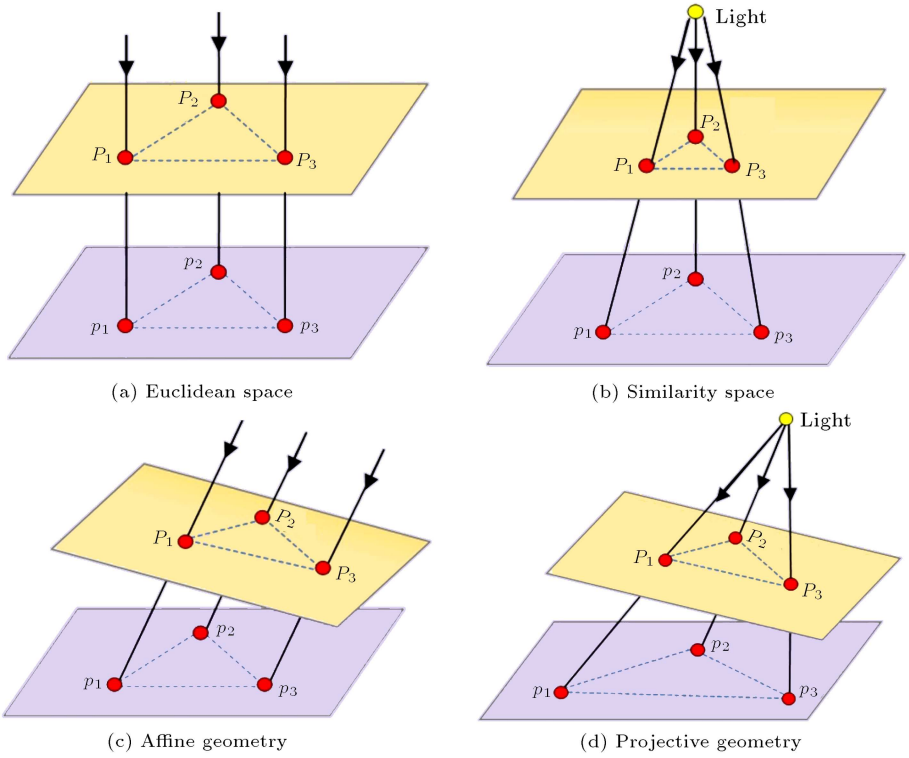
In the machine vision field, studies on depth estimation often use the imaging principle of finite camera models for perceptive projection to estimate depth. However, we still have to face real world (Euclidean space) and image plane (vision geometry space) transformation problems. Geometric space is displayed in 2D space, as shown in Figure 5, and diagrams of the results, after geometric transformations in visual geometric space, are as shown in Figure 6.

In visual geometry, direct reflection, diffusion and reflection of the light source are mixed. Similarity geometry, affine geometry and projective geometry are included and change at all times and places. Projective transformation can be divided into a combination of three transformations, as follows [15]:

$$\mathbf{M}_p = \mathbf{H}_s \mathbf{H}_a \mathbf{H}_p = \begin{bmatrix} s\mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{K} & 0 \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{v}^T & k \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{v}^T & k \end{bmatrix}, \qquad (11)$$

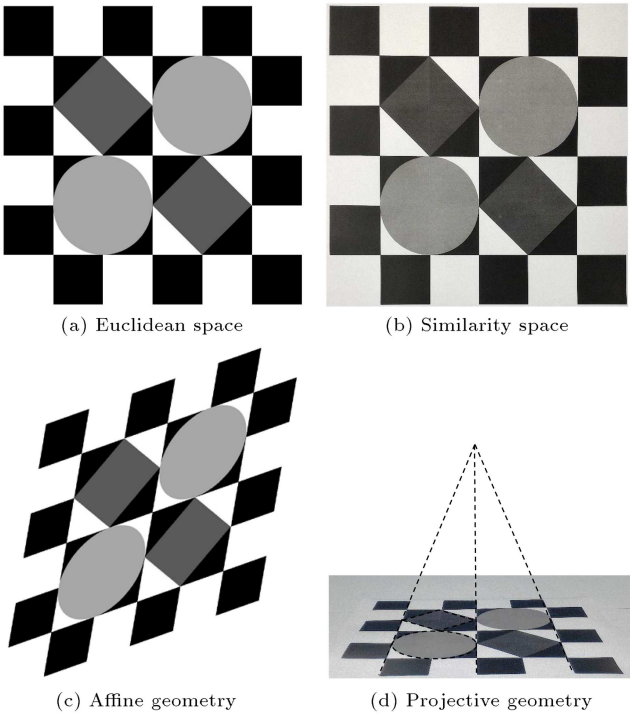
where $\mathbf{R}$ is a rotation matrix; $\mathbf{t}$ is a translation matrix;



(a) Euclidean space (b) Similarity space

(c) Affine geometry (d) Projective geometry

Figure 5. Diagrams of 2D geometric space.



(a) Euclidean space (b) Similarity space

(c) Affine geometry (d) Projective geometry

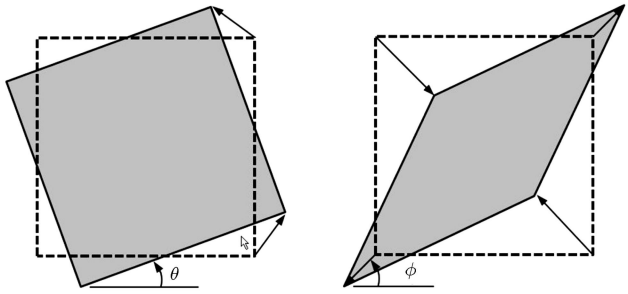
Figure 6. Diagrams of visual geometric space.

$\mathbf{A}$ is a non-singular matrix, and the given value is $\mathbf{A} = s\mathbf{RK}+\mathbf{tv}^T$, which represents the scaling and rotational deformation state; $\mathbf{K}$ is an upper-triangular matrix and $\det(\mathbf{K})=1$ after normalization; $\mathbf{v}^T$ is a 3×1 vector matrix, which represents that scaling and rotational deformation is extended at infinity; k is a homogeneous scale value, corresponding to the generated matrix model; $\mathbf{H}_s$ is the similarity transformation matrix; $\mathbf{H}_a$ is the affine transformation matrix with an increase in affine geometry; $\mathbf{H}_s\mathbf{H}_a$ is the affine transformation with an increase in affine geometric factors; $\mathbf{H}_p$ is the projection transformation matrix with an increase in projective geometric factors.

Digital image distortion is used to estimate the scaling and affine space warp on affine space along the direction of the Z axis, because the affine transformation matrix can be expressed as follows when no coordinate transformation occurs:

$$\mathbf{M}_a = \mathbf{H}_s\mathbf{H}_a = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix}. \tag{12}$$

The geometric effect of the affine transformation of linear component $\mathbf{A}$ involves two fundamental transformations, namely, rotational and non-isotropic scaling. Generally, the equivalent synthetic vision of the area-of-interest of the stereovision system is centered. Thus, only extrusion deformation, scaling and the rotation of the Z axis are considered, as shown in Figure 7. The affine matrix, $\mathbf{A}$, can be decomposed using singular value decomposition.



(a) Rotation (b) Deformation

Figure 7. Effects of two affine deformations on image plane.

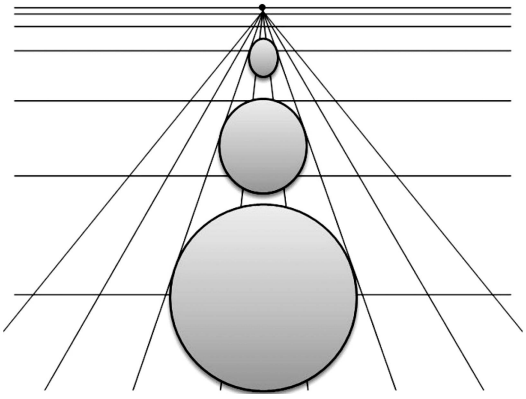


Figure 8. Extrusion deformation effect of Z axis towards the direction of Z axis may cause deformation of Y axis.

$$\mathbf{A} = \mathbf{UDV}^T = (\mathbf{UV}^T)(\mathbf{VDV}^T) = \mathbf{R}(\theta)\mathbf{R}(-\phi)\mathbf{DR}(\phi), \tag{13}$$

where $\mathbf{R}(\theta)$ represents rotation of the Z axis, $\mathbf{R}(\phi)$ represents space extrusion deformation caused by rotation of the Z axis, and D is a diagonal matrix, as follows:

$$\mathbf{D} = \begin{bmatrix} \lambda_x & 0 & 0 \\ 0 & \lambda_y & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{14}$$

where λ_x represents the scaling and deformation rate of the X axis, and λ_y represents the scaling and deformation rate of the Y axis.

As shown in Figure 8, in affine geometric space, the extrusion deformation effect of the Z axis on the direction of the Z axis may cause deformation in the direction of the Y axis. A circle may be deformed into a scaled elliptical shape; it is stretched longer and longer when it moves toward the direction of the Z axis, and the stretch will not stop until an elliptical shape disappears in the vision plane. Finally, all the circles vanish at an affine critical point. λ_x is assumed to have zoom scale, and thus, λ_y is space scaling, and the extrusion deformation rate (zoom in) of the Z axis is increased. When the scale ratio, λ_x, of 1:1 is proposed in similarly geometric space, the extrusion deformation rate (zoom in) of Z axis is approximately equal to the

ratio of the minimum to maximum axial lengths of the ellipse:

$$\lambda_{\frac{x}{y}} \approx \frac{\text{Minor axis length}}{\text{Major axis length}}. \quad (15)$$

Physically, for the smallest unit, δ, of images, the non-uniform spacing of the discrete depth level is enlarged, not in equal proportion, but in scale ratio, $\lambda_{\frac{y}{x}}$, in affine space. In consideration of what is viewed by binocular vision, the affine space warping coefficient, W_{arp}, represents the affine space warping viewed from left and right cameras:

$$W_{arp} = \frac{(\lambda_{\frac{x}{y}})_L + (\lambda_{\frac{x}{y}})_R}{2}, \quad (16)$$

where $(\lambda_{\frac{x}{y}})_L$ is the deformation rate viewed from the left camera and $(\lambda_{\frac{x}{y}})_R$ is the deformation rate viewed from the right camera, respectively. Thus, the spacing of depth interval, ΔZ, in correction Eq. (9) of the non-uniform discrete depth level without focus can be rewritten as follows:

$$\Delta Z \approx W_{arp} \frac{\delta W}{LBM} Z^2. \quad (17)$$

The above equation is called the correction equation of affine space warping. Thus, the modified stereovision algorithm can be expressed as:

$$X = \frac{B(I_L + I_R - M)}{2(I_L - I_R)},$$

$$Y = \frac{BMH(2J - N)}{2NW(I_L - I_R)},$$

$$Z = \frac{BLM}{W(I_L - I_R)} + W_{arp} \frac{\delta W}{LBM} \left[\frac{BLM}{W(I_L - I_R)}\right]^2. \quad (18)$$

4. Experimental results

In this paper, we designed the binocular vision system illustrated in Figure 9, and the values of the parameters

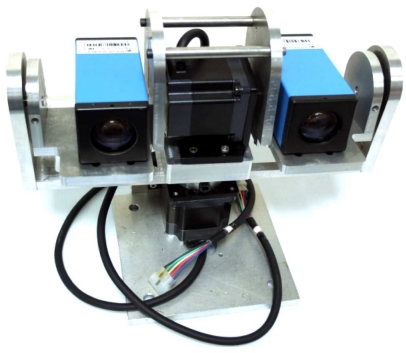


Figure 9. The experimental setup of binocular stereovision system.

Table 1. Camera parameters of the binocular vision system.

Image resolution	640×480
Image format	RGB
Baseline length	20 cm
Angle of view	42.8434°
Distance between image plane and virtual focal point	0.4715 cm
Average computing time (sec)	0.74 sec

of the system, such as image resolution, image format, baseline length, etc., are shown in Table 1. Two identical CCD cameras are mounted in parallel on a platform with two rotatory degrees of freedom. We applied the proposed stereovision algorithm to calculate the depth of the target centroid in two images, illustrated in Figure 10 and in the flow chart in Figure 11. The stereo images have a size of 640×480 pixels and the distance of the red target with diameter of 4.5 cm is increased from 70 cm to 270 cm with every interval of 20 cm, and then, from 300 cm to 500 cm with every interval of 100 cm. We can obtain the error in the correction equation of the non-uniform spacing of the discrete depth level and the error in the correction equation of the affine space warping, as shown in Tables 2 and 3. Next, we applied four types of stereovision model, respectively, to estimate the depth of the target centroid in two images, as shown in Figure 12, where OZ denotes the estimated depth value ($f = 45$ mm) in the traditional stereovision algorithm, MZ denotes the estimated depth value in the modified stereovision algorithm, $MZ1$ represents the estimated depth value obtained from the correction equation of the non-uniform spacing of the discrete depth level, and $MZ2$ represents the estimated depth value obtained from the correction equation of affine space warping. As a measurement of accuracy of the depth computed using the proposed method in this paper, an error parameter is defined as follows:

$$\text{error} = \frac{\text{computed depth} - \text{actual depth}}{\text{Actual depth}} \times 100\%. \quad (19)$$

In terms of the experiment results, from similarity geometric space to affine geometric space, the percentage error of the modified binocular vision algorithm in direct estimation of the sensing depth fell between -1.8-0.6%; the error increased from 0.7 cm to the vicinity of 6 cm. The percentage error of the correction equation of the non-uniform spacing of the discrete depth level fell between 0.01-1.43%, and the error increased from 0.01 cm to 3.87 cm. In proximal similarity space, the correction equation of the non-uniform spacing of the discrete depth level is very accurate. However, the affine space domain is not

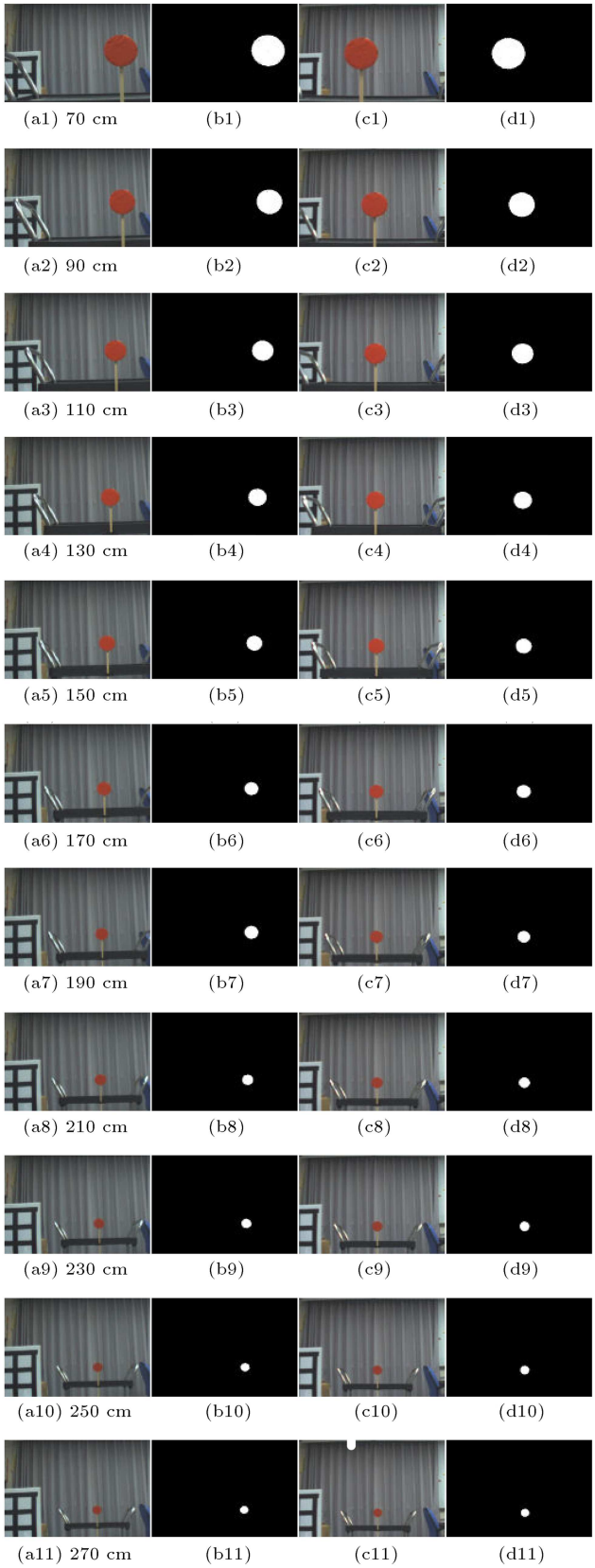


Figure 10. The stereo image pairs at different distances: (a) Captured by left camera; (b) after image processing; (c) captured by right camera; and (d) after image processing.

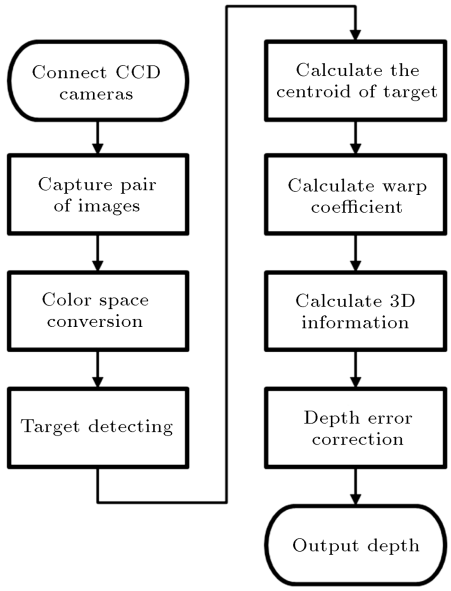


Figure 11. Process of stereoscopic image processing and depth error correction of affine geometric space distortion.

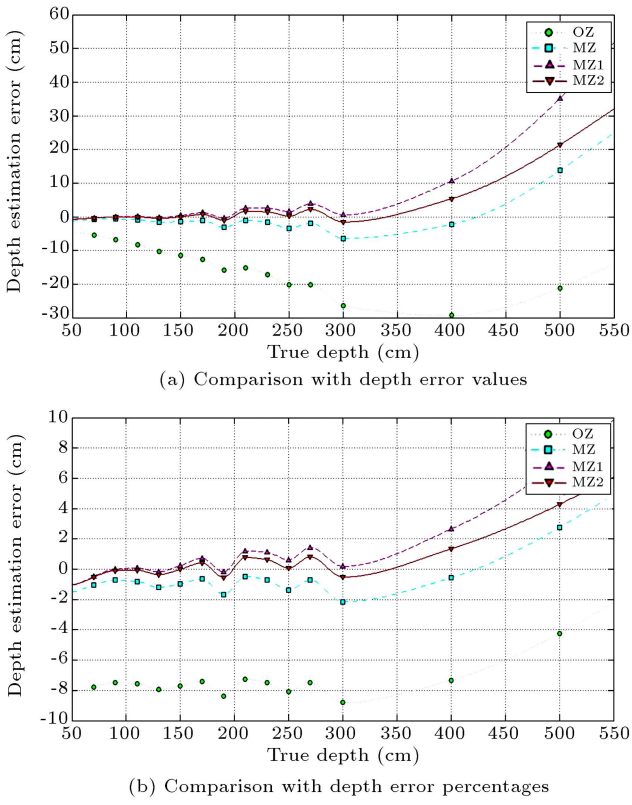


Figure 12. Comparison of depth estimation errors using four types of stereovision.

accurate. This is consistent with the assumption. The correction equation for the non-uniform spacing of the discrete depth level is an ideal error correction method, but fails to consider digital image deformation by geometric space.

The error correction equation for affine space

Table 2. Depth calculation and its error in correction equation of non-uniform spacing of discrete depth level.

Actual depth (cm)	Center-of-mass coordinates of left image	Center-of-mass coordinates of right image	Computed depth (cm)	Error (cm)	Error (%)
70.00	(171,368)	(131,190)	69.66	-0.34	-0.49%
90.00	(130,336)	(144,198)	89.99	-0.01	-0.01%
110.00	(138,325)	(151,212)	110.08	0.08	0.07%
130.00	(141,310)	(157,214)	129.77	-0.23	-0.17%
150.00	(147,300)	(158,217)	150.34	0.34	0.23%
170.00	(150,295)	(163,222)	171.22	1.22	0.72%
190.00	(152,290)	(165,224)	189.65	-0.35	-0.19%
210.00	(153,286)	(167,227)	212.52	2.52	1.20%
230.00	(155,282)	(168,228)	232.56	2.56	1.11%
250.00	(156,278)	(169,228)	251.53	1.53	0.61%
270.00	(157,278)	(171,232)	273.87	3.87	1.43%
300.00	(158,275)	(172,233)	300.56	0.56	0.19%
400.00	(161,266)	(174,235)	410.57	10.57	2.64%
500.00	(163,264)	(177,240)	535.15	35.15	7.03%

Table 3. Depth calculation and its error in correction equation of affine space warping.

Actual depth (cm)	Center-of-mass coordinates of left image	Center-of-mass coordinates of right image	Computed depth (cm)	Error (cm)	Error (%)
70.00	(171,368)	(131,190)	69.62	-0.38	-0.54%
90.00	(130,336)	(144,198)	89.91	-0.09	-0.10%
110.00	(138,325)	(151,212)	109.93	-0.07	-0.06%
130.00	(141,310)	(157,214)	129.53	-0.47	-0.36%
150.00	(147,300)	(158, 217)	149.98	-0.02	-0.02%
170.00	(150,295)	(163,222)	170.70	0.70	0.41%
190.00	(152,290)	(165,224)	188.97	-1.03	-0.54%
210.00	(153,286)	(167,227)	211.63	1.63	0.78%
230.00	(155,282)	(168,228)	231.45	1.45	0.63%
250.00	(156,278)	(169,228)	250.18	0.18	0.07%
270.00	(157,278)	(171,232)	272.21	2.21	0.82%
300.00	(158,275)	(172,233)	298.44	-1.56	-0.52%
400.00	(161,266)	(174,235)	405.35	5.35	1.34%
500.00	(163,264)	(177,240)	521.33	21.33	4.27%

warning is more accurate; the percentage error of most estimates fell between 0.06-0.82%, and the error value increased from 0.02 cm to 2.21 cm. When the proximal affine warping is not apparent, the correction equation for the non-uniform spacing of the discrete depth level on the similarity space domain is slightly better than that of the affine space warping. This is consistent with the assumption, although the difference is not great. When the affine transformation increases, the correction equation of affine space warping on the affine space domain is the stereovision algorithm with the highest accuracy so far.

5. Conclusions

In this paper, an efficient and accurate computational model of depth estimation based on a modified stereovision model is presented. The newly created correction equation for the non-uniform spacing of the discrete depth level is used to improve stereovision accuracy in

the similarity space. For the object depth, a smaller baseline may seriously restrict the depth resolution that can be achieved, and non-uniform spacing of the discrete depth level caused by the finite resolution of digital images, sharply increases with an increase in depth and distance. Thus, the correction equation of the non-uniform discrete depth level without focus is used to correct the error. Experimental results show that the proposed computational model has the ability to improve the accuracy of depth estimation. We can finally conclude that a slightly modified computational stereo model can extract the accurate depth information from a pair of stereo images. Besides, it is more important that this model is especially suitable for use in varifocal cameras.

Acknowledgment

This work was supported by the National Science Council of Taiwan (R.O.C.) under grant number NSC 102-2221-E-033-037.

References

1. Li, Z.N. and Hu, G. "Analysis of disparity gradient based cooperative stereo", *IEEE Transactions on Image Processing*, **5**(11), pp. 1493-1506 (1996).
2. Nitzan, D. "Three-dimensional vision structure for robot applications", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **10**(3), pp. 291-309 (1988).
3. Marr, D. and Poggio, T.A. "Cooperative computation of stereo disparity", *Science*, **194**(4262), pp. 283-287 (1976).
4. Barnard, S.T. and Fischler, M.A. "Computational stereo", *Computing Surveys*, **14**(4), pp. 553-572 (1982).
5. Brown, M.Z., Burschka, D. and Hager, G.D. "Advances in computational stereo", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **25**(8), pp. 993-1008 (2003).
6. Grosso, E. and Tistarelli, M. "Active/dynamic stereo vision", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(9), pp. 868-879 (1995).
7. Yau, W.Y. and Wang, H. "Fast relative depth computation for an active stereo vision system", *Real-Time Imaging*, **5**(3), pp. 189-202 (1999).
8. Ansari, M.E., Masmoudi, L. and Bensrhair, A. "A new regions matching for color stereo images", *Pattern Recognition Letters*, **28**(13), pp. 1679-1687 (2007).
9. Sengupta, S. "Effects of unequal focal lengths in stereo imaging", *Pattern Recognition Letters*, **18**(4), pp. 395-400 (1997).
10. Ren, H. and Wu, S.T. "Variable-focus liquid lens", *Optics Express*, **15**(10), pp. 5931-5936 (2007).
11. Chen, K.Y., Chien, C.C. and Tseng, C.T. "Improving the accuracy of depth estimation in binocular vision for robotic applications", *Applied Mechanics and Materials*, **284-287**, pp. 1862-1866 (2013).
12. Gonzalea, R.C. and Woods, R.E., *Digital Image Processing*, Addison-Wesley, New York, USA (1992).
13. Hsu, H.K., *Application of Stereo-Vision Navigation to the Acquisition of Position and Attitude of MAVs*, MS Thesis, Tamkang University, Taiwan (2011).
14. Hsieh, K.Y., *A Study on a Service Mobile Robot with Remote Control*, MS Thesis, Chung Yuan Christian University, Taiwan (2011).
15. Hartley, R. and Zisserman, A., *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, UK (2006).
16. Scharstein, D. "View synthesis using stereovision", PhD Dissertation, Cornell University, USA (1997).

Biographies

Kuan-Yu Chen received his BS and MS degrees in Mechanical Engineering from Chung Yuan Christian University, Taoyuan, Taiwan, in 1988 and 1995, respectively, and his PhD degree in Mechanical Engineering from the National Central University, Taoyuan, Taiwan, in 2008. He is currently Professor in the faculty of Mechanical Engineering at Chung Yuan Christian University, Taoyuan, Taiwan. His research interests are mainly in the area of machine intelligence, intelligent robotics, and automated optical inspection.

Chien-Hung Chen received his BS degree in Mechanical and Mechatronic Engineering from the National Taiwan Ocean University, Keelung, Taiwan, in 2007, and is currently a PhD degree student in Mechanical Engineering at Chung Yuan Christian University, Taoyuan, Taiwan. His research interests are mainly in the area of automated optical inspection, stereovision-based computational models, and, more recently, 3D reconstruction from line-scan cameras.

Cheng-Chin Chien received his BS, MS and PhD degrees in Mechanical Engineering from Chung Yuan Christian University, Taoyuan, Taiwan, in 1988, 2005 and 2013, respectively, and is currently senior software consultant at Galaxy Software Corporation, Taipei, Taiwan. His research interests are mainly in the areas of image processing and stereovision-based computational models, including, more recently, industry 4.0, big data implementations and cloud computing.