Model establishment and validation of non-similar stereo vision

Huang Fuyu[1], Zheng Zhe[2], Li Peijun[2], Xie Dabing[3], Huang Xinxin[1]

(1. Department of Electronic and Optics Engineering, Shijiazhuang Campus of Army Engineering University, Shijiazhuang 050003, China; 2. Unit 63850 of PLA, Baicheng 137001, China; 3. Unit 32181 of PLA, Xi′an 710000, China)

Abstract: By introducing the popular equidistant projection theory, a non-similar stereo vision was explored, and the three-dimensional (3D) location model and depth precision model were established. The correctness of proposed models was verified through carring out the 3D location experiment with a 115°×90° infrared stereo vision system. From the experimental results, the targets in different depths and large-airspace scene were located successfully. The depth error increases smoothly with increasing target depth, and it reached 1.32 m in the depth of 30 m. The error variation of 3D location model matched well with the results of depth precision model, and the whole error was obviously less than that of similar stereo vision model. The advantages of the proposed model is that the 3D coordinates of space target can be obtained directly without correcting the non-similar distorted images, and it contributes to enriching and improving the theory of stereo vision, and will help to spread the application of non-similar imaging theory in the fields of large-airspace situation awareness, target detection and so on.

Key words: stereo vision; non-similar imaging; 3D location; depth precision; large-airspace

CLC number: TN29; TP24 **Document code:** A **DOI:** 10.3788/IRLA201948.0825003

非相似立体视觉模型构建与验证

黄富瑜[1],郑 喆[2],李佩军[2],谢大兵[3],黄欣鑫[1]

(1. 陆军工程大学石家庄校区 电子与光学工程系,河北 石家庄 050003;
2. 中国人民解放军 63850 部队,吉林 白城 137001;
3. 中国人民解放军 32181 部队,陕西 西安 710000)

摘 要: 将非相似成像理论引入到立体视觉理论中,探索了一种非相似立体视觉机制,建立了空间目标三维定位模型和深度精度模型。根据流行的非相似等距投影理论,对非相似立体视觉三维定位方程进行了数理推导,并实验验证了所建模型的有效性。结果表明:利用三维定位模型可以准确地获取大空域场景中不同深度的多个目标位置信息,定位误差随着目标深度的增大而缓慢变大,在 30 m 深度处的定位误差达到 1.32 m,误差变化规律与深度精度模型结果吻合良好,且误差值远小于相似立体视觉模型的定位误差。所建模型的优势是无需对非相似畸变图像进行校正,便可较好地得到空间目标的三维坐标,研究工作对于拓展非相似成像理论在大空域态势感知、目标侦察等领域的应用具有重要意义。

关键词: 立体视觉; 非相似成像; 三维定位; 深度精度; 大空域

收稿日期:2019-03-11; **修订日期:**2019-04-21

基金项目:国家自然科学基金(61801507);陆军工程大学基础前沿创新项目

作者简介:黄富瑜(1985-),男,讲师,博士,主要从事计算机视觉与图像处理方面的研究。Email: hfyoptics@163.com

0 Introduction

As one of the important branches in computer vision systems, the binocular stereo vision can directly imitate the human eye and human visual perception, and has been widely used in industrial inspection[1], virtual reality [2], position detection[3], vision navigation[4], 3D non-contact measure[5] and so on. However, the study object of current stereo vision focuses on the binocular camera satisfying the law of similar imaging mostly, and it only provides a limited field of view (FOV) which is hard to meet the needs of wide-area vision navigation, space location and 3D reconstruction. To generate larger-airspace stereo video, the multi-camera mechanism is adopted in Refs. [6] and [7], and the omnidirectional vision is obtained. The multi-camera vision is relatively easy to realize and has been used in surrounding detection[8], but it is difficult to generate stereo vision for the same larger-airspace scene. The appearance of non-similar wide-angle camera can solve well the problem of smaller FOV of similar camera, and it is reported that the binocular non-similar camera has been used as obstacle avoidance system to perceive real-timely large-space forward-looking scene both in American Mars rover [9] and China′s "Hade Rabbit" lunar rover [10]. It is clear that the non-similar stereo vision will advantage the large-airspace, real-time situation awareness and obstacle avoidance, and the accuracy of 3D target location is important to above applications. In Refs.[11] and [12], the non-similar fisheye stereo cameras are mounted on the car and UAV for obstacle avoidance, but the distortion correction is done before locating the obstacle, and the pinhole location model is used to obtain the obstacle depth, which may influence the real-time efficiency and location accuracy. Thus, by introducing the non-similar imaging theory into the stereo vision research, the non-similar stereo vision model including the 3D location model and depth precision model will be established without dedistortion process, and the proposed model will be verified by carrying out the contrast experiment.

1 Theoretical model

1.1 3D location model

In a variety of non-similar imaging theories, the equidistant projection is considered to be the most popular imaging way, whose imaging formula satisfies[13]:

$$y'=f\omega \tag{1}$$

Where y' is the imaging height; f is the focal length of optical system; ω is the half-field angle of the incident ray.

Based on the equidistant imaging principle, the framework of non-similar stereo vision is constructed (see Fig.1), and it consists of left and right vision units. In each vision unit, O represents the camera optical center, o represents the image center, $OXYZ$ represents the camera coordinate system, o_ouv represents the image coordinate system, oxy represents the image-plane coordinate system, $O_WX_WY_WZ_W$ represents the world space coordinate system, and the image-plane coordinate axes ox and oy are in parallel with the camera coordinate axes OX and OY. B represents the baseline distance between two vision units, and the subscripts of L, R represent the affiliations of each symbol in the left and right coordinate system.

The operating principle of non-similar stereo vision is described as follows. Firstly, the space target $P(X_WY_WZ_W)$ in the world space system is mapped at the equivalent lens surfaces of left and right vision units, and the mapping points are $P_L'(\omega_L, \varphi_L)$ and $P_R'(\omega_R, \varphi_R)$ in the coordinate systems of $O_LX_LY_LZ_L$ and $O_RX_RY_RZ_R$, respectively.

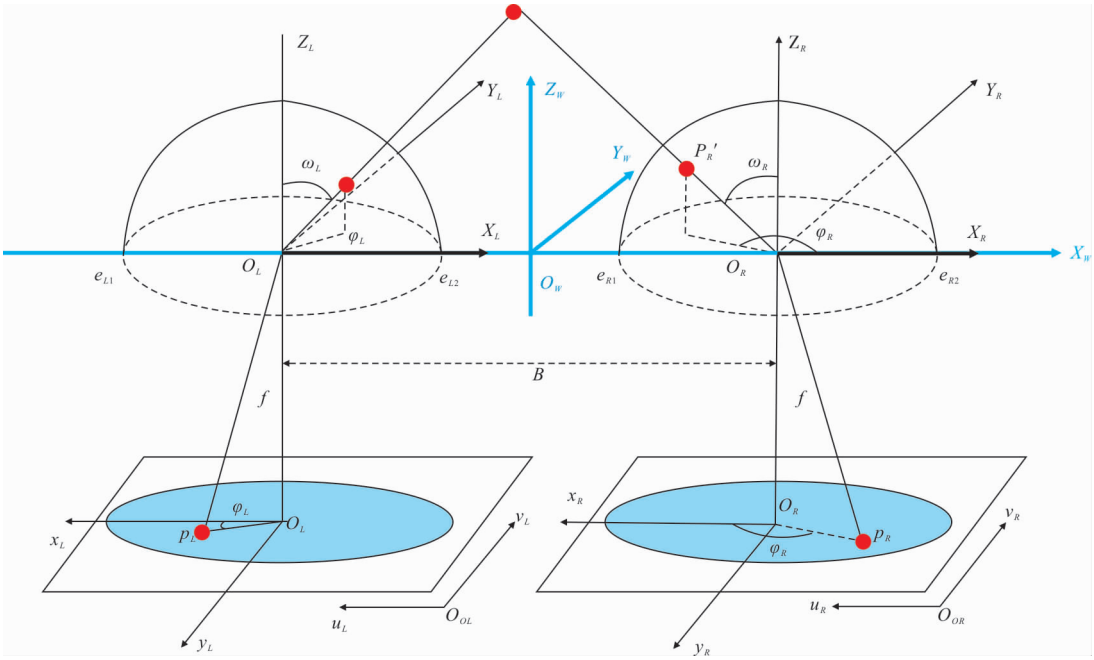

Fig.1 Framework of non-similar stereo vision

Secondly, the spherical points P_L' and P_R' are projected at the 2D image plane according to the non-similar imaging relation of Eq. (1), and the corresponding image-plane points are $p_L(x_L, y_L)$ and $p_R(x_R, y_R)$ in the coordinate systems of $o_L x_L y_L$ and $o_R x_R y_R$. Thirdly, calculate the 3D coordinates of space point P by using non-similar imaging theory.

In Fig.1, the angles ω_L, ω_R, φ_L and φ_R satisfy:

$$\begin{cases} \tan\omega_L = \dfrac{\sqrt{(X_W - B/2)^2 + Y_W^2}}{Z_W} \\ \tan\omega_R = \dfrac{\sqrt{(X_W - B/2)^2 + Y_W^2}}{Z_W} \\ \tan\varphi_L = \dfrac{y_L}{x_L} = \dfrac{Y_W}{X_W - B/2} \\ \tan\varphi_R = \dfrac{y_L}{x_L} = \dfrac{Y_W}{X_W - B/2} \end{cases} \tag{2}$$

By utilizing Eq.(1), the image-plane coordinates of (x_L, y_L) and (x_R, y_R) can be expressed as:

$$\begin{cases} x_L = f\omega_L \cos\varphi_L \quad y_L = f\omega_L \sin\varphi_L \\ x_R = f\omega_R \cos\varphi_R \quad y_R = f\omega_R \sin\varphi_R \end{cases} \tag{3}$$

By combining Eq.(2) and Eq.(3), the 3D location model of non-similar stereo vision can be deduced, which is:

$$\begin{cases} X_W = \dfrac{x_R/y_R + x_L/y_L}{x_R/y_R - x_L/y_L}\dfrac{B}{2} \\ Y_W = \dfrac{1}{x_R/y_R - x_L/y_L}B \\ Z_W = \dfrac{\sqrt{(x_R/y_R)^2 + 1}}{(x_R/y_R - x_L/y_L)\tan\left(\sqrt{x_R^2 + y_R^2}/f\right)}B \end{cases} \tag{4}$$

The equidistant projection of Eq.(1) describes the non-similar relationship between the object and its image, and the bigger the FOV is, the more serious the imaging distortion is. In the above derivation, the 3D location model of non-similar stereo vision is deduced from the non-similar imaging equation of Eq. (1), and the distortion characteristic is considered in Eq. (4). Thus, there is no need to worry about the problem of distortion correction.

1.2 Depth precision model

In Eq. (4), Z_W is the target depth, and it is relatively easy to obtain in the experiment. The depth precision can be used to analyze the model precision of Eq. (4). For a target with the same height Y_W equal to that of non-similar stereo vision system, the target depth Z_W can be

expressed as:

$$Z_W=\frac{B}{\tan\left(\sqrt{x_R^2+y_R^2}/f\right)-\tan\left(\sqrt{x_L^2+y_L^2}/f\right)} \tag{5}$$

Here, the expression of $\tan\left(\sqrt{x_R^2+y_R^2}/f\right)-\tan\left(\sqrt{x_L^2+y_L^2}/f\right)$ is defined as the non-similar disparity, and it can be seen that: the target depth is inversely proportional to the non-similar disparity, but varies nonlinearly along with the image-plane coordinates. From Eq.(5), the target depth is a synthetic function about the image-plane coordinates, the imaging focal length and baseline distance, and it can be written as

$$Z_W=F(X_R,X_L,Y_R,Y_L,B,f) \tag{6}$$

Thus, the target depth precision can be derived from Eq.(3), which is:

$$\Delta Z_W=\sqrt{\sum_i\left(\frac{\partial F}{\partial i}\right)^2} \tag{7}$$

Where, $i=X_R,X_L,Y_R,Y_L$, stands for the influence factor of each error; $\frac{\partial F}{\partial i}$ is the error transfer function. Thus, four error transfer functions can be deduced by differential operation, which are:

$$\frac{\partial Z_W}{\partial y_L}=\frac{B}{\left(\tan\frac{\sqrt{x_R^2+y_R^2}}{f}-\tan\frac{\sqrt{x_L^2+y_L^2}}{f}\right)}\cdot \sec^2\frac{\sqrt{x_L^2+y_L^2}}{f}\cdot\frac{y_L\left(x_L^2+y_L^2\right)^{-\frac{1}{2}}}{f} \tag{8}$$

$$\frac{\partial Z_W}{\partial x_L}=\frac{B}{\left(\tan\frac{\sqrt{x_R^2+y_R^2}}{f}-\tan\frac{\sqrt{x_L^2+y_L^2}}{f}\right)}\cdot \sec^2\frac{\sqrt{x_L^2+y_L^2}}{f}\cdot\frac{x_L\left(x_L^2+y_L^2\right)^{-\frac{1}{2}}}{f} \tag{9}$$

$$\frac{\partial Z_W}{\partial y_R}=\frac{B}{\left(\tan\frac{\sqrt{x_R^2+y_R^2}}{f}-\tan\frac{\sqrt{x_L^2+y_L^2}}{f}\right)}\cdot \sec^2\frac{\sqrt{x_R^2+y_R^2}}{f}\cdot\frac{y_R\left(x_R^2+y_R^2\right)^{-\frac{1}{2}}}{f} \tag{10}$$

$$\frac{\partial Z_W}{\partial x_R}=\frac{B}{\left(\tan\frac{\sqrt{x_R^2+y_R^2}}{f}-\tan\frac{\sqrt{x_L^2+y_L^2}}{f}\right)}\cdot \sec^2\frac{\sqrt{x_R^2+y_R^2}}{f}\cdot\frac{x_R\left(x_R^2+y_R^2\right)^{-\frac{1}{2}}}{f} \tag{11}$$

Inserting Eqs.(8)-(11) into Eq.(7), the target depth precision can be obtained, which is:

$$\Delta Z_W=k\sqrt{\left(\frac{\partial Z_W}{\partial x_L}\right)^2+\left(\frac{\partial Z_W}{\partial x_R}\right)^2+\left(\frac{\partial Z_W}{\partial y_L}\right)^2+\left(\frac{\partial Z_W}{\partial y_R}\right)^2} \tag{12}$$

Where, k is a subpixel correction coefficient ($0<k\leqslant 1$), and it is determined by the target localization accuracy.

2 Experiment and analysis

The experiment is carried out by utilizing a non-similar infrared stereo vision system whose horizontal, vertical and diagonal FOVs are 115°, 90° and 145°, respectively, as shown in Fig.2.

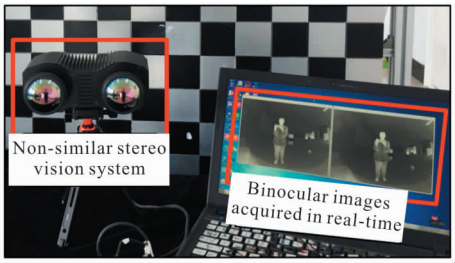

Fig.2 Non-similar stereo vision system in the experiment

The SIFT descriptor is one of the most successful and popular local image descriptor [14], and thus the SIFT descriptor is improved to match the feature points of binocular images by inducing the non-similar distortion feature. Fig.3 gives the matching effect of binocular images, and the matching ratio can reach 96.2% through multiple

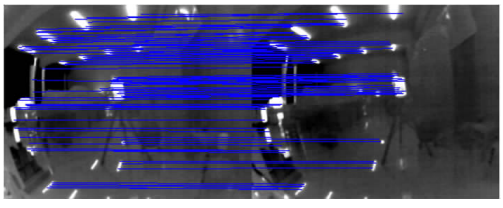

Fig.3 Stereo matching effect by improved SIFT algorithm

image matching verification. It illustrates the improved SIFT algorithm has a better applicability for non-similar infrared image.

Taking the person as the cooperative target, the target scenes are captured, and the targets of different depths(i.e. 5, 9, 10, 15, 16, 20, 30 m) are marked with rectangle boxes, as shown in Fig.4. In Fig.4, the left column images and middle column images are obtained by the left and right vision units, respectively, and the right column images are the corresponding disparity maps.

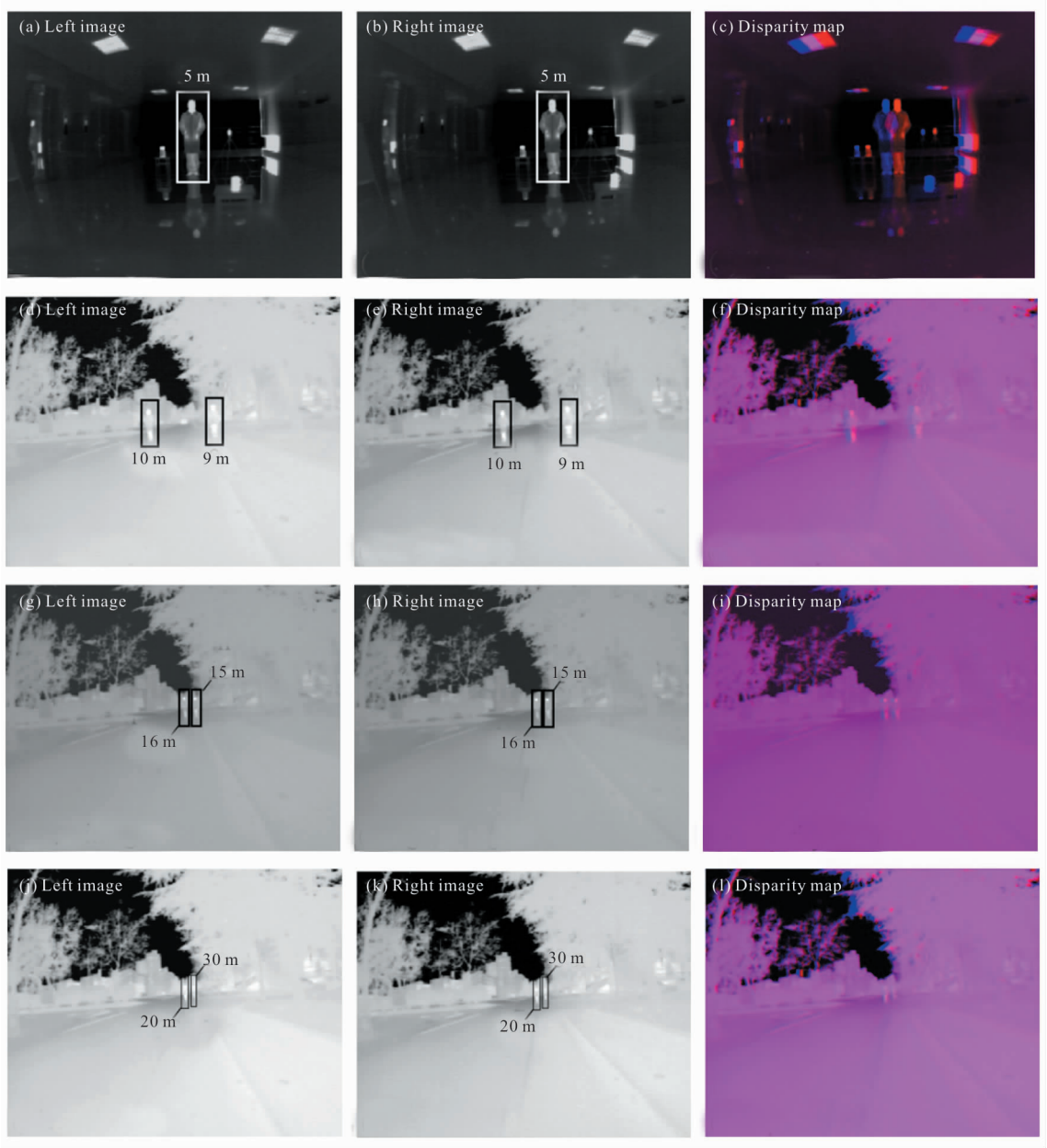

Fig.4 Captured target images in different target depths

In order to analyze the disparity more intuitively, the relative disparity curve of non-similar stereo vision is plotted from Fig.4, as shown in Fig.5, and it describes the relationship between the disparity and FOV. It can be seen: the disparity varying with the half-field angle ω is obviously different with the constant disparity of common similar stereo vision; the non-similar disparity curve changes smoothly when the full FOV 2ω is less than 30°, the curve declines sharply in the interval of $30°<2\omega\leqslant 55°$, and then

the curve declines slowly in the interval of $55°< 2\omega \leqslant 120°$; the disparity value of edge field is about 20% of central field which results from the non-similar compressive imaging.

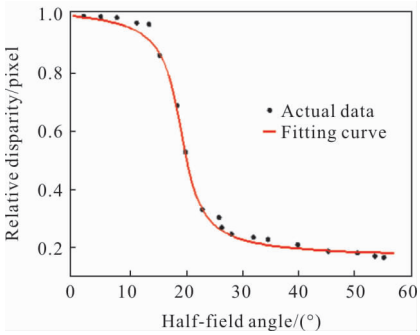

Fig.5 Relative disparity curve vs FOV

Using the improved SIFT algorithm to extract image-plane coordinates of targets, the coordinate pairs (x_L, y_L) and (x_R, y_R) of the same targets are obtained and inserted into Eq.(4), and then the 3D coordinates of target are calculated, as shown in Fig.6. In Fig.6, the 3D coordinates of space targets are displayed, and the multiple matching points of same target are marked with ellipse box while the mean depth is given around the ellipse box. It can be seen that the seven targets in different space positions are located successfully.

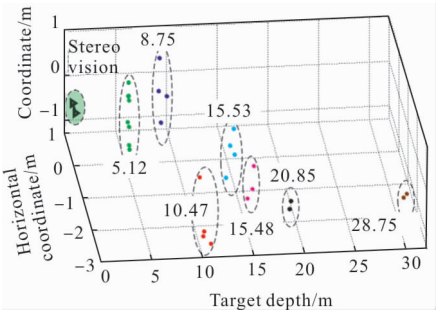

Fig.6 Target 3D coordinates calculated by proposed model

Since the target depths in the experiment are known, the results can be verified by comparing the calculated target depth with the actual target depth. Meanwhile, the similar stereo vision model can be used to calculate the target depth in order to further check the effect of the proposed model. In the similar stereo vision model, the target depth is usually expressed as[15]:

$$Z_W = \frac{f}{x_L - x_R} B \tag{13}$$

By applying the coordinates of matching points in Fig.3 into Eq.(5) and Eq.(12), the target depths of two models are calculated, as shown in Fig.7. The depth error curves which are derived from 3D location model of Eq. (4), traditional model of Eq. (13) and depth precision model of Eq.(12), are displayed in Fig.8.

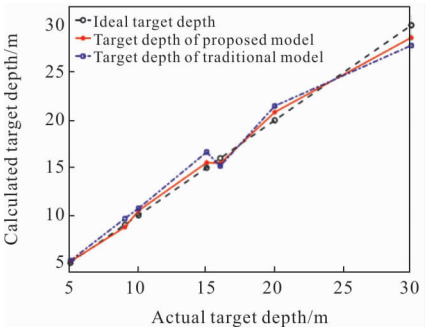

Fig.7 Curves of calculated target depths

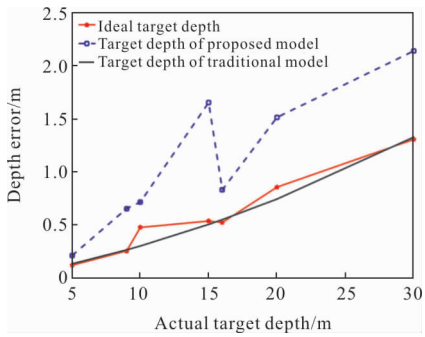

Fig.8 Error curves of calculated target depths

It can be seen from Fig.7 and Fig.8 that:

(1) The results of the non-similar stereo vision are closer to the actual values compared with that of the similar stereo vision, and the depth error of the proposed 3D location model is obviously less than that of the traditional similar model.

(2) The depth error from proposed model is basically less than 1.5 m within the target depth of 30 m, and is around 42.3% of the error from Eq. (13). In addition, the absolute value of depth error increases monotonically with the increasing

of the target depth, which satisfies the error law of stereo vision.

(3) The actual depth error of non-similar stereo vision model matches well with the results from the depth precision model of Eq.(12), and this verifies the correctness of proposed models again.

3 Conclusion

The concept of non-similar stereo vision is presented by introducing the equidistant projection theory into large-airspace stereo vision research, and the 3D location model and depth precision model are established. Through carrying out contrast experiment, the 3D coordinates of cooperative targets are obtained with a higher accuracy, and the location error is less than 1.5 m in 30 m depth, which is far less than that obtained from similar stereo vision model, and the error curve fits well with the results of depth precision model. The non-similar stereo vision can not only obtain larger airspace, but also locate the position of space target with a passive way. This research on non-similar stereo vision can enrich and improve the theory of stereo vision, and will help to expand the application of non-similar imaging theory in stereo vision fields, such as intelligent transportation, automatic driving, augmented reality, military reconnaissance and so on.

References:

[1] Jin Tao, Jia Hongzhi, Hou Wenmei, et al. Evaluating 3D position and velocity of subject in parabolic flight experiment by use of the binocular stereo vision measurement [J]. ***Chinese Optics Letters***, 2010, 8(6): 601-605.

[2] He Yu, Wang Lingxue, Cai Yi, et al. Research progress and prospect of catadioptric panoramic system [J]. ***Chinese Optics***, 2017, 10(5): 681-698. (in Chinese)

[3] Ni Zhangsong, Gu Yi, Liu Qingling, et al. Flexible calibration method for binocular stereo vision in large field of view [J]. ***Optics and Precision Engineering***, 2017, 25(7): 1882-1889. (in Chinese)

[4] Yang Ning, Shen Jingshi, Zhang Jiande, et al. Autonomous measurement of relative attitude and position for spatial non-cooperative spacecraft based on stereo vision [J]. ***Optics and Precision Engineering***, 2017, 25(5): 1331-1339. (in Chinese)

[5] Wang Xin, Wang Xiangjun. Multiple targets sparse matching for binocular vision positioning system with large field of view [J]. ***Infrared and Laser Engineering***, 2018, 47(7): 0726001. (in Chinese)

[6] Shi Qing, Li Chang, Wang Chunbao, et al. Design and implementation of an omnidirectional vision system for robot perception[J]. ***Mechatronics***, 2017, 41: 58-66.

[7] Qi Bingjie, Liu Jinguo, Zhang Boyan, et al. Research on matching performance of SIFT and SURF algorithms for high resolution remote sensing image [J]. ***Chinese Optics***, 2017, 10(3): 331-339. (in Chinese)

[8] Ameer H, Rehan H, Muhammad M K, et al. Stabilization of panoramic videos from mobile multi-camera platforms [J]. ***Image and Vision Computing***, 2015, 37: 20-30.

[9] Pedersen L, Bualat M, Kunz C, et al. Instrument deployment for Mars rovers [C]//Proc of IEEE, International Conference on Robotics & Automation, 2003, 2: 2535-2542.

[10] Wang Baofeng, Zhou Jianliang, Tang Geshi, et al. Research on visual localization method of lunar rover [J]. ***Sci Sinica Inform***, 2014, 44(4): 452-460.

[11] Kim D, Choi J, Yoo H, et al. Rear obstacle detection system with fisheye stereo camera using HCT[J]. ***Expert Systems with Applications***, 2015, 42 (17-18): 6295-6305.

[12] Johannes S, Cyrill S, Wolfgang F. On the accuracy of dense fisheye stereo[J]. ***IEEE Robotics and Autonation Letters***, 2016, 1(1): 227-234.

[13] Huang Fuyu, Shen Xueju, He Yongqiang, et al. Performance analysis of super-wide field of view imaging system used for space target detection [J]. ***Infrared and Laser Engineering***, 2015, 44 (10): 3135-3140.

[14] Wang Zhiqiang, Cheng Hong, Tan Haifeng, et al. Registration algorithm of aerial remote sensing images based on lateral inhibition competition[J]. ***Infrared and Laser Engineering***, 2018, 47(S1): S126005.

[15] Akhil V, Abdul B, Vishak P V. Stereo vision system implemented on FPGA [J]. ***Procedia Technology***, 2016, 24: 1105-1112.