Abstract—Stereo-vision plays an important part in planetary exploration. A high-precision calibration method is introduced for such system, which uses wide-angle lens. It makes use of planar homography constraint to estimate intrinsic and extrinsic parameters. On the basis of this initial guess, an optimization scheme is used to minimize a new cost function, 3D reprojection error sum. All Calibration parameters of two cameras are globally optimized simultaneously with Genetic Algorithm (GA). Both simulation and real image experiment results show this scheme have much higher precision than traditional methods.

Keywords—camera calibration, stereo vision, optimization

I. INTRODUCTION

Stereo vision plays an important role in planetary exploration, for it can percept and measure the 3-D information of the unstructured environment in a passive manner. It can provide consultant support for robotics control and decision-making. So it is applied in the field of rover navigation, real-time hazard avoidance, path programming and terrain modeling. In some cases, one stereo-vision system must accomplish both hazard detection and accurate localization with short baseline, i.e. 100-200mm in length. This seems to be a little ambivalent, for hazard detection needs wide view field, while accurate localization is on the contrary. Reconstruction precision is inverse proportion to focal length if the baseline is fixed. So researchers have to first select a compatible view angle, which guarantees the task workspace is within the view field. Then they must refine their camera calibration method in order to satisfy the accuracy requirement of rover localization, navigation and task operation.

In order to satisfy these requirements, wide angle lens is usually used. Lens distortion may reduce the precision of localization. So distortion parameter calibration plays an important part in such case. Moreover, calibration accuracy may also affect the complexity of the matching process. Tsai [2] proposed a method, in which a distorted parameter is used to describe the radial distortion of the lens. In [3][4], a five-parameter model is exploited to characterize several kinds of lens distortion. In [5], a more complicated model, CAHVORE, is introduced. Calibration becomes a nonlinear process if lens distortion is introduced. Usually camera calibration needs two steps. The first step generates an approximate solution using a linear technique, while the second step refines the linear solution using a nonlinear iterative procedure. The approximate solutions provided by the linear techniques must be good enough for the subsequent nonlinear technique to correctly converge. After the initial value has been obtained, the precision of the final result and convergence speed depends closely on optimization algorithm. Most of existing nonlinear methods minimize the geometric cost function using variants of conventional optimization techniques like gradient-descent, conjugate gradient descent Newton or Levenberg-Marquardt (LM) method. Therefore there are some problems in these circumstances. First, it is the commonly used cost function, reprojection error, which minimizes the distance between the measured image points and estimated image points. The points in each image are subject to noise, while the refined solution is only optimal to measured 2D image points, not to the real 3D points. So the final solution can inevitably be contaminated with large error, especially in depth direction when this refined solution is used for 3D reconstruction. Secondly, not all the parameters are optimized simultaneously. In most cases only part of the parameters are assumed to need refining while others are assumed to be correct and keep constant in optimization process, just like [2], which may result in the parameter not globally optimized. The third is the above techniques inherit well-known problems plaguing these differential-based methods, i.e. poor convergence and susceptibility to be trapped in local minimum. This is especially true for the objective function of camera calibration involves too many camera parameters and leads to a complex error surface. So the risk of local rather than global optimization might be severe with conventional methods.

To alleviate the problems in the existing calibration techniques, we develop an alternative paradigm based on a new cost function to conventional reprojection error cost function. And we try Genetic Algorithms (GA) in the searching process instead of differential method in order to get globally optimal solution in high-dimension parameter space and avoid trapping in local minimum.

This paper is organized as follow. In section 2, the camera model of stereo-vision system is proposed. Section 3 introduces the calibration method, which is based on planar homography constraint to get the initial solution. Section 4 gives the optimization strategy, including Reconstruction Error Sum, a new cost function, and GA searching process. Simulation and
real image experiment results are given in section 5. The article is concluded in section 6.

II. CAMERA MODEL

The finite projective camera, which often has pinhole model, is used in this paper just like [4]. As figure 1 shows, left and right cameras have intrinsic parameter matrices \( K_q \):

\[
K_q = \begin{bmatrix}
    k_{uq} & s_q & u_{0q} \\
    0 & k_{vq} & v_{0q} \\
    0 & 0 & 1
\end{bmatrix}, q = 1, 2
\] (1)

The subscript \( q = 1, 2 \) denotes left and right camera respectively. If the number of pixels per unit distance in the image coordinates are \( m_x \) and \( m_y \) in the x and y directions, \( f \) is the focal of length, \( k_{uq} = f m_x \) and \( k_{vq} = f m_y \) represent the focal length and right cameras have intrinsic parameter matrixes \( K_q \).

The model plane is observed in several positions, just like [6] introduced. At the beginning of calibration, image distortion is not considered. And the relationship between the 3D point \( P \) and its pixel projection \( p_q \) is:

\[
\lambda_q \tilde{p}_q = H_q P \quad \text{with} \quad H_q = K_q \begin{bmatrix} r_{1q} & r_{2q} & t_q \end{bmatrix} \] (8)

Where \( \lambda_q \) is an arbitrary factor. We assume the model plane is on \( Z = 0 \) of the world coordinate system. Then (6) can be changed into:

\[
\lambda_q \tilde{p}_q \equiv H_q P \quad \text{with} \quad H_q = K_q \begin{bmatrix} r_{1q} & r_{2q} & t_q \end{bmatrix} \] (8)

Here \( r_{1q} \) and \( r_{2q} \) are the first two columns of rotation matrix of two cameras, and \( H_q \) is the planar homography between two planes. If more than four pairs of corresponding points are known, \( H_q \) can be computed. Then we can use orthonormal constraint of \( r_{1q} \) and \( r_{2q} \) to get the closed-form solution of intrinsic matrix. Once \( K_q \) is estimated, the extrinsic parameters \( R_q, t_q \) and the scale factor \( \lambda_q \) for each image plane can be easily computed, as [6] indicated.

IV. OPTIMIZATION SCHEME

The above solution is obtained through minimizing the algebraic distance, which is not physically meaningful. The commonly used optimization scheme is based on maximum likelihood estimation:

\[
\sum_{i=1}^{n} \sum_{j=1}^{m} \left| p_{ij} - \hat{p}(K_q, k_{1q}, \ldots, k_{5q}, R_{iq}, t_{iq}, P_j) \right|^2
\] (9)

Where \( \hat{p}(K_q, k_{1q}, \ldots, k_{5q}, R_{iq}, t_{iq}, P_j) \) is the estimated projection of point \( P_j \) in image i, followed by distortion according to (3) and (4). The minimizing process is often solved with LM Algorithm. However, (8) is not accurate enough if it is used for localization and 3D reconstruction. The reason is just like section 1 described. Moreover, there are too many parameters to be estimated, namely, five intrinsic parameters, and five distorted parameters plus 6n extrinsic parameters for each camera. Each group of extrinsic parameter
might be only optimized for the points on the current plane, while it may deviate too much from its real value. So a new cost function is explored here, which is on the basis of Reconstruction Error Sum (RES).

A. Cost Function

Although the cost function using reprojection error is equivalent to maximum likelihood estimation, it has defect in recovering depth information, for it iteratively adjusts the estimated parameters to make the estimated image point approach the measured point as closely as possible. While for 3D points, it may be not. We use Reconstruction Error Sum (RES) as cost function:

\[
RES(b) = \sum_{i=1}^{n} \sum_{j=1}^{m} \left[ P_j - \prod (p_{y1,j}, p_{y2,j}, b_1, b_2) \right]^2 \quad (10)
\]

Where \( P_j \) is a 3D point in the world frame. Its estimated 3D coordinate can be denoted as: \( \prod (p_{y1,j}, p_{y2,j}, b_1, b_2) \), which is reconstructed through triangulation method with given camera parameters \( b_1, b_2 \) and image projections \( p_{ij1}, p_{ij2} \). \( b \) is a vector consisting 32 calibration parameters of both left and right cameras, including extrinsic, intrinsic and lens distortion described in (1), (2), (4), (5):

\[
b = \{b_1, b_2\} \quad (11)
\]

It is very important because appropriate searching scope can reduce computational complexity. The chromosome is generated randomly in the region near the initial value. The fitness function we chose here is (10). The whole population consists of \( M \) individuals, where \( M = 200 \). The full description of GA is below:

- **Initialization**: Generate \( M \) individuals randomly. Suppose the generation number \( t = 0 \), i.e.:
  \[
  G^0 = \{ b_1^0, \ldots, b_M^0 \}
  \]
  Where \( b \) is chromosome. Superscript is generation number. And subscript denotes individual number.

- **Fitness Computation**: Compute fitness value of each chromosome according to optimal selection and random selection.
  \[
  G^{t+1} = \{ b_1^{t+1}, \ldots, b_M^{t+1} \}
  \]
  \[
  \text{Fitness Computation: Compute fitness value of each chromosome.}
  \]

- **Selection operation**: Select \( k \) individuals according to optimal and random selection.
  \[
  G^t = \{ b_1^t, \ldots, b_M^t \}
  \]
  \[
  F(b_j^t) \leq F(b_j^{t+1})
  \]

- **Mutation operation**: Select \( p \) individuals from the new \( k \) individuals, and mutate part of genes randomly.
  \[
  G^{t+1} = \{ b_1^{t+1}, \ldots, b_k^{t+1}, b_k^{t+1}, \ldots, b_k^{t+1} \}
  \]

- **Crossover operation**: Perform crossover operation. Select \( I \) genes for crossover randomly. Repeat it \( M-K-p \) times.
  \[
  G^{t+1} = \{ b_1^{t+1}, \ldots, b_k^{t+1}, \ldots, b_k^{t+1} \}
  \]

- **Let** \( t = t+1 \). Select the best chromosome as current solution:
  \[
  b_{best} = \{ b_j^t \} \quad \text{if} \quad F(b_j^t) = \min_{j=1}^{M} (F(b_j^t)) \quad \text{if termination conditions are satisfied, i.e. } t \text{ is bigger than a predefined number or } F(b_{best}) < \varepsilon , \text{ search process will end.} \]

Otherwise, goto step 2.

V. EXPERIMENT RESULT

Both simulation and real image experiments have been done to verify the proposed method. Both left and right simulated cameras have the following parameter: \( k_{u1} = k_{v1} = 540, s_q = 0, u_{01} = 400, v_{01} = 300, q = 1,2 \). The length of the baseline is 200mm. World frame is bound at the midpoint on the line connecting the two optic centers. Rotation and translation between two frames are pre-defined. The distortion parameters of the two cameras are given. Some emulated points in 3D world, whose distances to the image center are about 1m, project on the image planes. These image points are added with Gaussian noise of different level. With these image projections and 3D points, we calibrate both emulation cameras with three different methods, Tsai method [2], Matlab method [4], and our scheme. A normalized error function is defined as:
\[
E(b) = \frac{1}{n} \sum_{i=1}^{n} (1 - \hat{h}_i / \bar{h}_i)^2
\]  

(12)

It is used to measure the distance between estimated cameras parameters and true cameras parameters so as to compare the performance of each method. Where \(\hat{h}_i, \bar{h}_i\) are the \(i^{th}\) element estimated and real values of (11) respectively, and \(n\) is the parameter number of each method. The performances of three methods are compared, and the results are shown in Table 1, where RES is our method. 1/8, 1/4, and 1/2 pixel noise is added in image points to verify the robustness of each method. From Table 1, it can be seen our method has higher precision and better robustness than Tsai and Matlab methods.

| Error | Scheme | Tsai | Matlab | RES |
|-------|--------|------|--------|-----|
| 1/8 pixel | 1.092 | 1.245 | 0.7094 |
| 1/4 pixel | 1.319 | 1.597 | 0.9420 |
| 1/2 pixel | 2.543 | 3.001 | 1.416 |

Table I. Normalized Error Comparison

Real image experiment is also performed on the 3D platform, which can translate in X, Y, Z direction with 1mm precision. The cameras used are IMPERX 2M30, which are working in the binning mode with 800×600 resolution, together with 4mm-focal-length lens. The length of baseline is 200mm. A calibration chessboard is fixed rigidly on this platform about 1m away from the camera. About 40 images, which are shown in figure 2, are taken every ten-centimeter on left, middle and right side of view field along depth direction. First we use all the corner points as control points for coarse calibration. Then 4 points of each image, altogether about 160 points are selected for optimization with (10). The rest 7000 points are used for verification. We use Pentium 1.7GHz CPU, and VC++ 6.0 developing environment, calibration process needs about 30 minutes. Calibration result obtained from Tsai method, Matlab toolbox and our scheme, are used to reconstruct these points. Error distribution histogram is shown in figure 3, in which top is the Tsai method, middle is Matlab scheme, and bottom is our method. The unit of horizontal axis is millimeter. Table 2 shows statistic reconstruction errors along X, Y, Z direction, including mean error \(A(X), A(Y), A(Z)\), maximal error \(M(X), M(Y), M(Z)\), and variance \(\sigma_x, \sigma_y, \sigma_z\). From these figure and table, it can be seen our scheme can have much higher precision than other method, especially in depth direction.

| Error | Scheme | Tsai | Matlab | RES |
|-------|--------|------|--------|-----|
| \(A(X)\) | 2.3966 | 3.4453 | 1.7356 |
| \(A(Y)\) | 2.1967 | 2.2144 | 1.6104 |
| \(A(Z)\) | 4.2987 | 5.2509 | 2.3022 |
| \(M(X)\) | 9.5756 | 13.6049 | 5.7339 |
| \(M(Y)\) | 9.8872 | 12.5877 | 7.3762 |
| \(M(Z)\) | 15.1088 | 19.1929 | 7.3939 |
| \(\sigma_x\) | 2.4499 | 2.7604 | 1.7741 |
| \(\sigma_y\) | 2.3873 | 3.0375 | 1.8755 |
| \(\sigma_z\) | 4.7211 | 4.8903 | 2.4063 |

Table II. Statistic Error Comparison

VI. CONCLUSION

In this paper, a high precision camera calibration method is proposed for stereo vision system using wide angle lens. It exploits 5 parameters to describe lens distortion. Genetic algorithm is used in searching process. Simulation and real image experiment show that this scheme has higher precision and better robustness than traditional method for space localization.

REFERENCES

[1] Murry D, Jennings C, “Stereo vision based mapping and navigation for mobile robots”, In: Proceedings of IEEE Conference on Robotics and Automation, 1997, pp.1694-1699.
[2] R. Y. Tsai, “A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the shelf tv cameras and lenses,” IEEE Journal of Robotics and Automation, vol.3, no.4, 1987, pp.323-344.

[3] J. Yunde, L. Hongjing, X. An, L. Wanchun, “Fish-Eye Lens Camera Calibration for Stereo Vision System”, Chinese Journal of Computers, vol.23, no.11, 2000, pp.1215-1219.

[4] Camera calibration toolbox for matlab, web site : http://www.vision.caltech.edu/bouguetj/calib_doc/.

[5] D. B. Gennery, “Least-squares camera calibration including lens distortion and automatic editing of calibration points”, Calibration and Orientation of Cameras in Computer Vision, T. Huang, Springer-Verlag, New York, 2001.

[6] Z. Zhang, “A flexible new technique for camera calibration,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.22, no.11, 2000, pp.1330-1334.

[7] M. L. Gong, and Y. H. Yang, “Genetic-based stereo algorithm and disparity map evaluation,” International Journal of Computer Vision, vol.47, no.1, 2002, pp.63-77.

[8] Zhou, Chuan, Tan, Da long, Zhu, Feng, “A High-Precision Calibration Method for Distorted Camera.” IROS 2004, Sep 28-Oct 2, Sendai, Japan, 2004, pp.2618-2623.