Stability Study of Camera Calibration Methods

Conference Paper · September 2005

4 authors, including:

Some of the authors of this publication are also working on these related projects:

PLANCLIMAC (MAC2/3.5b/244): PLANIFICACIÓN CONJUNTA, SEGUIMIENTO Y OBSERVACIÓN, MEJORA DEL CONOCIMIENTO Y SENSIBILIZACIÓN ANTE RIESGOS Y AMENAZAS DEL CAMBIO CLIMÁTICO EN LA MACARONESIA View project

ACUSQUAT View project

Stability Study of Camera Calibration Methods

J. Isern González, J. Cabrera Gámez, C. Guerra Artal, A.M. Naranjo Cabrera

Instituto Universitario de Sistemas Inteligentes y Aplicaciones Numéricas en Ingeniería (IUSIANI)

Universidad de Las Palmas de Gran Canaria

35017 Las Palmas de Gran Canaria

{jisern, jcabrera, cguerra}@iusiani.ulpgc.es, angel@jqbit.com

Abstract

Many methods have been developed to calibrate cameras, but very few works have been done to compare such methods or to provide the user with hints on the suitability of certain algorithms under particular circumstances. This work presents a comparative analysis of eight methods of calibration for cameras using a pattern as reference. This paper concentrates on the stability and accuracy of these methods when the pattern is relocated or the camera configuration varies. This study was carried out with real and synthetic images and using, whenever possible, the code made available by the methods' authors on the WWW. The experiments demonstrate that most of these methods are not stable in the sense that the intrinsic parameters returned by a calibration method suffered important variations under small displacements of the camera relative to the calibration pattern. Similar results were obtained for extrinsic parameters when the camera only changed its internal configuration (i.e. when it zooms in or out) but kept constant its relative position to the calibration pattern. In addition, this study shows that the image disparity is not an indicator of the reliability of the methods. In spite of the fact that the majority of the methods showed similar levels of global error, the calibrated values obtained for intrinsic and extrinsic parameters varied substantially among these methods for the same set of calibration images.

1 Introduction

In the field of computer vision, camera calibration is a procedure that tries to know how a camera projects an object on the screen. That is to say, the internal geometry of the camera and its optic features, as well as the translation and rotation of the reference object. This process is necessary in those applications where metric information of the environment must be derived from images. Some of these tasks are: making maps of the camera environment, tracking of objects, virtual reconstruction of these, etc. Also, if the camera is installed on a moving platform, then its position with respect to objects that are around it can be known. This allows a robot to move in its environment avoiding obstacles, heading to a specific object, or making the definition of the best trajectory to reach its destination easy.

This work has been dictated by practical needs. When a camera needed to be calibrated, it was observed that calibration results differed and also that different methods gave different results. The natural question was then whether there was a better calibration method (in some sense) than the others. The bibliography shows that many calibration methods have been proposed during the last few decades, and all of them, according to their authors, have a high accuracy. But, there are very few works where calibration methods have been studied impartially.

Tsai [10] gives some criteria that should be analyzed to select a suitable method: velocity, accuracy and autonomy. These were

used to compare theoretically a set of methods published before 1987. The difference in notation used by different authors is an additional difficulty when the appropriate method is searched. In [8], some methods, all of them published before 1994, are described in detail using a unified notation. These methods are compared in one experiment where only the distance between real points and reconstructed points (2D and 3D points) is analyzed. An aspect of the calibration problem that has been neglected by most surveys is the accuracy provided by a method with respect to the model parameters, due to the fact that usually only the accuracy in reconstruction is studied. One exception is Lai [7] who studies, from a theoretical and experimental point of view, how extrinsic parameters are affected by errors on some intrinsic parameters. This author shows that lens distortion and optic center deviation have little effect on the estimation of the extrinsic parameters while - on the other hand - the scale factor has a great influence. Another comparative work is [5] where five methods, proposed before 1991, are compared. In this work the precision achieved by each method could not be evaluated on ground basis because the experiments were carried out on images acquired with a camera whose setup was unknown. The present work is an extension of [4] in some specific aspects.

When this study started the first question that we needed to address was: Which methods should be analyzed? After reviewing the relevant literature, we decided to select a set of eight methods. Some of them were included for comparison because they were often used as a point of reference in the literature of this field, as is the case of Tsai's [9] and Faugeras' [3] methods. Others were selected due to their special type of computation: the use of non-iterative algorithms (Linear method), or neural networks (Ahmed's method [1]); Others by details of the camera model used, i.e. including high lens distortion (Heikkilä's method [6]), or optic center (optimized Tsai's method [8]); And, finally, others because of the type of pattern used for calibration: the majority of the methods need to use a pattern with two planes, knowing the relationship between them, usually a dihedral pattern, therefore methods that did not need to keep this restriction, like Batista's [2] and Zhang's [11] methods, were selected).

The experiments were carried out with real camera and with a simulator. The results provided by every method were compared and the accuracy in the pattern points reconstruction and the stability of the model parameters were analyzed. We have defined the stability of a calibration as the constancy of the results when one parameter or factor changes.

This paper is organized as follows. Next section deals with camera modelling. Section 3 briefly describes every method that has been used in the comparative. Some of the experiments are presented in section 4 and their results are analyzed. The last section is devoted to summary the main conclusions of this research.

2   Camera model

All the methods described in this paper follow the pin-hole camera model (figure 1). The parameters obtained from the calibration process can be classified as two types: *intrinsics*, those that define the camera configuration (internal geometry and optic features); and *extrinsics*, the rotation and the translation that would map the world coordinate system to the camera coordinate system.
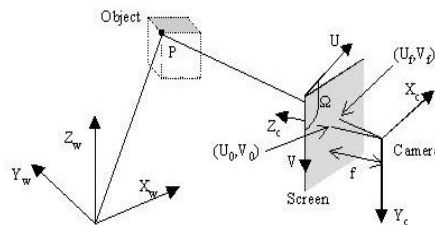

Figure 1: Camera model. Points projection on the image plane.

Calibration methods usually break down the transformation from 3D coordinates to 2D coordinates in the following four steps:

1. Transformation from 3D coordinates of world system (M_w) to camera system (M_c). In these methods, the rotation matrix is obtained applying successive turns around each axis followed by a translation, $M_c = R \cdot M_w + T$, where R is the rotation matrix, T is the translation array and $M = [X, Y, Z]^t$ are the 3D coordinates..

2. Projection on screen plane using the pinhole model.

$$u_p = f\frac{X_c}{Z_c} \qquad v_p = f\frac{Y_c}{Z_c} \qquad (1)$$

where f is the focal length.

3. Lens distortion. Two types of distortion, which depend on non-linear factors, are modelled: radial (u_r, v_r) and tangential (u_t, v_t):

$$\begin{aligned} u_d &= u_p + u_r + u_t \\ v_d &= v_p + v_r + v_t \end{aligned} \qquad (2)$$

4. Metric to screen units transformation:

$$\begin{aligned} u_f &= d_x s_x (u_d - \cot\Omega \cdot v_d) + u_0 \\ v_f &= d_y \frac{v_d}{\sin\Omega} + u_0 \end{aligned} \qquad (3)$$

where d_x and d_y are transformation factors, s_x is the pixel aspect ratio, and the optical center position is represented by u_0 and v_0. Some methods model the non-orthogonality between the image plane and the optic axis as a skew angle (Ω). Batista's method uses two angles to model this skew.

It is possible to simplify the formulation using a matrix notation: $\lambda \cdot m = P \cdot M$ where $M = [X_w, Y_w, Z_w, 1]^t$, $m = [u, v, 1]^t$, and λ is a non-zero scale factor. Finally, the 3x4 projection matrix can be split into two matrices: $P = AD$, one for extrinsic parameters ($D = [R|T]$) and the another for intrinsic parameters:

$$A = \begin{bmatrix} k_x & -k_x \cot\Omega & u_0 \\ 0 & k_y/\sin\Omega & v_0 \\ 0 & 0 & 1 \end{bmatrix} \qquad (4)$$

where $k_x = s_x d_x f$ and $k_y = d_y f$ are horizontal and vertical scale factors.

3 Calibration methods

The methods that have been studied in this work can be classified according to a number of criteria: computation complexity (iterative vs. non-iterative), resolution scheme (one single optimization step vs. multistep) or type of calibration pattern used (planar vs. non-planar or single view vs. multiple views). The following subsections contain a brief presentation of the methods considered.

3.1 Tsai's method [9]

This method splits the problem into two steps. It allows use of both planar and non-planar patterns. This method assumes that optical center is located at image center and that skew is null.

In the first step, the distortion is assumed null and a linear approach can be used to solve a subset of parameters: orientation (R), translation (T_x, T_y) and aspect ratio (s_x). In the second step, an initial guess of focal length (f) and pattern distance (T_z) can be estimated using a linear technique. Then focal length, pattern distance and distortion coefficient (k_1) can be refined by means of an optimization technique.

A variant has been proposed in [8] that adds two additional steps to the original method. Firstly, an iterative technique is used to refine the parameters returned by the original algorithm. Then, this process is repeated but the optic center is also included. The code created by X. Armangué and J. Salvi has been used in the experiments to implement this variant of Tsai's method. This code can be retrieved from "http://eia.udg.es/%7Earmangue/research/".

3.2 Faugeras' method [3]

This method obtains a projection matrix (P) using an optimization procedure where orthogonality of rotation matrix is hold. One non-planar pattern is necessary and lens distortion is not modelled. An initial estimate is necessary for the convergence of the method. In this work, the Linear method, described in section

3.3, was used to obtain this estimate. The calibration parameters can be recovered decomposing the projection matrix.

3.3 Linear method

It is a well-known classic method aimed at minimizing image disparity error. This is obtained minimizing $\|L \cdot x\|$, where:

$$L \cdot x = \begin{bmatrix} M^t & 0^t & -u \cdot M^t \\ 0^t & M^t & -v \cdot M^t \end{bmatrix} x = 0 \qquad (5)$$

with the vector x being the rows of projection matrix. The solution is the eigenvector corresponding to the smallest eigenvalue of $L^t L$, which may be conveniently found using the Singular Value Decomposition (SVD) of L. The result is normalized and each parameter can be obtained using the decomposition of the projection matrix, as in Faugeras' method. In this method one non-planar pattern is necessary and lens distortion is not modelled.

3.4 Ahmed's method [1]

This method is based on neural networks, consequently it is very slow. Initial values obtained from the Linear method (see section 3.3) were used in this research to accelerate the calibration process. Lens distortion is not considered by the algorithm and it needs one non-planar pattern.

A two-layer feedforward neural network is used, where world points coordinates are the input and image coordinates are the output. Each level has an associated weight matrix, one for extrinsic parameters (V) and another for intrinsic parameters (W).

The training process of the neural network is repeated until error converges to a minimum. The measure of error depends on two factors: image disparity of each point and non-orthonormality of rotation matrix. Weighting matrices are updated following the descent gradient rule.

Input and output values of the net are normalized (S_1, S_2). When the method converges, a projection matrix is obtained as: $P = S_1 W V S_2$. Parameter values can be extracted from this matrix decomposing the projection matrix, using i.e. the indications of Faugeras' method.

3.5 Heikkilä's method [6]

This method consists of four steps, but in this research only the first two have been utilized because the other two are related exclusively to image acquisition. It models both radial and tangential distortion of the lens.

Initially, a linear method (DLT algorithm) is used to obtain initial estimates for some parameters (extrinsics, focal and optic center). From the projective matrix thus obtained, parameter values are extracted using RQ decomposition. Finally, the Levenberg-Marquardt method is used to minimize: $\sum (u - u_f)^2 + \sum (v - v_f)^2$ (image disparity).

The Calibration toolbox v3.0, developed by Janne Heikkilä, has been used in this work. It can be downloaded from "http://www.ee.oulu.fi/~jth/calibr/".

3.6 Batista's method [2]

This method needs only one image of a planar pattern to calibrate a camera. It is a multistep and iterative method that uses a least squares technique at each step. Distortion is modelled using only one coefficient. Initially, this method assumes that distortion and skew are null, aspect ratio is one, the optic center is coincident with image center, and transformation factors and focal length are obtained from the manufacturer.

The procedure comprises two loops: the first repeats the three early steps described below and it finishes when aspect ratio converges to one; the second repeats all the steps and it ends when the mean of image disparities goes below a threshold.

This algorithm comprises these four steps: 1) Obtaining rotation angles, assuming that translation is null and using four points forming a rectangle. 2) Translation and aspect ratio can be obtained using all pattern points. 3) Distance and distortion can be calculated using a linear approach. Focal length is resolved using the Gauss lens model. 4) The intrinsic

parameters not obtained in previous steps are only allowed to vary in the second loop.

3.7 Zhang's method [11]

This method needs at least three different views of a planar pattern. Displacements between these views can be unknown. Lens distortion is modelled using two coefficients.

This method comprises the following steps: 1) Projective matrix (P_j) is calculated from each image. An initial estimate is obtained using a non-iterative approach. Then the Levenberg-Marquardt technique is used to refine P_j where image disparity is minimized. 2) The intrinsic parameters matrix (A) is obtained from the absolute conic $(B = A^{-t}A^{-1})$, which is invariant to object displacements. Two constraints on the intrinsic parameters are imposed for each image and the equation system thus obtained is solved using SVD. 3) Computation of extrinsic parameters (D). 4) An initial estimate of radial distortion coefficients are calculated using a least squares technique. 5) All parameters are refined using an optimization procedure that minimizes mean of image disparity.

4 Experimental results

The experiments discussed in this paper have focused on characterizing the stability of the estimation of the model parameters and the accuracy in the pattern points reconstruction. These experiments were performed both with a real camera and with a simulator. Whenever possible, the original code (i.e. made available by the author(s) of the respective method) was used.

Different measures were used in the experiments to estimate the accuracy and stability of the methods. The global error (2D-Error and 3D-Error) is the estimation of point disparity. That is the difference between real coordinates and estimated coordinates of points using the results of calibration. The mean distance between ground truth value and estimated value was used to calculate the error in each parameter. We have defined the stability of a method

as the constancy of its calibration results when one parameter or factor changes. The maximum difference of the values returned by a method was used as an indicator of the stability of that method. The stability of the methods was studied from two points of view. First, the stability of intrinsic parameters was analyzed when the location of the pattern changed but the camera setup was kept constant. Conversely, the variation of extrinsic parameters was analyzed when the camera setup changed but the placement of the pattern relative to the camera was held fixed.

4.1 Experiments with a real camera

As the ground truth values of each parameter are unknown, only the analysis of stability was carried out with real data. The camera utilized in the experiments was a Sony EVI-G21, which has a motorized zoom. This camera was connected to an Imaging IC-PCI-COMP digitalizer card. Image resolution was set at 768x576 pixels and the focal length varied between 4.5 and 13.5 mm (manufacturer data) due to zoom lens. The auto-focus was disabled in all of the experiments.

4.1.1 Stability of extrinsic parameters

Four sets of views of a static pattern were used to establish the stability of extrinsic parameters. The zoom was moved between its extreme values along each sequence.

	ΔT_x (mm)	ΔT_z (mm)	ΔR_x $(°)$	ΔR_y $(°)$
Linear	7.6	39.3	0.61	0.50
Faugeras	6.0	39.2	0.31	0.39
Ahmed	7.6	38.7	0.61	0.48
Tsai	2.0	25.8	0.14	0.10
Tsai-LM	13.6	39.8	1.12	0.69
Heikkilä	15.4	15.4	1.23	0.74
Batista	1.9	23.1	0.32	0.53
Zhang	11.0	27.8	0.44	0.46

Table 1: Stability of some extrinsic parameters with real data.

Table 1 shows the mean of the maximum

differences of the four sets of views that were calibrated. It can be observed that, in spite of the fact that only an intrinsic parameter was changed, most of the extrinsic parameter showed important variations, especially the distance to the pattern (T_z).

4.1.2 Stability of intrinsic parameters

Three sets of views with a constant internal camera configuration were used to test the stability of intrinsic parameters. In each sequence a pattern with 108 points was displaced horizontally.

	Δu_0 (pix.)	Δv_0 (pix.)	Δk_x (pix.)	Δk_y (pix.)
Linear	35.2	6.4	16.3	16.7
Faugeras	35.3	6.7	16.3	16.7
Ahmed	35.1	6.4	16.2	16.7
Tsai	-	-	2213	2257
Tsai-LM	373.8	445.5	117.9	117.2
Heikkilä	124.1	336.1	121.6	123.9
Batista	0.2	0.7	1.0	0.8
Zhang	42.3	18.2	41.7	39.6

Table 2: Stability of some intrinsic parameters with real data.

Table 2 shows the mean of the maximum differences of the results for the three sequences with horizontal displacement of the pattern. It can be seen that the intrinsic parameters are not constant, especially the horizontal component of the optic center. It also reveals that Heikkilä's method and the optimized version of Tsai's method exhibit considerable variability in all the intrinsic parameters. Moreover, the huge variability of focal length obtained by Tsai's method is particulary outstanding. Finally, Batista´s method presents the most stable results because the majority of the intrinsic parameters are estimated in the last step of the algorithm when the rest of the parameters have been adjusted, thus this method obtains results close to the initial guess values.

Figure 2 shows that the variation of the optic center in many methods is directly proportional to the displacement of the pattern on the screen.
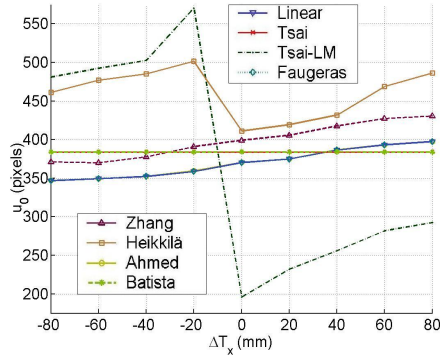

Figure 2: Stability of u_0 ($u_0 = 384 pixels$) with real data.

4.2 Experiments in the simulator

When the calibration is performed with real data, ground truth values for each parameter of the camera model are unknown so it is impossible to determine their accuracy. To overcome this difficulty, a simulator was developed using Matlab.

4.2.1 Stability of extrinsic parameters

This experiment was carried out in similar conditions to a real camera. Low distortion ($K_1 = -0.0015 \frac{1}{mm^2}$) and noise level ($\sigma_{2D} = 0.15 pix; \sigma_{3D} = 0.4mm$) were injected into images. Fifty sets of views or sequences of the pattern were used to eliminate the random effects of noise. The pattern was kept static ($T_z = 1150mm$) and the focal length of the camera was varied between 4.5 to 13.5 mm in each sequence.

Table 3 shows the mean of the maximum differences of the fifty sequences that were calibrated. It can be observed that this result is very similar to the calibration result of the real camera. The extrinsic parameters, particularly the distance to the pattern (T_z), are not constant.

Moreover, it was observed that all the methods produced similar levels of global error under these conditions. However, when the error in each parameter was analyzed separately

	ΔT_x (mm)	ΔT_z (mm)	ΔR_x ($^\circ$)	ΔR_y ($^\circ$)
Linear	3.1	37.5	0.28	0.15
Faugeras	3.7	37.7	0.27	0.19
Ahmed	3.1	37.7	0.27	0.15
Tsai	2.4	37.3	0.12	0.12
Tsai-LM	11.6	13.9	0.83	0.6
Heikkilä	42	30.7	2.93	1.99
Batista	2.2	9	0.39	0.32
Zhang	18.3	27.6	0.45	0.78

Table 3: Stability of some extrinsic parameters with synthetic data.

then it was observed that each method obtained different levels of error (Figure 3). Also, the experiments have shown that there is not relation between the global accuracy and the stability of the methods.

4.2.2 Stability of intrinsic parameters

This experiment was carried out in the same conditions as the previous one. In this case the camera setup was kept constant while the pattern was displaced horizontally.

The results are summarized in Figure 4 where the location of optic center varies proportionally to the displacement of the pattern on the screen. It is observed that Heikkilä's method and the optimized version of Tsai's method have the largest variability.

5 Conclusions

This paper has presented the analysis of an objective comparison of the stability and global accuracy carried out with eight camera calibration methods that use a pattern as reference.

Theoretically, when calibration is effected using different views of the same pattern located at different places and the camera setup is constant, the values of the intrinsic parameters estimated in the calibration process should be the same. If we consider the experimental results, this assumption does not hold in practice, which has serious consequences from a practical point of view. It is possible to
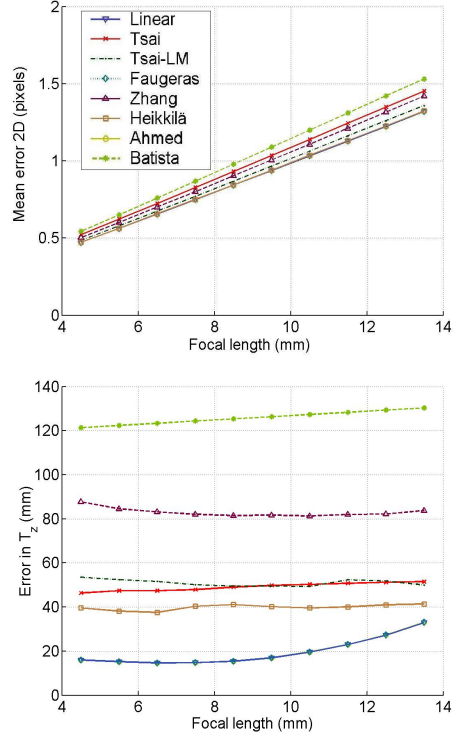

Figure 3: Global and T_z accuracy.

find large discrepancies between the obtained values in the calibration of each view. Similar conclusions have been obtained with the reciprocal situation of a static camera that changes its internal setup.

The comparison of the methods results in these experiments have shown that Batista's method presents, in general, the most stable results. Tsai´s method has obtained very stable results for the extrinsic parameters, but the estimation of focal length is very variable. Heikkilä's method has obtained the most stable results in the estimation of T_z, but it has high variability in the results of the rest of the parameters.

The experiments have also shown that the values of the parameters obtained in the calibration process depend on the method used. In other words, differences exist between the
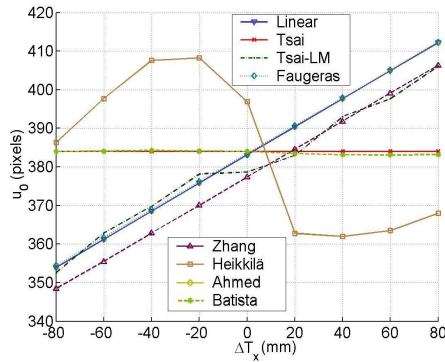
Figure 4: Stability of u_0 ($u_0 = 384 pixels$) with synthetic data.

estimated value for a specific parameter obtained by each method, using the same images to carry out the calibration.

Finally, the error function minimized by the methods, usually the image disparity, yields similar error levels with all them. But this does not guarantee that the parameter estimations converge to the ground truth values, which is a serious problem if we want to use the calibrated camera in mobile applications.

References

[1] M. Ahmed, E. Hemayed, and A. Farag. Neurocalibration: a neural network that can tell camera calibration parameters. In *IEEE International Conference on Computer Vision*, pages 463–468, September 1999.

[2] J. Batista, H. Araújo, and A.T. de Almeida. Iterative multi-step explicit camera calibration. In *Proc. Sixth International Conference on Computer Vision ICCV98*, volume 15, pages 709–714, January 1998.

[3] Olivier Faugeras. *Three-dimensional computer vision: a geometric viewpoint*. MIT Press, 1993.

[4] J. Isern González. *Estudio comparativo de métodos de calibración y autocalibración de cámaras*. PhD thesis, University of Las Palmas de Gran Canaria, Las Palmas, Spain, July 2003.

[5] P. Grattoni, G. Pettiti, and F. Pollastri. Geometric camera calibration: a comparison of methods. In *Proc. 5th Int. Conf. on Advanced Robotics*, pages 1775–1779, 1991.

[6] J. Heikkilä. Geometric camera calibration using circular control points. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1066–1077, October 2000.

[7] Jim Z.C. Lai. On the sensitivity of camera calibration. *Image and Vision Computing*, 11(10):656–664, December 1993.

[8] J. Salvi, X. Armangué, and J. Batlle. A comparative review of camera calibrating methods with accuracy evaluation. *Pattern Recognition*, pages 1617–1635, June 2002.

[9] Roger Y. Tsai. A versatile camera calibration technique for high-accuracy 3d machin vision metrology using off-the shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation*, RA-3(4):323–344, August 1987.

[10] Roger Y. Tsai. Synopsis of recent progress on camera calibration for 3d machine vision. *The Robotics Review*, pages 147–159, 1989.

[11] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactios on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, November 2000.