CAMERA CALIBRATION USING COMPOUND GENETIC-SIMPLEX ALGORITHM

G. G. Savii[*]

"Politehnica" University of Timisoara, Faculty of Mechanical Engineering, Department of Mechatronics, 1 Mihai Viteazul Blvd., 300222 Timisoara, Romania

The paper presents a novel camera calibration method using a two-step approach. First, a genetic algorithm is used to find a good enough approximation of the solution. Then a multidimensional unconstrained nonlinear minimization (Nelder-Mead simplex) algorithm is used to refine the solution. This approach avoids errors due to linearizations and automatically finds a very good initial point for the error minimizing procedure. All the camera parameters (intrinsic and extrinsic) are determined simultaneously, giving a consistent solution. Tested on several cases, the proposed method proved to be an efficient tool for determining the camera parameters needed for various applications, like analytical photogrammetry, 3-D space reconstruction from 2-D images and vision-based head tracking.

1. Introduction

Three-dimensional (3-D) digitizing sensors are intensively used in photogrammetry and 3-D coordinate measurement [1,2], vision-based head-tracking [3,4], extracting 3-D models (that is, geometries and topologies) of physical objects in a facility [5] for computer aided design [6].

Tracking head movements is important in the design of an eye-controlled human/computer interface. There are various position tracking methods, most used being mechanical, magnetic and optical. Mechanical position trackers have a low lag, are much less sensitive to the environment than magnetic position trackers, and tend to be affordable. However, they have a small working volume, and their comfort is low because mechanical linkages create motion restrictions [3,4]. Magnetic position trackers are generally very flexible since they are small enough so that they can be attached to heads. However, they have possible important disadvantages, presenting erroneous readings caused by magnetic interference from devices such as radios or monitors, and inaccuracies caused by large objects of ferrous metals that interfere with the electromagnetic field [3,4,7]. Optical position trackers are able to work over a large area, but they need to maintain a line of sight from the set of reference points to the camera [3,4].

2. Camera calibration

Camera calibration is the process of determining the parameters of a mathematical camera model. Essentially, it is a system identification process. This allows one to match any location in the image with a line-of-sight in the real world. The line-of-sight is a ray that extends from the camera and includes every point in the world that could possibly be projected onto a point in an image. In other words, a many-to-one relationship exists between the possible 3D points along the ray and the single 2D imaged point.

[*] Corresponding author: george.savii@mec.utt.ro

For conventional cameras, a perspective projection model is frequently used, commonly known as the pinhole camera model. Camera parameters can be divided into two categories, namely intrinsic parameters and extrinsic parameters. Intrinsic parameters are independent of the camera pose. They may include the effective focal length, the width and the height of photo-sensor cell, and the image centre (i.e. the image coordinates of the intersection of the optical axis and the image sensor plane). Extrinsic parameters are essentially the camera pose (camera's position and rotation in the world coordinate frame). Hence, they are independent of intrinsic parameters [8].

The camera's lens system and its wider-than-a-pinhole aperture typically create discrepancies or distortion. As a result, the camera model additionally includes internal parameters that describe the distortion. The present paper does not take into consideration this distortion, assumed negligible.

Many techniques have been developed for camera calibration because of the strong demand of applications. Abdel-Aziz and Karara [9] introduced a direct linear transformation (DLT) method, an approach that has been used in analytical photogrammetry. The DLT method that does not consider lens distortion is the one that estimates a DLT matrix, which consists of the composite parameters made by intrinsic and extrinsic camera parameters. If necessary, given the DLT matrix, camera parameters can be easily determined. This method is used, for example, in biomechanical analysis to perform 3-D space reconstruction from 2-D images.

In the three-dimensional reference object-based calibration, camera calibration is performed by observing a calibration object whose geometry in 3-D space is known with very good precision. Calibration can be done very efficiently [10]. The calibration object usually consists of two or three planes orthogonal to each other.

The self-calibration techniques do not use any calibration object. Just by moving a camera in a static scene, the rigidity of the scene provides in general two constraints [11] on the cameras' internal parameters from one camera displacement, by using image information alone. Therefore, if images are taken by the same camera with fixed internal parameters, correspondences between three images are sufficient to recover both the internal and external parameters which allow the reconstruction of the 3-D structure up to a similarity. While this approach is very flexible, it is not yet mature. Because there are many parameters to estimate, one cannot always obtain reliable results.

3. Camera model

To obtain the pinhole camera model, the systems of coordinates must be specified (Fig. 1). All coordinate systems are Cartesian. The image plane is behind the optical centre F (in the lens plane), which is the origin of the camera coordinate system (CCS). The origin of the image coordinate system is the point corresponding to the centre of the frame memory. In the image plane, the coordinates are expressed in pixels, while in the other systems – in length units (e.g. meters).
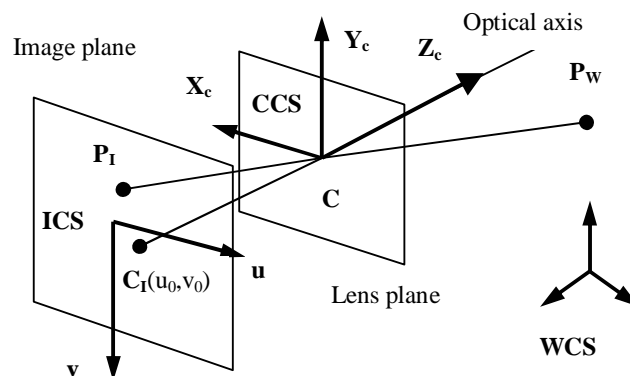

Fig. 1. Systems of coordinates for the camera model.

Given a point P (Fig. 1), its coordinates in the world coordinate system (WCS) are written as $\mathbf{P}_W = [X_W\ Y_W\ Z_W]^T$. It corresponds to the point $\mathbf{P}_C = [X_C\ Y_C\ Z_C]^T$ in the CCS, and is projected as the point $\mathbf{P}_I = [X_I\ Y_I]^T$ in the image plane (ICS).

The transformation matrix between the WCS and the CCS can be written as:

where $\mathbf{T}_C^W = [t_x\ t_y\ t_z]^T$ is the translation vector, and $\mathbf{R}_C^W$ is the 3x3 rotation matrix determined by

$$\mathbf{H}_C^W = \begin{bmatrix} \mathbf{R}_C^W & \mathbf{T}_C^W \\ 0 & 1 \end{bmatrix}$$

three Euler angles ($\phi_x,\ \phi_y,\ \phi_z$):

$$\mathbf{R}_C^W = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} = \begin{bmatrix} c_x c_y & s_x c_z + c_x s_y s_z & s_x s_z - c_x s_y c_z \\ -s_x c_y & c_x c_z - s_x s_y s_z & c_x s_z + s_x s_y c_z \\ s_y & -c_y s_z & c_y c_z \end{bmatrix}$$

where $s_x = \sin\phi_x$, $c_x = \cos\phi_x$, etc.

The point on the image plane is obtained by a perspective projection from a 3-D object point in the CCS to a 2-D point using the combined transformation matrix:

$$\mathbf{H}_F^C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/f & 1 \end{bmatrix}$$

where f is the focal length (and approximate distance between lens plane and image plane – effective focal length).

The coordinates in the image plane are obtained by scaling and translation, using the transformation matrix:

$$\mathbf{H}_I^F = \begin{bmatrix} du & 0 & u_0 \\ 0 & -dv & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

where du and dv are the horizontal and vertical resolutions of the image sensor (the minus sign takes care of the y axis inversion); u_0 and v_0 are the coordinates of the image centre (principal point - image plane intersection with optical axis) with respect to the ICS (in pixels).

The intrinsic parameters are: f, du, dv, u_0, v_0. The extrinsic parameters are: ϕ_x, ϕ_y, ϕ_z, t_x t_y t_z, making a total of eleven parameters for the camera model.

Performing a camera calibration with point-to-point correspondences requires a mathematical relationship between the world and image coordinates. For a point P, the relation can be written as:

$$\mathbf{P}_I = \mathbf{H}_I^F \cdot \mathbf{H}_F^C \cdot \mathbf{H}_C^W \cdot \mathbf{P}_W$$

4. The new camera calibration procedure

The matrix relation between the point coordinates in WCS and ICS is equivalent to two scalar equations, which means that in order to determine the value of the eleven unknown camera parameters, at least six points are needed with known coordinates in the world frame and in the image frame. A system of at least eleven highly nonlinear equations is obtained. One approach to solve this system is using the DLT method mentioned above [9]. In order to get the linear form, some assumptions are made, so instead of the focal length and sensor resolutions, the effective focal length (lens-image distance) in image coordinates, and the ratio of horizontal and vertical resolutions are computed [3,4,9,12], or one of the two resolutions is assumed known [3,4]. Usually a large number of points are used to create an overdetermined linear system. Moreover, it is important to note that a strictly coplanar set of target points requires a different formulation, since coplanar features will reduce the rank of the system matrix to eight. Under these circumstances, there is no explicit formulation of the solution. Any linear least-squares solver like QR Decomposition or Singular Value Decomposition can find a good solution, but is time consuming.

To overcome the problem of knowing a good set of initial values for the parameters to be used by the solver, a method based on genetic algorithms is proposed (Fig. 2).

As input data, the coordinates of the target points are given, both in the world coordinate system ($\mathbf{P}_W$) and on the image ($\mathbf{P}_I$). Then the limits of the search space (11 dimensions), and the values for the parameters of the genetic algorithm (number of chromosomes in the population, number of bits per unknown variable, crossover and mutation probabilities) are established. The first population is randomly generated and checked for feasibility, applying repairs if needed.

In order to calculate the fitness values, needed for sorting the chromosomes, a genotype to phenotype conversion is performed, i.e. the binary codes in the chromosomes are converted to real (geometric) values. The fitness values are computed as the sum of Euclidian distances between the actual target (given) points on the image and the calculated points using the phenotypes.

The selection process of a chromosome in the creation of a new generation is controlled by a

$$\mathrm{sel}(k) = \mathrm{ceil}\left(\frac{\sqrt{8k+1}}{2} - \frac{1}{2}\right) + m - 1, \quad k = 1,2,\dots,m$$

probability that depends on the fitness value, i.e. on the rank after sorting. The probability is simulated using a special vector:
where ceil is the ceiling function (round towards plus infinity) and m is the number of chromosomes in the population. The vector has the form: [m m-1 m-1 m-2 m-2 m-2…].

The iterative creation of new populations is interrupted when the expected best fitness is attained (e.g. zero sum of distances), or the allowed number of iterations is reached. The best chromosome of the last generation is converted to phenotypes (geometric values), which are used as initial point for the multidimensional unconstrained nonlinear minimization (Nelder-Mead simplex algorithm). A refined solution to the calibration method is obtained.

If the error of the Nelder-Mead solution is acceptable, the results are displayed. In the other case, the process is restarted with the genetic algorithm. Now the new population is initiated by using the best chromosome from the last population and a chromosome obtained by phenotype to genotype conversion of the Nelder-Mead solution.

5. Results

The proposed algorithm was tested on several cases. The appearance of one calibration object on the image plane is presented in Fig. 3 and the target points coordinates in Table 1.
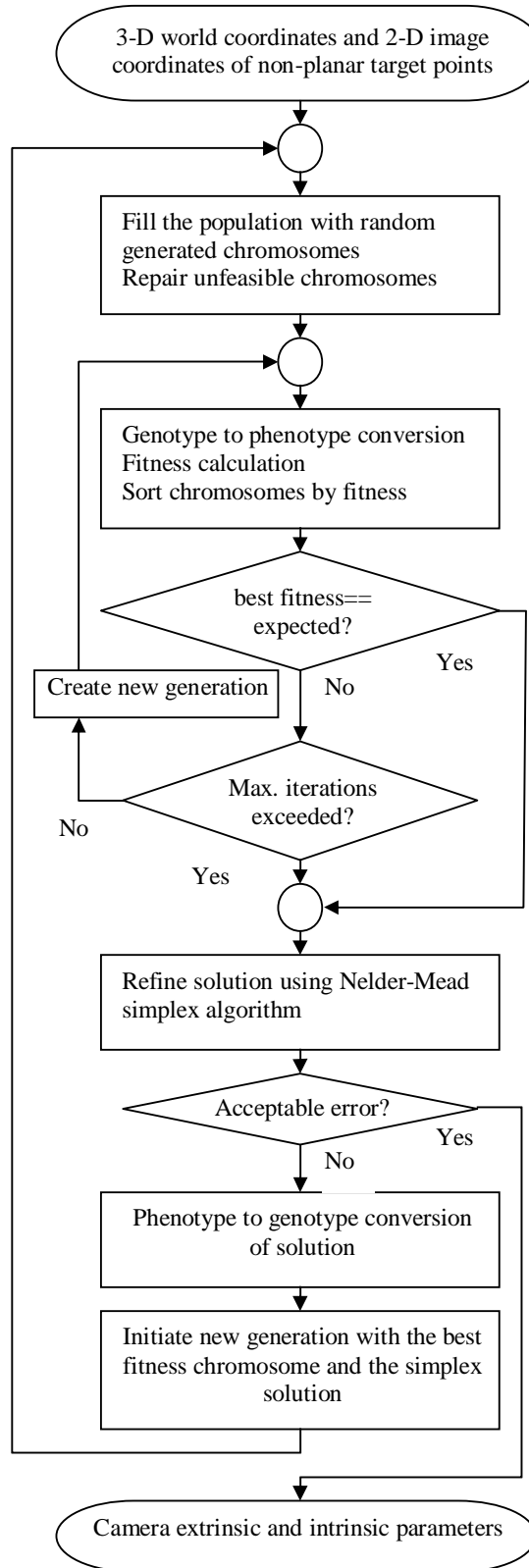
Fig. 2. Genetic-simplex algorithm for camera
calibration.

Fig. 4 presents the evolution of the errors during calibration for a population of 40 chromosomes, using 16-bit genes to represent each unknown parameter, a mutation probability of 0.01 and crossover probability of 0.90. The final errors, when iterations stopped after 12 "epochs" (restarts of the genetic algorithm), were 1.265 after genetic algorithm and 1.216 after Nelder-Mead refinement. The results were used to simulate the camera. The individual errors after simulation, for each pixel, are under half the size of a pixel (Table 1).
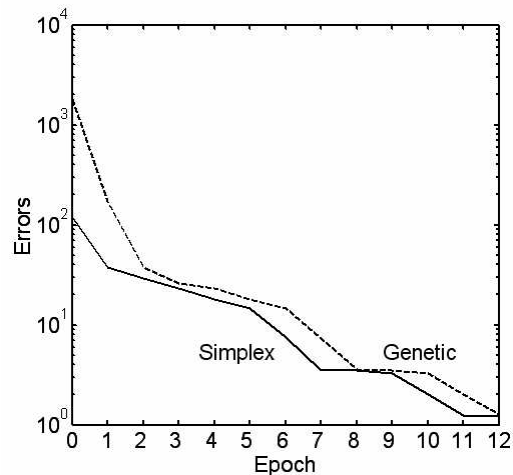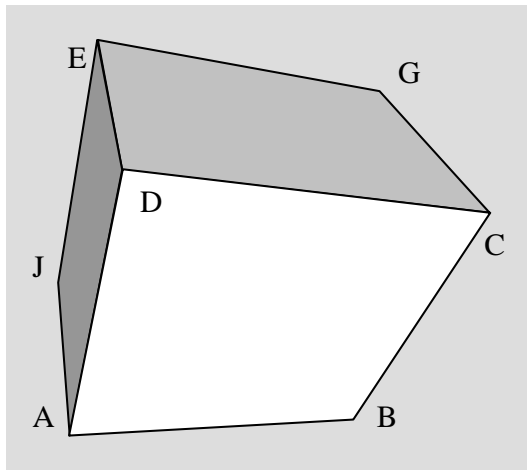

Fig. 3. Image of the calibration object with target points.

Fig. 4. Error evolution during calibration process.

Table 1. Target points and resulting errors after simulation with calibration results.

Point		A	B	C	D	E	G	J
WCS Coord.	X_W	0	933	933	0	0	933	0
	Y_W	0	0	877	877	877	877	0
	Z_W	0	0	0	0	-660	-660	-660
ICS Coord.	X_I	299	1098	1478	445	378	792	280
	Y_I	1126	1083	501	386	16	598	678
Error	dX	0.0004	-0.0528	0.0189	0.0000	0.3420	0.0661	-0.1548
	dY	0.0020	-0.0953	0.2047	0.0000	0.3574	-0.2087	0.1035

The calculated intrinsic parameters were:

$f = 7.3356$, $du = 226.81$, $dv = 227.97$, $u_0 = 792$, $v_0 = 598$,

and the extrinsic parameters:

$\mathbf{T}_C^W = [-199.48 \ 1496.3 \ 988.58]^T$, $\phi_x = -0.67115$, $\phi_y = -0.38519$, $\phi_z = 0.068227$.

For comparison, the camera (CCD sensor) resolutions given by the manufacturer were $du = 226.57$ and $dv = 226.06$, the focal length indicated by the camera $f = 7.09$, and the picture size 1600x1200.

6. Conclusions

Tested on several cases, the proposed method proved to be an efficient tool for camera calibration, with advantages over other known methods in many cases.

The implementation makes it possible to work with different number of unknown parameters and different number of target points.

References

[1] R. J. Tsai, IEEE Journal of Robotics and Automation, **3**, 323 (1987).

[2] Tzung-Sz Shen, Chia-Hsiang Menq, IEEE Transactions on Robotics and Automation **17**, 502 (2001).

[3] K. S. Park, C. J. Lim, Int. J. Human-Computer Studies **52**, 879 (2000).

[4] K. S. Park, C. J. Lim, Int. J. Human-Computer Studies **54**, 319 (2001).

[5] D. Zetu, P. Schneider, P. Banerjee, IIE Transactions **30**, 597 (1998).

[6] G. G. Savii, Encyclopedia of Information Systems**,** Academic Press 2002, p.171.

[7] K. Meyer, H. L. Applewhite, F. A. Biocca, Presence, **1**, 173 (1992).

[8] S. W. Shih , Y. P. Hung, W. S. Lin, Technical Report, TR-IIS-96-003, Institute of Information & Science, Academia Sinica, Nankang, Taipei, Taiwan, 1996.

[9] Y. I. Abdel-Aziz, H. M. Karara, Symposium on Close-Range Photogrammetry, Urbana, Illinois, 1971, p. 1.

[10] O. Faugeras, Three-Dimensional Computer Vision: A Geometric Viewpoint, MIT Press (1993).

[11] S. J. Maybank, O. D. Faugeras, Internat. J. Computer Vision **8,** 123 (1992).

[12] G. C. Yerem, Master of Science Thesis, University of Tennessee, Knoxville (2001).