CAMERA CALIBRATION METHOD FOR STEREO MEASUREMENTS

**Sergiu Nedevschi, Tiberiu Marita, Mircea Vaida, Radu Danescu,
Dan Frentiu, Florin Oniga, Ciprian Pocol, Daniel Moga**

Technical University of Cluj-Napoca, 15 C. Daicoviciu Str., 3400 Cluj-Napoca
E-mail:Sergiu.Nedevschi@cs.utcluj.ro, Marita.Tiberiu@cs.utcluj.ro, rdanescu@vision.utcluj.ro

Abstract: *Camera calibration is an important and sensitive step in 3D environment reconstruction by stereovision. Small errors in the initial estimation of the camera parameters (especially in the estimation of the principal point) could rise to high errors in the 3D measurements, which are increasing with the working distance. There are many general-purpose calibration methods for calibration of individual cameras. We propose a method for refining the results of such methods by inferring stereo information from a controlled 3D scene. The parameters of the two cameras composing a stereovision system are recalibrated together by minimizing the depth reconstruction error of the control points from the reference scene.*

Key words: *camera calibration, camera model, stereovision, depth estimation, 3D reconstruction.*

1. INTRODUCTION

Camera calibration in the context of three-dimensional machine vision is the process of determining the internal geometric and optical characteristics (intrinsic parameters) of the camera and/or the 3D position and orientation of the camera coordinate system relative to a certain world coordinate system (extrinsic parameters).

The correct estimation of these parameters is crucial for applications based on artificial vision in order to perform a correct 3D reconstruction from 2D images of the environment for far working ranges. The performances of these applications are strongly affected by their precision especially when stereovision algorithms are used.

1.1. *General-purpose camera calibration methods*

Several methods for general-purpose geometric camera calibration are presented in the literature. Some of their assessment criteria are: autonomy, accuracy, efficiency and versatility [1]. They can be classified in the following categories [2,3,4]:

a. By assumptions on estimated parameters:

Photogrammetric calibration – By observing a calibration object whose geometry in 3-D space is known with very good precision.

Self-calibration – Does not use any calibration object. Images are taken by the same camera with fixed internal parameters by moving a camera in a static scene. Correspondences between three images are sufficient to recover both the internal and external parameters.

b. Considering different forms of calculated parameters:

Explicit calibration – All parameters attached to the camera should be calculated explicitly (the traditional approach for the camera calibration).

Implicit calibration – Only some intermediate parameters are required (e.g. camera matrix), while no need to find out the exact value for each physical parameter of the camera used.

c. According to the complexity and complete level of the estimation process:

Linear methods – The imaging process is described in terms of a simple linear equation.

The parameters of the linear equation are functions of the physical camera parameters. Lens distortions are usually not modeled.

Combined methods (linear and nonlinear) – Two-step methods: the initial parameter values are computed linearly, then a nonlinear camera model is considered (including noise and lens distortion). The final parameters values are obtained by solving a nonlinear data fitting problem in a least squares fashion by minimizing the error function between the estimated and observed data.

The **Direct Linear Transformation (DLT)** [5] is the simplest camera calibration method based on the pinhole camera. Does not consider any radial and tangential distortions and is used for an initial estimation of the camera parameters and thus still plays an important role in computer vision [2].

Tsai's calibration algorithm [1] is based on the pinhole model and it assumes that only the first order radial distortion occurs in the camera. Under that assumption, a pixel in the image is only distorted along the radial direction, thus the radial alignment constraint (RAC) model can be employed to make the calibration process linear.

Zhang's calibration algorithm [3] requires the camera to observe a planar pattern shown at a few (at least two) different orientations. Radial lens distortion is modeled. An estimation of the calibration parameters is done using a closed-form (linear) solution using homographies followed by a nonlinear optimization technique based on the maximum likelihood criteria.

Heikkila's calibration algorithm [4] uses the same pinhole camera model and both radial and tangential lens distortions are modeled. Initial parameters estimation is done using the direct linear transformation (DLT), ignoring lens distortions. Parameters are refined using nonlinear estimation (when both noise and lens distortions are considered). The final parameters are obtained by minimizing the total squared error between the model and observation data.

Bouguet's method (implemented in the CalTech Calibration Toolbox**)** [6,7] uses Heikkila's intrinsic camera model. In the initialization phase the homographies for all points on series of images are found and the intrinsic parameters are initialized by setting the distortion to 0. Then the extrinsic parameters are found for each image of the pattern. The final nonlinear optimization using Maximum Likelihood estimation is identical to Zhang,s method, augmented with the tangential distortion coefficients.

1.2. Calibration method suited for stereo measurements

A camera calibration method is good as long as it permits the stereo algorithm to establish good 3D parameters, given the correspondence of features in stereo image pairs. The general-purpose calibration methods are used to calibrate each camera individually and do not aim to infer stereo information. The only success criterion for the calibration is that the reprojected image of some control points (grid pattern of known size) should coincide with their actual position in the image.

It was demonstrated however that even if the projection errors were near zero, the 3D reconstruction from the stereo algorithm was not precise enough for measurements at far ranges. A solution oriented towards stereo must be developed, but this requires a lot of time and work. Thus we propose a combined solution which estimates the camera parameters in the classical way and then refines them using stereovision information:

a. The initial estimation is done using an adapted version of the Bouguet algorithm. The main drawback of this method is that is highly dependent on the calibration image set (size and position of calibration pattern, lighting conditions, camera quality – noise level and stability, number of images) and errors are encountered in the estimation of the principal point coordinates. Our 3D reconstruction experiments with our stereo camera setup proved that an error of 1 pixel in horizontal direction or 3 pixels in vertical direction of the principal point coordinates rises to an error of 10% in the depth estimation because erroneous intrinsic parameters rise also to errors in the extrinsic parameters estimation.

b. Then these initial parameters are refined/corrected by inferring stereo information. For this second step we propose **the brute force calibration method** which performs 3D reconstruction on a set of feature pairs of known distance (depth), and refines the initial calibration parameters by minimizing the estimated distance error to those control points.

2. THE CAMERA MODEL

The pinhole camera model [8] augmented with the lens distortions [4] is used (Fig.1).

The following camera parameters are considered:

Intrinsic parameters – describing the internal geometric and optical characteristics of the camera (those that specify the camera itself):

- focal length, that is, the distance between the camera lens and the image plane: ***f***;
- the principal point - location of the image center in pixel coordinates: (u_0, v_0);
- the radial and tangential distortion coefficients of the lens.

Extrinsic parameters - the 3-D position and orientation of the camera frame relative to a certain world coordinate frame (are needed to transform world coordinates to a camera centered coordinate frame or vice versa):

- the translation vector $\boldsymbol{T} = [\ Tx,\ Ty,\ Tz\ ]^T$;
- the rotation vector $\boldsymbol{r} = [\ Rx,\ Ry,\ Rz\ ]^T$ or its equivalent rotation matrix $\boldsymbol{R}$.

The projection of a 3D point *P*, having coordinates (X, Y, Z) in the camera reference frame, on the image plane is computed using the intrinsic parameters as follows:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = A \cdot \begin{bmatrix} x_D \\ y_D \\ 1 \end{bmatrix} \tag{1}$$

where:

$[u\ \ v\ \ 1]^T$ – is the pixel coordinate vector in the image frame;

$[x_D\ \ y_D\ \ 1]^T = [x+\delta x\ \ \ y+\delta y\ \ \ 1]^T$ – is the distorted image point vector in camera coordinate frame;

$[x\ \ y]^T$ – is the undistorted image point vector in camera coordinate frame; the relation between $[x\ y]$ and $[X\ Y\ Z]$ is given by the fundamental equations system of the perspective camera model are [8]:

$$x = f_X \cdot \frac{X}{Z}; \qquad y = f_Y \cdot \frac{Y}{Z}. \tag{2}$$

f_X , f_Y – are the focal lengths expressed in units of horizontal and vertical pixels.

$\boldsymbol{A}$ – is the internal camera matrix:

$$A = \begin{bmatrix} f_X & 0 & u_0 \\ 0 & f_Y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

$[\delta x\ \ \delta y\ ]$ – is the total offset caused by the radial and tangential distortions of the lens [4]:

$$\begin{bmatrix} \delta x \\ \delta y \end{bmatrix} = \begin{bmatrix} \partial x^r + \partial x^t \\ \partial y^r + \partial y^t \end{bmatrix} \tag{4}$$

$[\delta x^r\ \ \delta y^r\ ]$ – is the radial distortion, modeled using the first and second order radial distortion coefficients k_1 and k_2:

$$\begin{bmatrix} \partial x^r \\ \partial y^r \end{bmatrix} = \begin{bmatrix} x \cdot \left(k_1 \cdot r^2 + k_2 \cdot r^4 + \ldots\right) \\ y \cdot \left(k_1 \cdot r^2 + k_2 \cdot r^4 + \ldots\right) \end{bmatrix} \tag{5}$$

where: $r^2 = x^2 + y^2$;

$[\delta x^t\ \ \delta y^t\ ]$ – is the tangential distortion modeled by the tangential distortion coefficients p_1 and p_2:

$$\begin{bmatrix} \partial x^t \\ \partial y^t \end{bmatrix} = \begin{bmatrix} 2p_1 \cdot xy + p_2(r^2 + 2x^2) \\ p_1(r^2 + 2y^2) + 2p_2 \cdot xy \end{bmatrix} \tag{6}$$

The transformation between the coordinates of the point *P* expressed in the world reference frame $\boldsymbol{XX_W} = [\ X_W\ ,\ Y_W\ ,\ Z_W\ ]^T$ and its coordinates in the camera reference frame $\boldsymbol{XX_C} = [\ X\ ,\ Y\ ,\ Z\ ]^T$ (direct mapping) can be performed trough the following matrix equation:

$$XX_C = R_{WC} \cdot XX_W + T_{WC} \tag{7}$$

where:

$\boldsymbol{T_{WC}}$ – is the world to camera translation vector – position of the world reference frame in the camera's one.

$\boldsymbol{R_{WC}}$ – is the world to camera rotation matrix.

The rotation is represented using Euler angles R_X (pitch), R_Y (yaw), and R_Z (tilt) which define a sequence of three elementary rotations of the world frame around *x, y, z*-axis respectively. The angles signs are given by the right hand rule.

For the reverse transformation (inverse mapping - 3D reconstruction), from the camera reference frame to the world reference frame the equation can be deduced from (7):

$$XX_W = R_{WC}^{-1} \cdot (XX_C - T_{WC}) \tag{8}$$

Because the rotation matrix is an orthogonal one [8], the inverse of $\boldsymbol{R}$ is equal with its transpose:

$$XX_W = R_{WC}{}^T \cdot (XX_C - T_{WC}) \quad (9)$$

$$XX_W = R_{CW} \cdot (XX_C + T_{CW}) \quad (10)$$

where:

$\boldsymbol{T_{CW}}$ – is the camera to world translation vector;

$\boldsymbol{R_{CW}}$ – is the camera to world rotation matrix:

$$R_{CW} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \quad (11)$$

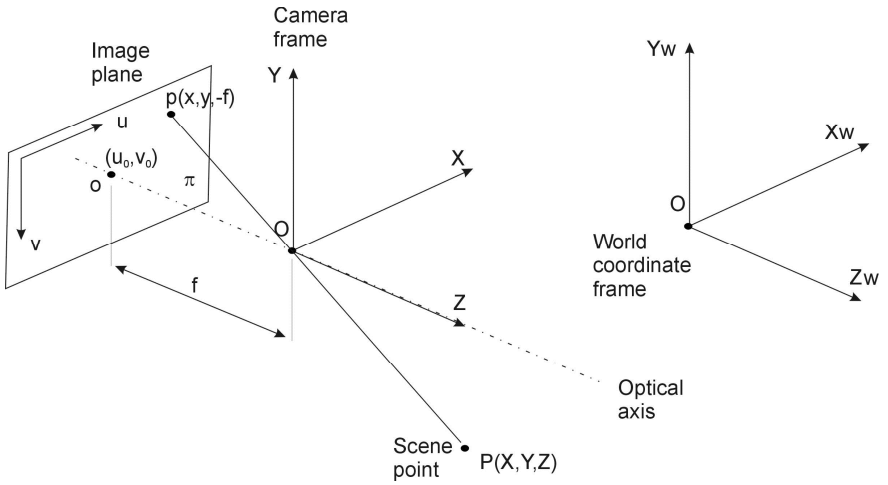

Fig. 1. The pinhole camera model

3. INITIAL ESTIMATION OF THE CALIBRATION PARAMETERS

The initial estimation is done using an adapted version of the original Caltech Calibration Toolbox. The Bouguet (CalTech) method [6, 7] uses the intrinsic camera model described in chapter 2. In the initialization phase the homographies for all control points on series of images are found and the intrinsic parameters are initialized by setting the distortion to 0. Then the extrinsic parameters are found for each image of the pattern. The final nonlinear optimization using Maximum Likelihood estimation is identical to Zhang's method [3], augmented with the tangential distortion coefficients.

As calibration pattern, a plane plate with black and white squares of known sizes is used. Each camera is calibrated individually. For intrinsic parameters a set of at least 20 images must be taken at several orientations and distances, with the pattern covering the maximum area from the image, as possible (Fig. 2). A good set of input images gives smaller uncertainties (less then 5 pixels) in the estimation of the parameters and determines the algorithm to converge in fewer steps.

Then the extrinsic parameters are estimated. For this task the cameras are mounted on a stereo rig, and the pattern is positioned to a known position relative to the cameras. For each camera a single image is taken and the extrinsic parameters of each camera, relative to a world reference frame related to the calibration pattern, are computed.

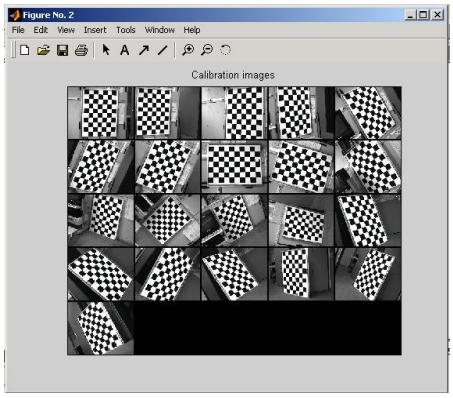

Fig. 2. Set of images for intrinsic calibration

The main drawback of this method is that is highly dependent on the calibration image set (size and position of calibration pattern, lighting conditions, camera and lens quality, lens focusing and aperture, number of images). If the proper intrinsic calibration methodology is not kept, errors could be encountered especially in the estimation of the principal point coordinates. The principal point estimation errors can degrade drastically the quality of the extrinsic parameters and the stereo reconstruction process.

4. THE BRUT FORCE CALIBRATION METHOD

The method is based on the hypothesis that the main source of errors is the position of the principal point of the two cameras. Finding the right position of this point has proven a difficult task, and the oscillations of this position in different calibration sessions had lead us to believe that this position is the main source for errors.

The main steps of the method are:

i. Initial intrinsic parameters calibration using the method described in chapter 3.

ii. Initial extrinsic parameters calibration using the method described in chapter 3. The pair of acquired images and the distance from the cameras to the calibration pattern (the distance estimation can be considered accurate enough for this purpose) are kept for the next steps.

iii. Take stereo image pairs with objects at known positions. The distance to these objects should be accurately measured using a laser device. The origin of this distance must be the world reference system, that is the position of the pattern used for extrinsic calibration. The distance distribution must be as uniform as possible.

iv. Matching corresponding features of the objects (control points) in the image pairs and providing distances to the objects. A database is build, containing:

- Coordinates of the point in the left image;
- Coordinates of the point in the right image;
- Distance to the point from the world coordinate system.

v. Recalibrating the camera parameters: for any possible position of the principal point of the two cameras in a given search interval depending on the current search step (serch_interval = search_magnitude x search_step) around the original position (given by the classical method) the following operations are done:

- Perform extrinsic calibration using the acquired extrinsic images in step **ii** ;
- For any point in the database, compute its distance using the general stereo reconstruction algorithm (chapter 5), with the parameters obtained from the current extrinsic calibration;
- Sum the relative errors of the points. Keep the minimum total relative error, its principal point coordinates and the extrinsic parameters that lead to it.

vi. Diminish the search interval by halving the search step and continue the search around the new principal point providing minimum error by repeating step **v.** until the search step is small enough (about 0.05).

vii. The last found principal points coordinates providing minimum error are kept and the intrinsic and extrinsic parameters are updated.

5. THE 3D RECONSTRUCTION ALGORITHM

The problem of mapping a pair of correspondent image points (p_L, p_R) into a unique 3D point P_W (X_W, Y_W, Z_W**)** is called 3D reconstruction. The stereo camera configuration used is a general one, because it was proved that is very difficult to obtain a canonical one due to the mechanical imprecision of the cameras and the mounting rig.

Solving the problem is straight forward: P_W is positioned at the intersection of the two projection rays that generates $p_L(x_L, y_L)$ and $p_R(x_R, y_R)$. First the 3D equations of the two projection rays (which are passing trough P_W and the two optical centers of the two cameras) should be found. This is a widely covered subject [9,10].

Computing the intersection can be done in a geometric fashion or an algebraic fashion. Algebraically, the intersection of the rays is obtained by solving equation system (12). The distortion coefficients of the lenses were ignored.

$$A * \begin{vmatrix} X_W \\ Y_W \\ Z_W \end{vmatrix} = B \qquad \Leftrightarrow$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{vmatrix} * \begin{vmatrix} X_W \\ Y_W \\ Z_W \end{vmatrix} = \begin{vmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{vmatrix} \qquad (12)$$

where:

$$A = \begin{vmatrix} r_{13}x'_L + r_{11}f_L & r_{23}x'_L + r_{21}f_L & r_{33}x'_L + r_{31}f_L \\ r_{13}y'_L + r_{12}f_L & r_{23}y'_L + r_{22}f_L & r_{33}y'_L + r_{32}f_L \\ r'_{13}x'_R + r'_{11}f_R & r'_{23}x'_R + r'_{21}f_R & r'_{33}x'_R + r'_{31}f_R \\ r'_{13}y'_R + r'_{12}f_R & r'_{23}y'_R + r'_{22}f_R & r'_{33}y'_R + r'_{32}f_R \end{vmatrix}$$

r_{ij} *si* r_{ij}' – are the elements of the camera-to-world rotation matrices for the left ($\boldsymbol{R}_{CL}$) and the right camera ($\boldsymbol{R}_{CR}$) respectively ($\boldsymbol{R}_{CW}$ matrix from chapter 2).

f_L , f_R – average focal lengths of the left and right camera respectively: $(f_x + f_y)/2$;

(x_L', y_L'), (x_R', y_R') – coordinates of image points p_L and p_R relative to the image plane centers (principal points):

$$\begin{cases} x_L' = x_l - u_{0L} \\ y_L' = y_L - v_{0L} \\ x_R' = x_R - u_{0R} \\ y_R' = y_R - v_{0R} \end{cases}$$

$$B = \begin{vmatrix} a_{11}X_{CL} + a_{12}Y_{CL} + a_{13}Z_{CL} \\ a_{21}X_{CL} + a_{22}Y_{CL} + a_{23}Z_{CL} \\ a_{31}X_{CR} + a_{32}Y_{CR} + a_{33}Z_{CR} \\ a_{41}X_{CR} + a_{42}Y_{CR} + a_{43}Z_{CR} \end{vmatrix}$$

and

$\boldsymbol{T}_{CL} = \begin{vmatrix} X_{CL} \\ Y_{CL} \\ Z_{CL} \end{vmatrix}$ and $\boldsymbol{T}_{CR} = \begin{vmatrix} X_{CR} \\ Y_{CR} \\ Z_{CR} \end{vmatrix}$ are the positions of the two camera coordinate systems in the World reference system ($\boldsymbol{T}_{CW}$ vector from chapter 2).

Because the number of equations in system (12) is greater then the number of variables and due to the fact that the intrinsic and extrinsic parameters are determined with some errors, this system must be solved using a method that should minimize the indetermination (for example LSE – Least Squares Estimation). Geometrically, this means that the projection rays do not intersect each other and the 3D point must be determined as the middle of the common orthogonal segment for the two rays [8].

6. EXPERIMENTAL RESULTS

A comparison of the calibration results using the classical method based on the Caltech Calibration Toolbox and using the brut force calibration method is presented. The estimated distance by stereo reconstruction to some control points were the assessment criteria.

The set of control points were the four visible corners of a cubical object, placed at several distances, uniformly distributed in a range from 14 to 107 m (fig. 3) measured in world reference frame coordinates. These control point were also used in the brut force calibration method for building the database with the needed data for inferring stereo information in the recalibration process. The distances to the control points were measured using a laser measurement device. The origin of the world reference frame (related to the pattern position in the extrinsic calibration process) was at about 6.8m in front of the cameras.

Results of the classical calibration procedure are presented in Table 1. The principal point position was wrong estimated (due to an improper set of input images for the intrinsic calibration procedure).

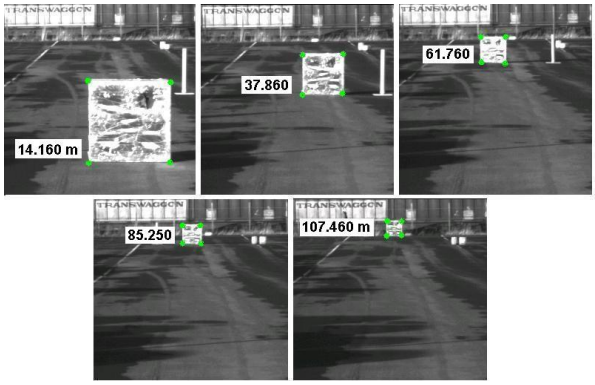

Fig. 3. Example of reference pictures with objects placed at different known distances and the selected control points (green). Only left images are shown

In figure 4 are presented the results of the depth estimation by stereo reconstruction for the previously considered 20 control points. Because the principal point was wrong estimated, also the extrinsic parameters were wrong estimated. This lead to high errors in the depth estimation, which are increasing linearly with the distance (Fig. 4.b):

Table 1. Intrinsic parameters from the classical calibration method:

	fx [pixels]	**fy [pixels]**	**u_0 [pixels]**	**v_0 [pixels]**
Left camera	2387.8	2387.3	353.2960	209.9128
Right camera	2391.6	2390.5	423.6890	218.8417

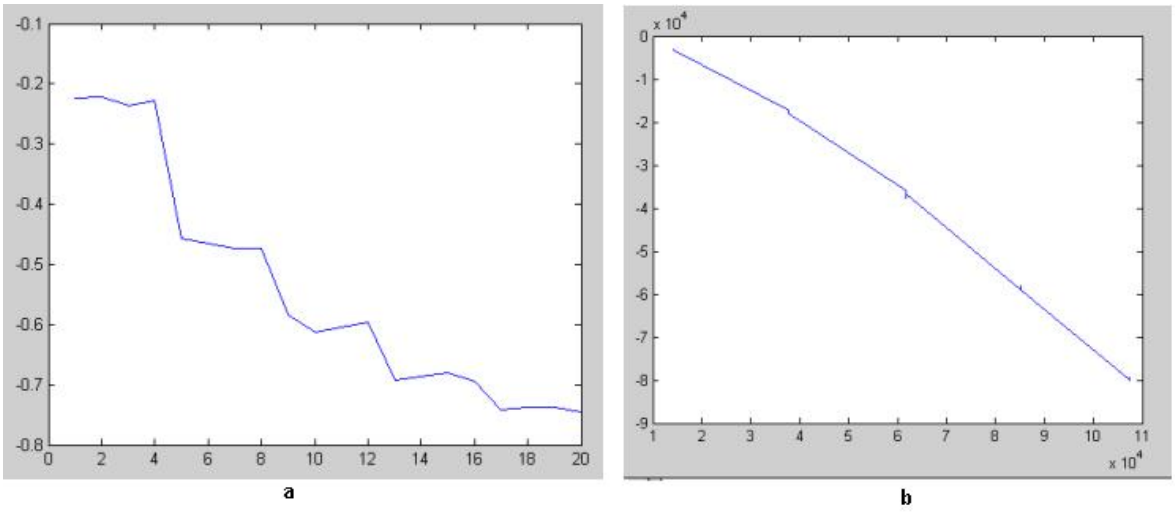

Fig. 4. Error of the depth reconstruction using parameters from the classical calibration method: a). relative error vs. control point; b). absolute error vs. depth [mm]

Results of the brut force calibration procedure are presented next. The principal point position was corrected by inferring stereo information. The correction was done only for the horizontal coordinates due to the configuration of the stereovision system (with cameras displaced on horizontal direction - horizontal baseline) and therefore in the reconstruction process only the horizontal disparity was used. For the correction of the principal point on vertical direction a vertical displacement of the two cameras should be used, the principle of the method remaining the same.

The corrected coordinates of the principal point of the two cameras are:

- Left : (u_0, v_0) = (**444.11628,** 209.91278)
- Right: (u_0, v_0) = (**467.34139,** 218.8417)

The corrected extrinsic parameters for 3D reconstruction are shown below:

$$T_{CL} = \begin{bmatrix} 209.748 \\ -493.988 \\ -6849.142 \end{bmatrix}; \quad T_{CR} = \begin{bmatrix} 584.437 \\ -493.516 \\ -6855.708 \end{bmatrix};$$

$$R_{CL} = \begin{bmatrix} 0.999876 & 0.001427 & 0.015696 \\ -0.002655 & 0.996908 & 0.078534 \\ -0.015535 & -0.078566 & 0.996788 \end{bmatrix};$$

$$R_{CR} = \begin{bmatrix} 0.999562 & 0.007656 & 0.028590 \\ -0.009748 & 0.997228 & 0.073763 \\ -0.027946 & -0.074009 & 0.996866 \end{bmatrix}.$$

In figure 5 are presented the results of the depth estimation by stereo reconstruction for the same 20 control points. The absolute error is under 2m (with two exceptions 3 m at 85.250 m and -4m at 107.460m), and does not depend on depth.

7. CONCLUSIONS

The proposed calibration method can be used to refine the parameters obtained with general-purpose calibration algorithms. Their correct estimation is essential in the stereo reconstruction process, which is sensitive to the sum of all errors in the calibration process. Thus stereo information from a controlled 3D scene was used to perform the calibration improvements. The experimental results proved that even a set of poor initial calibration parameters (in terms of principal point) can be corrected using this method in order to perform precise 3D reconstruction of the working environment for far depth ranges.

The limitation of this method is in the case when significant errors are encountered in the initial estimation of the other camera intrinsic parameters (focal length, distortion coefficients). In this case a solution that estimates all parameters of the cameras in a stereo fashion, similar to the method presented here, is foreseen.

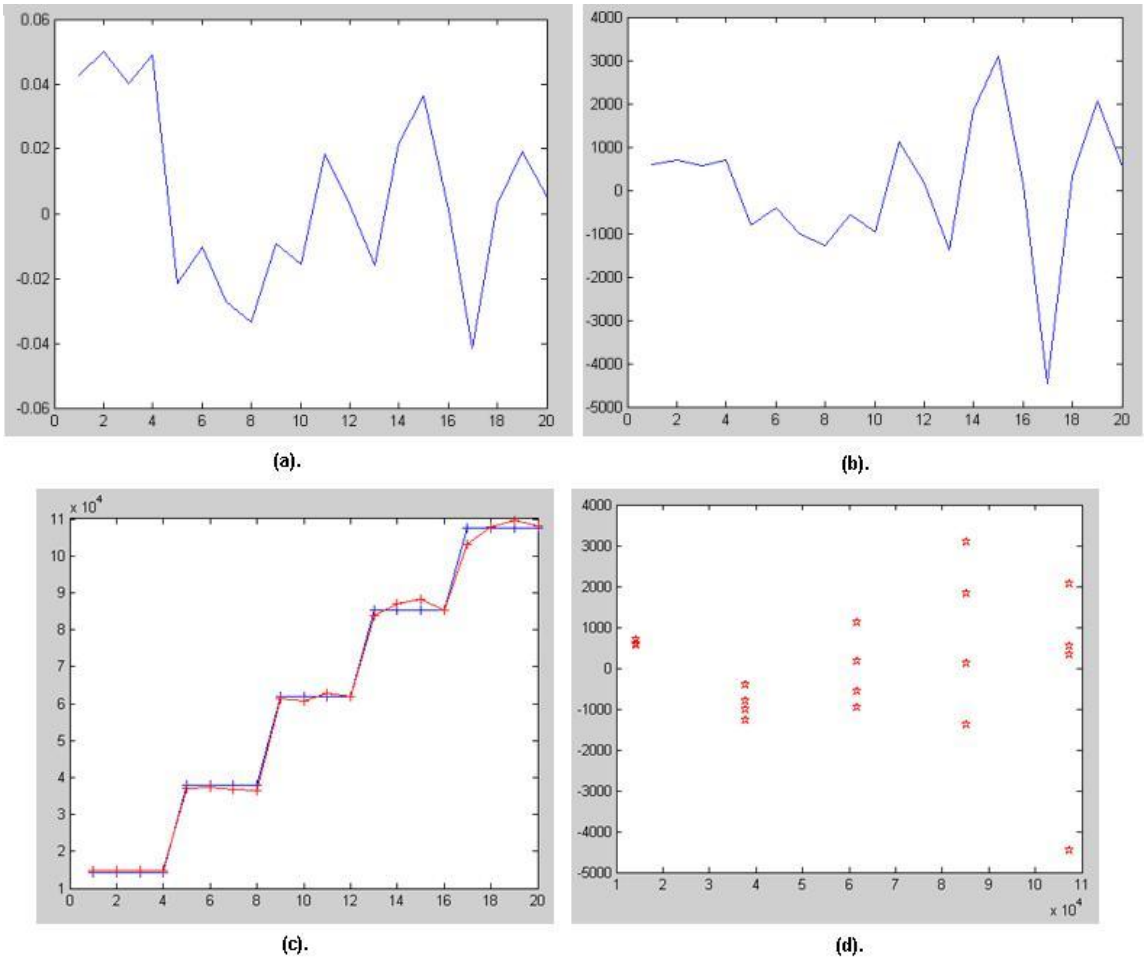

Fig. 5. Error of the depth reconstruction using parameters from the brut force calibration method for the 20 control points: a). Relative error vs. control point; b). Absolute error vs. control point [mm]; c). Measured (blue) and estimated (red) depths [mm] vs. control point; d). Absolute error vs. depth [mm]

REFERENCES

[1] Tsai, R. Y.(1987), "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*, RA-3(4)/1987: pp.323-344.

[2] B.J. Lei, E. A. Hendriks, M.J.T. Reinders - Reviuwing Camera Calibration and Image Registration Techniques, Technical Report for Calibration and Registration, 1999.

[3] Zhang Z (1999), "Flexible Camera Calibration By Viewing a Plane From Unknown Orientations." *International Conference on Computer Vision (ICCV'99)*, Corfu, Greece, September 1999, pp. 666-673

[4] Heikkila, J.; Silven, O. (1997), "A four-step camera calibration procedure with implicit image correction," *Proc. IEEE Computer Society Conf.*, 1997 pp.1106 –1112.

[5] Abdel-Aziz, Y. I. & Karara, H. M. (1971), "Direct linear transformation into object space coordinates in close-range photogrammetry". *Proc. Symposium on Close-Range Photogrammetry*, Urbana, Illinois, pp. 1-18.

[6] Caltech Vision Group offical site: http://www.vision.caltech.edu/.

[7] Intel Corporation (2001), "Open Source Computer Vision Library." *Reference Manual,* http://www.intel.com/research/mrl/research/opencv/

[8] Trucco E., Verri A. (1998), "Introductory techniques for 3D Computer Vision", Prentice Hall, 1998.

[9] R. C. Gonzalez, R. E. Woods, "Digital image processing." pg. 148-156. *Addison-Wesley Publishing Company* 1992.

[10] J. Ramesh, R. Kasturi, B. G. Schunk, "Machine Vision", *McGraw-Hill, Inc.,* 1996.