NON-LINEAR OPTIMIZATION-BASED BATCH CALIBRATION WITH ACCURACY EVALUATION

Sidney Antonio A. Viana Jacques Waldmann Fábio de Freitas Caetano

Departamento de Sistemas e Controle – Divisão de Engenharia Eletrônica
Instituto Tecnológico de Aeronáutica
CTA-ITA-IEE-IEES 12.228-900 – São José dos Campos – SP
Fax:(012)347-5878 Tel.:(012)347-5993
sidney@ele.ita.cta.br, jacques@ele.ita.cta.br, caetano@ele.ita.cta.br

RESUMO O presente trabalho expõe a aplicação de uma técnica "batch" à calibração de um sistema de visão ativa, objetivando a identificação dos parâmetros que caracterizam a óptica (parâmetros intrínsecos), a posição e a atitude da câmera (parâmetros extrínsecos). Esta identificação permite a realização de dois propósitos fundamentais: a recuperação 3D, associada à localização de objetos no ambiente ou a determinação do movimento da câmera, assim como a correção de erros de imageamento devido a distorções ópticas e montagem imperfeita do conjunto câmera-lente. Inicialmente determina-se uma estimativa inicial dos parâmetros mediante o emprego de um modelo *pin-hole* dotado de óptica ideal, produzindo uma solução analítica baseada em mínimos quadrados. Em seguida, busca-se refinar este resultado iterativamente mediante o emprego de um modelo de distorção óptica radial. A validação dos resultados é feita com base em um índice que reduz o efeito de quantização espacial. Resultados da calibração usando dados sintéticos e reais são apresentados. Observou-se que erros de medida nos pontos 3D de controle e em suas projeções no plano-imagem influenciam detrimentalmente a calibração, impondo requisitos quanto à qualidade destas medidas. Uma avaliação estatística foi levada a cabo visando determinar intervalos de confiança para os parâmetros estimados a partir dos dados sintéticos. O valor calibrado da distância focal da câmera real foi usado com sucesso no rastreamento de objetos por movimentos sacádicos, indicando que a calibração efetuada apresenta exatidão adequada à esta tarefa visual.

Palavras-chave: calibração de câmeras, distorções ópticas, otimização, visão computacional.

ABSTRACT The application of a batch technique is presented for the calibration of an active vision head. The approach aims at the identification of both the intrinsic parameters related to the camera optics and the extrinsic parameters that describe the camera position and attitude. Successful identification allows one to attend to two fundamental purposes: firstly, the recovery of 3D structure, usually accompanied by object location in an unstructured environment or the detemination of camera motion, and secondly, the correction of imaging errors caused by optical distortions and misalignments in the lens-camera assembly. An initial estimate of the parameters is produced by assuming a pin-hole camera and ideal optics, thus leading to a closed-form least-squares solution. The latter is further refined by means of a radial distortion model of the actual optics. An accuracy index validates the results of the calibration method. This index is robust to spatial quantization errors that occur in the imaging process and thus accomodates a comparison of calibration performance in distinct vision systems. Results based on synthetic and real data are presented. Measurement errors in both 3D control points and their projections onto the image plane detrimentally affect the calibration accuracy. Requirements upon the quality of such measurements are stated. A statistical evaluation was carried out to determine the average performance and confidence intervals for the parameter estimation from synthethic data. The calibrated focal length has been used in conjunction with a saccadic motion control algorithm for visual tracking. Successful completion of tracking under various circumstances indicates that the calibration accuracy is adequate for this specific visual task.

Keywords: camera calibration, computer vision, optical distortions, optimization.

1 INTRODUCTION

Camera calibration has two important purposes to attend to: (1) to recover the 3D structure from 2D images, thus allowing the location of objects in space or the determination of camera position and attitude relative to a fixed reference system; and (2) to correct imaging errors caused by lens distortion and misalignments in the camera-lens assembly. The former is useful for autonomous navigation and guidance whereas the latter finds applications in the field of photogrametry and vision-based metrology. An analysis covering a diversity of methods is found in Tsai (1989) and in Weng *et alii* (1992). It is stressed that batch calibration methods use all the measurements available to produce an off-line identification of system parameters, as opposed to recursive methods which on-line improve the identification results as new measurements are acquired during the system operation. Calibration methods are further classified in three categories:

1. Closed-form solution methods: A non-linear function of the original parameters is employed to yield a new

identification problem which is linear on the transformed parameters. Least-squares is used and the original parameters are recovered via inverse mapping. However, the inclusion of optical distortion models is rather cumbersome in this class of methods.

2. Direct optimization methods: Parameters are iteratively estimated via optimization of the residual errors that relate the 3D control points and the corresponding projections onto the image plane under the estimated parameters. Refined camera and optical distortion models are easily accounted for but an adequate initial solution is called for to circumvent local minima or even divergence.

3. Two-step methods: A composition of the former two classes. Typically, a closed-form solution for some or all of the parameters is used as an initial guess for the iterative optimization. The approach proposed in Weng *et alii* (1992) comprises a closed-form solution for all parameters which are then iteratively improved in the optimization phase in which optical distortion is accounted for.

The literature manifests a constant concern regarding the attempt to accommodate two contradictory aspects: the simplicity of the method *vis-à-vis* the resulting accuracy. Numerical instability arises as one adds to the complexity of the lens model. Therefore, the complexity of the distortion model and that of the selected method should match the required accuracy for the desired visual task as well as the available computational resources. In photogrametry and vision-based metrology, for instance, strict requirements are imposed upon the calibration accuracy of intrinsic parameters such as focal length , image center coordinates and scale factors. Vision-based navigation aided by inertial sensors, on the other hand, gives way to the relaxation of requirements regarding the calibration of extrinsic parameters such as camera orientation and attitude.

One resorts to simpler methods when dealing with applications in which most of the intrinsic parameters remain fixed whereas focal length and the extrinsic parameters, for instance, are allowed to change. In Wang and Tsai (1991) a closed-form solution for such a case is presented with claims that satisfactory results have been achieved in autonomous guidance. A simpler version of a direct method is applied in Abidi and Chandra (1995) to the real-time calibration of a steerable autofocus camera in which four 3D control points lying on a tetrahedron are used. In Chaterjee and Roychowdhury (1993) a closed-form solution is proposed which does not account for optical distortion and errors in both the 3D control points and the image plane measurements. No claims are stated, however, regarding its accuracy and robustness in real applications. A non-iterative method for the calibration of the scale factor is reported in Penna (1991) claiming adequate results . As for optical distortion, it is emphasized in Jackowski and Goshtasby (1997) how crucial its identification is and a compensation technique is proposed for the purpose of pattern classification in satellite images that contain land, forest and water. Distortion is then distinguished according to two main causes: one originates in the non-linear optics and assembly misalignments whereas the other arises from chromatic aberration due to photosensor varying efficiency when subject to changing wavelength and illumination.

The impact of measurement errors, camera position and attitude, the characteristics of the imaging apparatus and the number of images on the accuracy of a calibration method is generally evaluated by an index that is usually based on statistics such as the standard deviation of the estimated parameters or the RMS value of the errors in the image-plane projections (Weng et alii, 1992; Tu and Dubuisson,1992; Li and Lavest 1996). An initial evaluation of accuracy often makes use of virtual camera with selected parameter values and of synthetic data. The application of a calibration method to a real vision apparatus, nevertheless, presents a challenge to accuracy evaluation since the actual parameter values are not known and thus one should utilize image-plane projection errors. However, such projections are affected by error sources that are not due to limitations of the calibration method but rather caused either by limitations of the vision system, such as spatial quantization or the quality of the measurements of the 3D control points.

The calibration procedure presented here is based on the two-step metod (Weng *et alii*, 1992). It has been motivated by the development of a system for the visual tracking of moving objects at the ITA-INPE Active Computer Vision and Perception Laboratory (ACVPL). Off-line identification of focal length is required because the selected approach to visual tracking is based on the compensation of background motion induced by the camera motion. Image regions whose motion is inconsistent with that of the camera are segmented as moving objects (Murray and Basu, 1994) . The structure of the paper is as follows. Section 2 describes the vision system and camera-lens model, first assuming ideal optics and then the inclusion of optical distortions. Section 3 presents the methodology of calibration and some accuracy indexes for its evaluation. Implementation and analysis of results obtained from both synthetic and real data are exposed in Section 4. Finally, Section 5 presents the conclusions and suggestions for future work.

2 THE VISION SYSTEM AND THE CAMERA-LENS APPARATUS MODEL

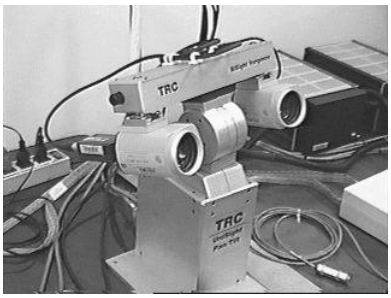
2.1 The Vision System

The vision system is composed of a Helpmate Robotics BiSight Vergence head equipped with monochromatic Hitachi KP-M1 cameras and servo-actuated Fujinon H10x11E-MPX31 lenses. The vision head has pan, tilt and assymetric vergence as its degrees of freedom. Aperture, zoom and focus are presently adjusted off-line acording to the requirements of the visual task. Lens parameters are adjustable by means of two AD/DA signal acquisition boards which communicate with the corresponding lens motor drives for the purpose of closed-loop control of focal length and focus. The operator can change system parameters, receive status information and issue off-line commands via the graphic interface. The vision head is depicted in Figure 1.

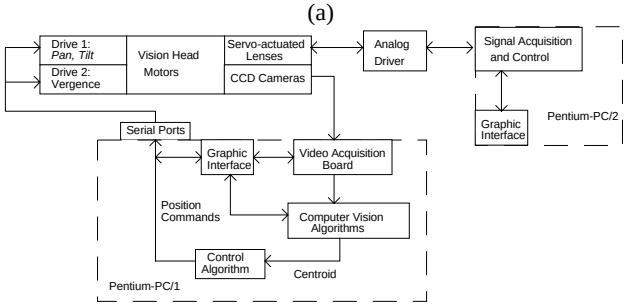
2.2 Modelling The Camera-Lens Assembly

2.2.1 Ideal Optics

A pin-hole camera model is used. The representation of 3D points and the respective projections on the image plane utilize four reference frames, as depicted in Figure 2.



(a)



(b)

Figure 1: (a) The Otelo vision head. (b) The active vision system.

Let $[x_w \;\; y_w \;\; z_w]^T$ denote the representation of a point **P** in 3D space according to the $\{O_w x_w y_w z_w\}$ reference frame which is fixed. Let $[x_c \;\; y_c \;\; z_c]^T$ represent **P** relative to the $\{O_c x_c y_c z_c\}$ reference frame which is fixed to the camera and moves along with it. Its origin O_c is assumed coincident with the center of projection and likewise axis z_c to coincide with the optical axis. The image plane containing the CCD array is considered to be perpendicular to the optical axis, parallel to the $\{x_C y_C\}$ plane and displaced from it by the length *f*. Reference frames $\{O_w x_w y_w z_w\}$ and $\{O_c x_c y_c z_c\}$ are related by:

$$[x_c \;\; y_c \;\; z_c]^T = \mathbf{R}[x_w \;\; y_w \;\; z_w]^T + \mathbf{T} \tag{2.1}$$

$\mathbf{R} = (r_{ij})_{3\times 3}$ is the rotation matrix that rotates from reference frame $\{O_w x_w y_w z_w\}$ to a reference frame parallel to $\{O_c x_c y_c z_c\}$ and $\mathbf{T} = [t_x \;\; t_y \;\; t_z]^T$ is the translation vector that takes the origin from the former to the origin of the latter. **R** and **T** are the extrinsic parameters which define camera position and attitude in the 3D space. Two reference frames are further defined on the image plane: {*O'uv*} and {*rc*}, the latter with its origin at the top left corner of the image. *O'* represents the principal point and {*uv*} axes are parallel to {*rc*} and $\{x_c y_c\}$. Pixel location is supplied by the image acquisition system in terms of (*r*,*c*) coordinates, with (r_0,c_0) denoting the image center where the principal point *O'* lies. The following imaging relationships are based on this model:

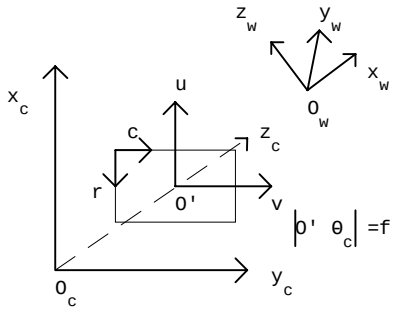


Figure 2: Reference frames.

$$u = f\frac{x_c}{z_c} \;\; ; \;\; v = f\frac{y_c}{z_c} \tag{2.2}$$

$$r - r_0 = s_u u \;\; ; \;\; c - c_0 = s_v v \tag{2.3}$$

where s_u and s_v are scale factors that originate in the difference in scanning frequency between the CCD and the video acquisition board. Notice that s_u is negative and s_v is positive due to the relation between *O'uv* and *rc*. Images are captured in a 320×240 pixel resolution and the CCD dimensions are 0.88cm×0.66cm. The scaling factors are then $|s_u|=|s_v|=363.63$pixel/cm. They are both equal, thus representing square pixels. Combining equations (2.1),(2.2) and (2.3) yields the following relations between the camera parameters, the 3D control points and the corresponding projections onto an image plane that is normalized by focal length *f*:

$$\frac{u}{f} = \frac{r - r_0}{f_u} = \frac{r_{11}x_w + r_{12}y_w + r_{13}z_w + t_x}{r_{31}x_w + r_{32}y_w + r_{33}z_w + t_z} \tag{2.5a}$$

$$\frac{v}{f} = \frac{c - c_0}{f_v} = \frac{r_{21}x_w + r_{22}y_w + r_{23}z_w + t_y}{r_{31}x_w + r_{32}y_w + r_{33}z_w + t_z} \tag{2.5b}$$

$$f_u = f s_u \;\; ; \;\; f_v = f s_v \tag{2.6}$$

Equations (2.5) and (2.6) constitute the mathematical model of a pin-hole camera equipped with ideal optics. The intrinsic parameters are $[c_0, r_0, f_u, f_v]$. Since the scale factors are known, the estimation of *f* is straightforward.

2.2.2 The Optical Distortion Model

One reason for optical distortion is the occurrence of errors either in lens design and manufacture or in the assembly of the lens-camera apparatus. As a result, the projections on the image plane become distorted. Physically, one has:

$$u' = u + \delta_u(u,v) \;\; ; \;\; v' = v + \delta_v(u,v) \tag{2.7}$$

where *u'*,*v'* are the observed projections on the image plane and *u*,*v* are the non-observable ideal projections. $\delta_u(u,v), \delta_v(u,v)$ denote the projection errors due to optical distortion. Four types of optical distortion are concisely described as radial distortion, tangential distortion, misalignment distortion and prismatic distortion in Weng *et alii* (1992). In order to keep the model from becoming excessively complex, the present work considers lens radial distortion solely, depicted in Figure 3, and modelled according to:

$$\delta_u(u,v) = \alpha_1 u(u^2 + v^2); \qquad \delta_v(u,v) = \alpha_1 v(u^2 + v^2) \tag{2.8}$$

with α_1 denoting the radial distortion coefficient.

Equations (2.3) and (2.7) yield:

$$\frac{r - r_0}{s_u} = u' = u + \delta_u(u,v) \;\; ; \;\; \frac{c - c_0}{s_v} = v' = v + \delta_v(u,v) \tag{2.9}$$

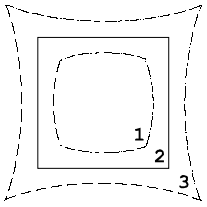


Figure 3: Effect of radial distortion on the imaging process (Weng *et alii* , 1992): (1) negative distortion; (2) no distortion; (3) positive

The image plane coordinates normalized by the focal length are:

$$\hat{u} = u'/f \quad ; \quad \hat{v} = v'/f \tag{2.10}$$

such that

$$\frac{u}{f} = \hat{u} - \frac{\delta_u(u,v)}{f} \quad ; \quad \frac{v}{f} = \hat{v} - \frac{\delta_v(u,v)}{f} \tag{2.11}$$

As u and v are non-observable, $\delta_u(u,v), \delta_v(u,v)$ above should be expressed in terms of the observable coordinates $\hat{u},\hat{v}$. Therefore (2.11) is rewritten as:

$$\frac{u}{f} = \hat{u} + \delta'_u(\hat{u},\hat{v}) \quad ; \quad \frac{v}{f} = \hat{v} + \delta'_v(\hat{u},\hat{v}) \tag{2.12a}$$

$$\delta'_u(\hat{u},\hat{v}) = k_1\hat{u}(\hat{u}^2+\hat{v}^2) = (k_1/f^3)u'(u'^2+v'^2) \tag{2.12b}$$

$$\delta'_v(\hat{u},\hat{v}) = k_1\hat{v}(\hat{u}^2+\hat{v}^2) = (k_1/f^3)v'(u'^2+v'^2) \tag{2.12c}$$

Assuming that the distortion at the observed position (u',v') is approximately the same as at the ideal position (u,v), i.e.:

$$\delta_u(u,v)/f \cong \delta_u(u',v')/f = \alpha_1 u'(u'^2+v'^2)/f \tag{2.13}$$

then from (2.11), (2.12) and (2.13) one obtains the normalized optical distortion factor

$$k_1 = -\alpha_1 f^2 \tag{2.14}$$

and from (2.5) and (2.12) results the camera model with optical distortion:

$$\hat{u} + k_1\hat{u}(\hat{u}^2+\hat{v}^2) = \frac{r_{11}x_w + r_{12}y_w + r_{13}z_w + t_x}{r_{31}x_w + r_{32}y_w + r_{33}z_w + t_z} \tag{2.15a}$$

$$\hat{v} + k_1\hat{v}(\hat{u}^2+\hat{v}^2) = \frac{r_{21}x_w + r_{22}y_w + r_{23}z_w + t_y}{r_{31}x_w + r_{32}y_w + r_{33}z_w + t_z} \tag{2.15b}$$

One proceeds to formulate the calibration problem as follows: given a set of 3D control points $[x_{w,i} \;\; y_{w,i} \;\; z_{w,i}]^T; i = 1,\ldots,n$ and the corresponding observed projections (r_i, c_i) in pixels, estimate in some optimal sense the intrinsic and extrinsic parameter vector and the optical distortion coefficient $\mathbf{m} = [r_0 \;\; c_0 \;\; f_u \;\; f_v \;\; \mathbf{T} \;\; \mathbf{R_1} \;\; \mathbf{R_2} \;\; \mathbf{R_3}]^T$ and $\mathbf{d}=[k_1]$, respectively. $\mathbf{R}_1$, $\mathbf{R}_2$ and $\mathbf{R}_3$ are column-vectors in the rotation matrix $\mathbf{R}$.

3 THE CALIBRATION METHODOLOGY

The two-step calibration methodology proposed here has been adapted from Weng *et alii* (1992) with some modifications to improve its robustness to 3D point measurement errors. Furthermore, it draws from this reference an accuracy index for the evaluation of method performance. This index includes a normalization by the image spatial resolution, thus allowing the comparison of the attained calibration accuracy in distinct vision systems. The first step is direct and non-iterative, consisting of a least-squares parameter estimation under the assumption of ideal optics, i.e., $\mathbf{d}=k_1=0$. This initial estimate, $\bar{\mathbf{m}}$, is then further refined in the second step which attempts to correct deviations of the image-plane projections caused by optical distortions and image digitization. Non-linear optimization is employed to find the optimal values ($\mathbf{m}^*$,$\mathbf{d}^*$) in a sense to be later defined. In order to avoid divergence or convergence to local minima, it shall be seen that the 3D control points and respective projections should be measured with adequate accuracy such that the initial estimate $\bar{\mathbf{m}}$ results in the vicinity of the global optimum. The contribution of the second step to the improvement of calibration accuracy depends fundamentally on the quality of the measurements and on the initial least-squares solution.

3.1 The Solution to the Calibration Problem

Let Ω represent the set of all 3D control points and ω the set of corresponding image points. The identification problem is defined as follows, where $F(\cdot,\cdot,\cdot,\cdot)$ is later defined.

$$\min_{\mathbf{m},\mathbf{d}} F(\Omega,\omega,\mathbf{m},\mathbf{d}) \tag{3.1}$$

Firstly, an initial estimate $\bar{\mathbf{m}}$ is sought via least-squares assuming ideal optics, i.e., $\mathbf{d}=\mathbf{0}$. However, $\mathbf{m}^*$ depends on knowledge of $\mathbf{d}$. Still assuming ideal optics, non-linear optimization of F is initiated taking into account only those 3D control points with their respective projections lying within some region around the image center. Optical distortion should not be significant there, thus yielding an improved estimate $\tilde{\mathbf{m}}$. A circular region with radius equal to a fourth of the square image side is employed in Weng *et alii* (1992). Here a rectangular region with adjustable dimensions has been used. The effect of varying the region dimensions upon calibration accuracy shall be discussed in Section 4.3. In the second step, the estimate $\tilde{\mathbf{m}}$ of $\mathbf{m}^*$ is kept fixed and F is minimized over $\mathbf{d}$, thus producing $\mathbf{d}_i^*$ which represents the i-th estimate of $\mathbf{d}^*$. $\mathbf{m}_i^*$ is subsequently estimated by keeping $\mathbf{d} = \mathbf{d}_i^*$ fixed and minimizing F over $\mathbf{m}$. This procedure is iterated k times yielding the estimates $\mathbf{m}_k^*$ and $\mathbf{d}_k^*$. This artificial decoupling between $\mathbf{m}$ and $\mathbf{d}$ is justified in Weng *et alii* (1992) as a means of circumventing convergence to local minima. An adequate initial estimate $\bar{\mathbf{m}}$ is required to produce adequate estimates $\mathbf{m}_k^*$ and $\mathbf{d}_k^*$. Our implementation of the method includes two modifications to the original algorithm:

1. The orthonormalization of rotation matrix $\mathbf{R}$ has been skipped since minor improvements have been observed which did not justify the additional computational workload;

2. Pruning of 3D control points which caused excessive error in its corresponding estimated projection on the image plane as the least-squares approach used for non-linear optimization is inherently sensitive to inconsistent data.

3.1.1 The First Step

The first step comprises an initial estimation of $\mathbf{m}$ via least-squares subject to $\mathbf{d}=\mathbf{0}$, resulting in the initial estimate $\bar{\mathbf{m}}$. It is followed by non-linear optimization of F over $\mathbf{m}$ subject to $\mathbf{d}=\mathbf{0}$ with $\bar{\mathbf{m}}$ as the initial guess, thus yielding $\tilde{\mathbf{m}}$. A detailed description follows.

3.1.1.1 PART I: Least-squares estimation of m subject to d=0

Under the assumption of ideal optics, $\mathbf{d}=\mathbf{0}$, and taking into account solely central points, the camera model (2.5) provides two equations for each 3D control point $[x_w \;\; y_w \;\; z_w]_i^T$ and its projection (r_i,c_i) on the image plane. The non-linear equations in the camera parameters are:

$$(r_i'-r_0)(r_{31}x_{w,i}+r_{32}y_{w,i}+r_{33}z_{w,i}+t_z)=f_u(r_{11}x_{w,i}+r_{12}y_{w,i}+r_{13}z_{w,i}+t_x)$$
$$(c_i'-c_0)(r_{31}x_{w,i}+r_{32}y_{w,i}+r_{33}z_{w,i}+t_z)=f_v(r_{21}x_{w,i}+r_{22}y_{w,i}+r_{23}z_{w,i}+t_y)$$

Each point provides two equations and there are 16 parameters to estimate. A mapping is then sought that transforms the above non-linear identification problem into a linear one in auxiliary parameters that can be solved via least-squares (Weng *et alii*, 1992). The transformation is:

$$\mathbf{W}_1=f_u\mathbf{R}_1+r_0\mathbf{R}_3\ ;\ \mathbf{W}_2=f_v\mathbf{R}_2+c_0\mathbf{R}_3\ ;\ \mathbf{W}_3=\mathbf{R}_3 \tag{3.2a}$$

$$w_4=f_ut_x+r_0t_z\ ;\ w_4=f_ut_x+r_0t_z\ ;\ w_6=t_z \tag{3.2b}$$

where $\mathbf{R}_1$, $\mathbf{R}_2$ and $\mathbf{R}_3$ are the columns of $\mathbf{R}$. The resulting $2n$ equations that relate to the n central points can be expressed as $\mathbf{AW}=\mathbf{0}$, where $\mathbf{W}^{\mathrm{T}}=[\mathbf{W}_1\ \mathbf{W}_2\ \mathbf{W}_3\ w_4\ w_5\ w_6]^{\mathrm{T}}$ is the vector of 12 auxiliary parameters and

$$\mathbf{A}=\begin{bmatrix} -x_{w1} & -y_{w1} & -z_{w1} & 0 & 0 & 0 & r_1'x_{w1} & r_1'y_{w1} & r_1'z_{w1} & -1 & 0 & r_1' \\ 0 & 0 & 0 & -x_{w1} & -y_{w1} & -z_{w1} & c_1'x_{w1} & c_1'y_{w1} & c_1'z_{w1} & 0 & -1 & c_n' \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -x_{wn} & -y_{wn} & -z_{wn} & 0 & 0 & 0 & r_n'x_{wn} & r_n'y_{wn} & r_n'z_{wn} & -1 & 0 & r_n' \\ 0 & 0 & 0 & -x_{wn} & -y_{wn} & -z_{wn} & c_n'x_{wn} & c_n'y_{wn} & c_n'z_{wn} & 0 & -1 & c_n' \end{bmatrix} \tag{3.3}$$

Among all possible solutions of the above homogeneous linear system, the aequate solution should satisfy the following requirements: (1) the 2-norm of $\mathbf{W}_3$ must equal unity, since it is a column of the orthonormal rotation matrix $\mathbf{R}$; and (2) the sign of $w_6=t_z$ must be compatible with the location of the origin of the camera reference frame $\{O_c x_c y_c z_c\}$ relative to the world reference frame $\{O_w x_w y_w z_w\}$, having in mind that the translation vector is represented in the camera reference frame. The homogeneous system can be rewritten as a nonhomogeneous one by temporarily imposing $w_6=t_z=1$ thus yielding:

$$\mathbf{A'W'}+\mathbf{B}=\mathbf{0} \tag{3.4}$$

where $\mathbf{A'}$ is a matrix with the eleven first columns of $\mathbf{A}$; $\mathbf{B'}$ is the last column of $\mathbf{A}$ and $\mathbf{W'}$ is vector $\mathbf{W}$ correspondingly reduced of $w_6=t_z=1$. The least-squares solution for $\mathbf{W'}$ in (3.4) is then obtained. In order to enforce the aforementioned requirements, the following normalization is carried out (Weng *et alii*, 1992):

$$\mathbf{S}=\begin{bmatrix}\mathbf{S}_1\\ \mathbf{S}_2\\ \mathbf{S}_3\\ s_4\\ s_5\\ s_6\end{bmatrix}=\pm\frac{1}{\|\mathbf{W}_3\|}\begin{bmatrix}\mathbf{W}_1\\ \mathbf{W}_2\\ \mathbf{W}_3\\ w_4\\ w_5\\ 1\end{bmatrix} \tag{3.5}$$

and the parameters $\overline{\mathbf{m}}=[r_0\ c_0\ f_u\ f_v\ t_x\ t_y\ t_z\ \mathbf{R}_1\ \mathbf{R}_2\ \mathbf{R}_3]^{\mathrm{T}}$ are determined according to:

$$r_0=\mathbf{S}_1{}^{\mathrm{T}}\mathbf{S}_3\ ;\ c_0=\mathbf{S}_2{}^{\mathrm{T}}\mathbf{S}_3\ ;\ f_u=-\|\mathbf{S}_1-r_0\mathbf{S}_3\|\ ;\ f_v=\|\mathbf{S}_2-c_0\mathbf{S}_3\| \tag{3.6a}$$

$$t_x=(s_4-r_0s_6)/f_u\ ;\ t_y=(s_5-c_0s_6)/f_v\ ;\ t_z=s_6 \tag{3.6b}$$

$$\mathbf{R}_1=(\mathbf{S}_1-r_0\mathbf{S}_3)/f_u\ ;\ \mathbf{R}_2=(\mathbf{S}_2-c_0\mathbf{S}_3)/f_v\ ;\ \mathbf{R}_3=\mathbf{S}_3 \tag{3.6c}$$

3.1.1.2 PART II: Non-linear optimization of the initial estimate $\overline{\mathbf{m}}$ subject to d=0

Let Ω denote the set of 3D control points, $\omega=[u_1\ v_1\ u_2\ v_2\ \cdots\ u_n\ v_n]^{\mathrm{T}}$ the set of the respective ideal projections of the 3D control points on the image plane and $\omega'=\omega+N$ the set of observed projections. Let the actual imaging process be described by a non-linear function $f(\mathbf{m},\mathbf{d})$. Assuming that $\mathbf{d}$ is known and letting $\mathbf{m}=\overline{\mathbf{m}}$, the observed projections on the image plane are initially predicted as $\overline{\omega}=f(\overline{\mathbf{m}},\mathbf{d})$. Linearizing $f(\cdot,\cdot)$ with relation to $\mathbf{m}$ at $\mathbf{m}=\overline{\mathbf{m}}$ yields:

$$\omega'=f(\mathbf{m},\mathbf{d})\cong f(\overline{\mathbf{m}},\mathbf{d})+(\mathbf{m}-\overline{\mathbf{m}})\frac{\partial f(\mathbf{m},\mathbf{d})}{\partial\mathbf{m}}\bigg|_{\mathbf{m}=\overline{\mathbf{m}}}=\overline{\omega}+\mathbf{J}(\mathbf{m}-\overline{\mathbf{m}}) \tag{3.7}$$

with $\mathbf{J}$ as the Jacobian of $f(\cdot,\cdot)$. The minimum variance estimator in the vicinity of $\overline{\mathbf{m}}$ is then given by $\tilde{\mathbf{m}}=\min_{\mathbf{m}} E[\|\mathbf{m}-\overline{\mathbf{m}}\|_2^2]$. Assuming N as zero-mean uncorrelated observation noise, it is claimed that, for $\mathbf{d}=\mathbf{0}$, the above cost function to be minimized is equivalent to the sum of squared errors (SSE):

$$F(\Omega,\omega',\mathbf{m},\mathbf{d}=\mathbf{0})=E[\|\mathbf{m}-\overline{\mathbf{m}}\|_2^2]\big|_{\mathbf{d}=\mathbf{0}}=$$

$$=\alpha\sum_{i=1}^{n}\{[r'_i-r_i(\mathbf{m},\mathbf{d}=\mathbf{0})]^2+[c'_i-c_i(\mathbf{m},\mathbf{d}=\mathbf{0})]^2\} \tag{3.8}$$

where α is a constant, $(r_i(\mathbf{m},\mathbf{d}),c_i(\mathbf{m},\mathbf{d}))$ are the estimated projections of the 3D control points on the image plane according to the imaging function $f(\mathbf{m},\mathbf{d})$ and (r'_i,c'_i) are the measured projections, all within the central region. The optimization procedure determines the new estimate $\tilde{\mathbf{m}}$ which minimizes $F(\cdot,\cdot,\cdot,\cdot)$ subject to $\mathbf{d}=\mathbf{0}$, with $\overline{\mathbf{m}}$ as initial guess:

$$\tilde{\mathbf{m}}=\min_{\mathbf{m}}\left\{\sum_{i=1}^{n}\{[r'_i-r_i(\mathbf{m},\mathbf{d}=\mathbf{0})]^2+[c'_i-c_i(\mathbf{m},\mathbf{d}=\mathbf{0})]^2\}\right\} \tag{3.9}$$

This is a non-linear optimization problem due to the non-linear imaging function $f(\mathbf{m},\mathbf{d})$.

3.1.2 The Second Step

The second step, which is iterative, comprises three parts, namely: 1) least-squares estimation of $\mathbf{d}=\mathbf{d}_j^*$, subject to $\mathbf{m}=\mathbf{m}_{j-1}^*$, $\mathbf{m}_0^*=\tilde{\mathbf{m}}$; (2) elimination of image points with excessive error; and (3) estimation of $\mathbf{m}=\mathbf{m}_j^*$ subject to $\mathbf{d}=\mathbf{d}_j^*$. This step utilizes all 3D control points that have not been eliminated due to excessive projection errors.

3.1.2.1 PART I: Least-squares estimation of d subject to $\mathbf{m}=\mathbf{m}_{j-1}^*$, $\mathbf{m}_0^*=\tilde{\mathbf{m}}$

The estimation of $\mathbf{d}$ utilizes the camera model in equations (2.5) to (2.15). This model is linear with relation to the optical radial distortion k_1, and therefore one is to solve a linear identification problem via least-squares. The cost function to optimize is again the SSE defined in (3.8). According to the camera model, the error in the i-th estimated projection on the image plane (in pixels), i=1,...,n denoting the n control points, is:

$$r_i(\mathbf{m},\mathbf{d}) - r'_i = r_0 + f_u\left[\frac{r_{11}x_{w,i} + r_{12}y_{w,i} + r_{13}z_{w,i} + t_x}{r_{31}x_{w,i} + r_{32}y_{w,i} + r_{33}z_{w,i} + t_z} - k_1\hat{u}_i(\hat{u}_i^2 + \hat{v}_i^2)\right] - r'_i \tag{3.10a}$$

$$c_i(\mathbf{m},\mathbf{d}) - c'_i = c_0 + f_v\left[\frac{r_{21}x_{w,i} + r_{22}y_{w,i} + r_{23}z_{w,i} + t_y}{r_{31}x_{w,i} + r_{32}y_{w,i} + r_{33}z_{w,i} + t_z} - k_1\hat{v}_i(\hat{u}_i^2 + \hat{v}_i^2)\right] - c'_i \tag{3.10b}$$

Each image point provides two equation that are linear in k_1. Analogously to Section 3.1.1.1 the above equations can be rewritten as **Qd**+**C**=**0** with **Q** and **C** defined as:

$$\mathbf{Q}_{(2n\times 1)} = \begin{bmatrix} \hat{u}_1(\hat{u}_1^2 + \hat{v}_1^2) \\ \hat{v}_1(\hat{u}_1^2 + \hat{v}_1^2) \\ \vdots \\ \hat{u}_n(\hat{u}_n^2 + \hat{v}_n^2) \\ \hat{v}_n(\hat{u}_n^2 + \hat{v}_n^2) \end{bmatrix} \quad \mathbf{C}_{(2n\times 1)} = \begin{bmatrix} \frac{r_0 - r'_1}{f_u} - \frac{r_{11}x_{w,1} + r_{12}y_{w,1} + r_{13}z_{w,1} + t_x}{r_{31}x_{w,1} + r_{32}y_{w,1} + r_{33}z_{w,1} + t_z} \\ \frac{c_0 - c'_1}{f_v} - \frac{r_{21}x_{w,1} + r_{22}y_{w,1} + r_{23}z_{w,1} + t_y}{r_{31}x_{w,1} + r_{32}y_{w,1} + r_{33}z_{w,1} + t_z} \\ \vdots \\ \frac{r_0 - r'_n}{f_v} - \frac{r_{11}x_{w,n} + r_{12}y_{w,n} + r_{13}z_{w,n} + t_x}{r_{31}x_{w,n} + r_{32}y_{w,n} + r_{33}z_{w,n} + t_z} \\ \frac{c_0 - c'_n}{f_v} - \frac{r_{21}x_{w,n} + r_{22}y_{w,n} + r_{23}z_{w,n} + t_y}{r_{31}x_{w,n} + r_{32}y_{w,n} + r_{33}z_{w,n} + t_z} \end{bmatrix} \tag{3.11}$$

At the j-th iteration the solution that minimizes $\|\mathbf{Qd} + \mathbf{C}\|_2^2$ subject to $\mathbf{m} = \mathbf{m}^*_{j-1}$, $\mathbf{m}^*_0 = \tilde{\mathbf{m}}$ yields

$$\mathbf{d}^*_j = \min_{\mathbf{d}}\left\{\sum_{i=1}^{n}\left\{[r'_i - r_i(\mathbf{m}^*_{j-1},\mathbf{d})]^2 + [c'_i - c_i(\mathbf{m}^*_{j-1},\mathbf{d})]^2\right\}\right\} \tag{3.12}$$

3.1.2.2 PART II: Elimination of image-points with excessive error in their the estimated projections

The measured projections (r'_i, c'_i) on the image plane are compared to the estimated projections $(r_i(\mathbf{m}^*_{j-1}, \mathbf{d}^*_j), c_i(\mathbf{m}^*_{j-1}, \mathbf{d}^*_j))$. 3D control points giving rise to an estimation error above a selected threshold are discarded in order to reduce the detrimental impact of inconsistent data in the measurement sets Ω and ω'.

3.1.2.3 PART III: Non-linear optimization of m subject to $\mathbf{d} = \mathbf{d}^*_j$

This is analogous to equation (3.9) in Section 3.1.1.2 except for the modified constraint $\mathbf{d} = \mathbf{d}^*_j$ and initial guess $\mathbf{m}^*_{j-1}$, thus resulting the estimate $\mathbf{m} = \mathbf{m}^*_j$. The overall calibration algorithm is as follows.

1. Given the 3D control points and the corresponding image-plane projections, consider only those with projections lying within the central region;
2. Obtain the least-squares solution of the linear estimation problem defined in (3.2), (3.3), (3.4) and the estimate $\overline{\mathbf{m}}$ from equations (3.5) and (3.6);
3. Refine the above estimate via the non-linear optimization in equation (3.9) with $\overline{\mathbf{m}}$ as the initial guess and **d**=**0** yielding the estimate $\tilde{\mathbf{m}}$;
4. Consider now all 3D control points and the corresponding image-plane projections;
5. $j \leftarrow 1$, $\mathbf{m}^*_0 = \tilde{\mathbf{m}}$, select a pruning threshold, iterate k times;
6. Obtain the least-squares solution of the linear estimation problem defined in (3.11) subject to $\mathbf{m} = \mathbf{m}^*_{j-1}$ producing the estimate $\mathbf{d}^*_j$;
7. Prune the 3D control points which give rise to projection errors above threshold;
8. Refine estimate $\mathbf{m}^*_{j-1}$ via the non-linear optimization in equation (3.9) with $\mathbf{m}^*_{j-1}$ as the initial guess and subject to $\mathbf{d} = \mathbf{d}^*_j$ yielding the estimate $\mathbf{m}^*_j$;
9. $j \leftarrow j+1$;

3.2 Accuracy Evaluation

In Tsai (1987) accuracy evaluation has been carried out based on the 3D reconstruction error. Generally, some particular criterion for accuracy evaluation is proposed which, depending on the vision system and experimental setup, arises difficulties regarding the comparison of results of calibration algorithms that have been implemented in distinct vision systems. The proper formulation of an adequate accuracy index is quite convenient. Accuracy evaluation based solely on the grounds of 3D reconstruction error is unfortunately affected by the capabilities of the underlying vision system and the quality of the measurement equipment and experimental setup. Among such factors one should notice the spatial resolution of image acquisition and errors in the mesurement of 3D control points and respective projections on the image plane. Let the pixels be backprojected on the scenery such that to each backpropagated pixel corresponds a certain area of the imaged surface at a given depth z_C (Figure 4). The area denotes the position uncertainty of a 3D point at depth z_C due to the finite resolution of the image.

Letting a and b represent the dimensions of the backpropagated pixel upon the object plane perpendicular to the optical axis at depth z_c relative to the center of projection, the following relations hold:

$$\frac{a\,[cm]}{z_c[cm]} = \frac{1\,[pixel]}{|f_u|\,[pixel]} \;;\; \frac{b\,[cm]}{z_c[cm]} = \frac{1\,[pixel]}{|f_v|\,[pixel]} \tag{3.13}$$

The vertical digitization error ε_v (in centimeters) of a pixel backprojected to depth z_c may be modelled as a random variable with uniform probability density in the interval $[-a/2; a/2]$, thus yielding the corresponding mean $\mu_{\varepsilon v} = 0$ [cm] and variance $\sigma^2_{\varepsilon v} = a^2/12$ [cm^2]. Proceeding likewise for the horizontal digitization error ε_h yields the mean $\mu_{\varepsilon h} = 0$ [cm] and variance $\sigma^2_{\varepsilon h} = b^2/12$ [cm^2]. The total uncertainty in the 3D position of a point at a distance z_c from the camera due to image digitization is modelled as $\varepsilon = \sqrt{\varepsilon_h^2 + \varepsilon_v^2}$ [cm], i.e. the vertical and horizontal errors are uncorrelated. $\sigma^2_\varepsilon = E[\varepsilon^2] = E[\varepsilon_v^2] + E[\varepsilon_h^2] = \sigma^2_{\varepsilon v} + \sigma^2_{\varepsilon h}$ [cm^2] is thus the variance of the total 3D uncertainty at depth z_c and can be rewritten as:

$$\sigma^2_\varepsilon = \frac{(a^2 + b^2)}{12} = \frac{z_c^2}{12}\left(\frac{1}{f_u^2} + \frac{1}{f_v^2}\right) \tag{3.14}$$

The Normalized Calibration Error (NCE) is an accuracy index defined as the RMS value of the ratio between the backprojected position error at depth z_C of a given 3D point and its corresponding total uncertainty

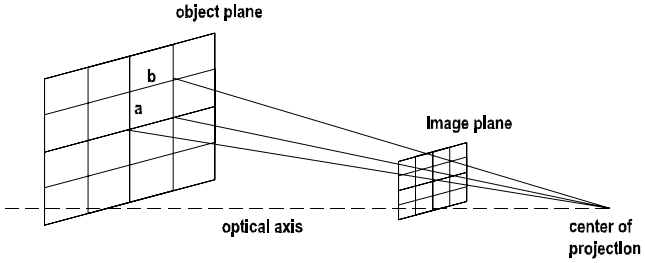


Figure 4: Backprojection on the object of a pixel on the image plane.

as given in (3.14) (Weng *et alii*, 1992). The NCE is therefore expressed as:

$$\mathrm{NCE}=\left(\frac{1}{n}\sum_{i=1}^{n}\frac{12[(x_{c,i}-\hat{x}_{c,i})^2+(y_{c,i}-\hat{y}_{c,i})^2]}{z_{c,i}^2(f_u^{-2}+f_v^{-2})}\right)^{1/2} \tag{3.15}$$

where $[x_{c,i}\ \ y_{c,i}\ \ z_{c,i}]^{\mathrm{T}}$, i=1,..., n denote the measured i-th 3D point coordinates represented in the camera reference frame and $[\hat{x}_{c,i}\ \ \hat{y}_{c,i}\ \ z_{c,i}]^{\mathrm{T}}$ are the backpropagated estimates at depth $z_{c,i}$ computed according to the estimated camera parameters. The transformation from the world reference frame where the 3D control points are originally represented to the camera reference frame defined in (2.1) employs the estimated extrinsic parameters **T** and **R**. The NCE index is useful for evaluating the calibration accuracy of a monocular system because it does not require the reconstruction of the 3D control points. This index is robust to image spatial resolution and the distance to the 3D control points, since they affect equally both numerator and denominator of (3.15). NCE ≤ 1 indicates that quantization errors are more significant than the calibration errors on the backpropagated position. In this case the calibration method successfully extracts the useful information embedded in the measurement set up to the limit imposed by image resolution. The effect of errors in the measurement of 3D control points on calibration accuracy is then negligible as long as the error magnitude remains inferior to that of the backprojected digitization error at depth z_c. On the other hand, NCE >> 1 signals a poor calibration in which errors in the estimated parameters result in backprojected estimates far more incorrect than the expected effect of backprojected digitization errors.

An additional calibration accuracy index is the SSE between measured image points and estimated image points according to the identified camera parameters. The image-plane projection error SSE is given by Weng *et alii* (1992):

$$\mathrm{SSE}=\sum_{i=1}^{n}\left([r'_i-r_i(\mathbf{m}^*,\mathbf{d}^*)]^2+[c'_i-c_i(\mathbf{m}^*,\mathbf{d}^*)]^2\right) \tag{3.16}$$

Finally, another accuracy index is the comparison between the theoretical standard deviation of the image-plane projection error μ and its estimate μ' for a normalized image plane with focal length f = 1 cm. The variance σ_ε^2 defined in (3.14) consists of the location uncertainty of a 3D point at depth z_c from the camera due to the finite resolution of the image. Therefore one has the adimensional standard deviation of the projection error at depth f = 1 cm as $\mu_0=(f_u^{-2}+f_v^{-2}/12)^{1/2}$. Figure 5 shows that to each pixel corresponds infinite 3D points within the backprojection volume of that pixel. Therefore, minute 3D measurement errors will be masked by the image resolution. The quality of 3D measurements is as critical for adequate calibration accuracy as that of the image-plane measurements (Li and Lavest, 1996). For the latter, subpixel accuracy is often recommended and implemented by means of image processing for the detection and location of the relevant image points. For image points provided with subpixel accuracy by image processing techniques $\mu=\mu_0/5$ is proposed as a more realistic adimensional estimate of the standard deviation of the positional error on the normalized image plane (Weng *et alii*, 1992).

From the estimates $\mathbf{m}^*$ and $\mathbf{d}^*$, the estimate μ' of μ is:

$$\mu'=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{r'_i-r_i(\mathbf{m}^*,\mathbf{d}^*)}{f_u}\right)^2+\left(\frac{c'_i-c_i(\mathbf{m}^*,\mathbf{d}^*)}{f_v}\right)^2\right]} \tag{3.17}$$

An accurate calibration yields $\mu'(\mathbf{m}^*,\mathbf{d}^*)$ close to μ. The required accuracy for the 3D measurements is directly related to image resolution, which is dictated by pixel size, to the field of view, which depends on focal length, and to the distance between the object and the camera. From Figure 4 and assuming that the distance to the object is kept constant, the improvement of image resolution requires smaller pixels if the focal length remains unchanged. On the other hand, if the pixel size remains constant, an increase in focal length produces a smaller field of view per pixel and therefore the 3D coverage per pixel results in a more detailed image (zoom-in effect). One concludes that the 3D measurement accuracy should match the image resolution dictated by pixel size and the 3D coverage defined by the focal length in order to attenuate the detrimental effects on calibration accuracy. As the distance to the object z_c increases, the backpropagated image digitization error at such distance increases proportionally and as a result the calibration method will be more compliant to 3D measurement errors. For instance, at z_c = 120 cm, image resolution 320×240 pixels, CCD size 0.88cm×0.66cm and with f = 1.1 cm, the acceptable 3D measurement error magnitude in the vertical direction as indicated by (3.13) is at most:

$$a=\frac{z_c}{|f_u|}=\frac{z_c}{f|s_u|}=\frac{120}{(1.1)(240)/(0.66)}=0.3\ \mathrm{cm} \tag{3.18}$$

Since the focal length in pixels is the same in both directions the same value stands as the maximum acceptable horizontal error magnitude in 3D measurement. One concludes as well that the larger the focal length the higher the sensitivity to 3D measurement errors.

4 IMPLEMENTATION AND RESULTS

A rectangular grid pattern has been used for the selection of the 3D control points as depicted in Figure 5(a). The vision head was positioned in such a way that the camera optical axis was perpendicular to the grid. Four stations were defined for the grid, with a 10 cm distance between each one. The origin of the world reference frame $\{O_w x_w y_w z_w\}$ was fixed at the closest station from the camera with plane $\{O_w x_w y_w z_w\}$ parallel to the grid. The rotation matrix **R** in this setup was close to identity ($\mathbf{R}\approx\mathbf{I}_3$). The dimension of the grid cells have been measured and their vertices used as 3D control points. Station 1 was positioned around 120 cm from the camera. Images acquired at the four stations were used to provide depth information to the set of control points. Image-plane projections were manually selected as those pixel coordinates that corresponded to sharp vertices. The set of image-plane projections was composed of 164 pixel positions with resolution 320×240 as shown in Figure 5(c). Figure 5(d) shows the 3D control points.

4.1 Calibration with Synthetic Data

Calibration with synthetic data allows the initial evaluation of the algorithm accuracy as the parameters of a virtual camera are perfectly known. The procedure is suitable for the preliminary analysis of the impact of optical distortion and image digitization upon calibration accuracy. It consists of the following steps: (1) determine the virtual image-plane projections of the 3D points using the known parameters of the virtual camera; (2) estimate the camera parameters from the 3D measurements and respective image-plane projections; (3) compare the calibrated parameters with the actual values; and

(4) project the 3D points on the virtual image-plane using the estimated parameters and compare with the results of step 1. Simulation of optical distortion and image digitization may be included in step 1.

Virtual image resolution was selected as 320×240 with ideal optics and camera-lens assembly thus resulting the image principal pont (r_0,c_0) at the geometrical center of the image plane. Ground-truth values for the extrinsic parameters **T** and **R** were selected such that the resulting image points resembled those in the actual images. Ground-truth world and camera reference frames were assumed as perfectly aligned. The following ground-truth values for the virtual camera were used: principal point (r_0,c_0)=(120,160) pixels, focal lengths $(f_u,f_v)=(-2s_u,2s_v)$ pixels, $\mathbf{R}=\mathbf{I}_3$, $\mathbf{T}=[-11\ 8\ 130]^T$, k_1=0.05, the latter being used according to circumstances as follows. The calibration method was tested under three circumstances: (1) ideal optics and no digitization errors, (2) optical distortion by use of k_1 and no digitization error, and (3) optical distortion and digitization errors. The digitization model employed is based on rounding off the pixel coordinates of an image-plane projection to the nearest integer value. For the virtual camera one has $\mu\approx 2.25\times10^{-4}$.

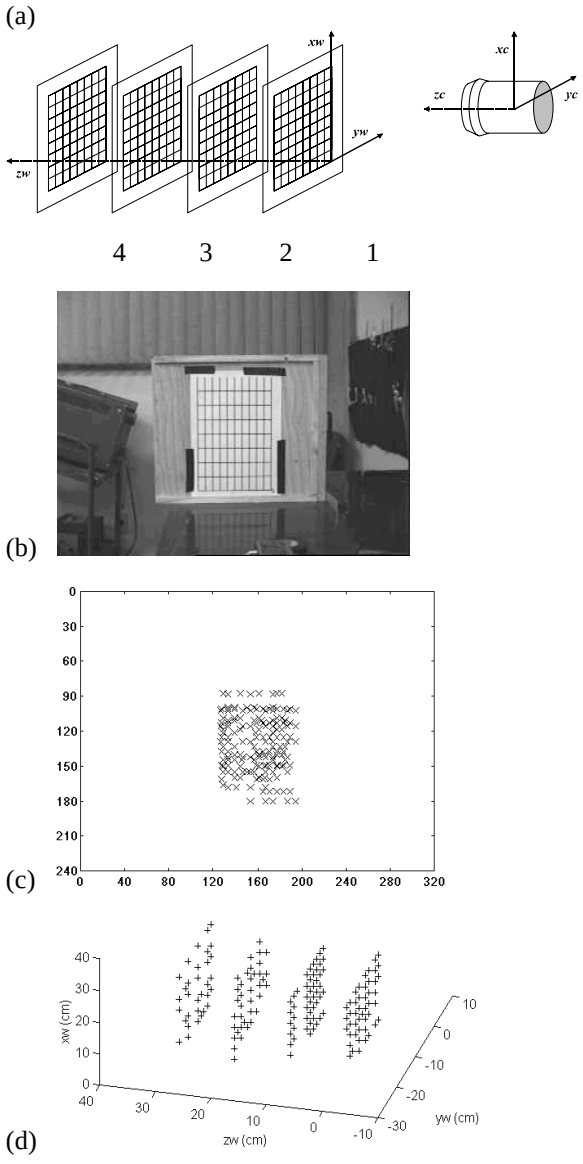


Figure 5: (a) Experimental setup scheme. (b) Image acquired of grid at station 1. (c) Actual image-plane projections. (d) 3D control points in the world

The results related to ideal optics and no image digitization are discussed in the following. A rectangular central region sized 1/4 of the image dimensions was used yielding 69 central image points. The solution achieved by the end of Part I in step 1 was perfect. Part II of step 1 of the algorithm iterated once with this solution as the initial guess yielding practically the same solution and moreover SSE = μ' = 0. The use of perfect information thus resulted in exact estimates except for $\hat{k}_1 = -6\times10^{-6}$. Other ground-truth values for the camera parameters were tried out also yielding perfect results.

In more realistic conditions, however, measurements are expected to become corrupted by noise and parts of the scenario to get out of focus because of finite field of depth. In order to evaluate the effect of optical distortion, the 3D projections were simulated but this time including the radial distortion model. Less accurate estimates were attained with SSE = 0.21 and $\mu'=3.3\times10^{-5}$ which is an order of magnitude smaller than the theoretical μ related to image resolution. Figure 6(a,b) shows the image-plane projection errors. Since no excessive error occurred by the end of step 1, no image points were excluded from further processing. One concludes that in this case the effect of distortion upon calibration was minor.

In actual applications the image resolution may seriously affect calibration accuracy. Table 4.1 summarizes the results for the virtual camera with spatial quantization effects included. The central region was again selected with size 1/4 of the acquired image side and ground-truth radial distortion coefficient k_1=0.05. Figure 6(c) shows that the detrimental effect of digitization is far more important than that arising from optical distortion. The high SSE value by the end of step 1 is in agreement with this assertion. Figure 6(d) depicts the estimation error after one iteration of step 2 with 1.14 pixel as the largest error. 3D measurements with corresponding projection errors above 0.7 pixel were then discarded, therefore removing fifteen 3D measurements and the corresponding projections from the measurement set. A second iteration of step 2 with this pruned measurement set significantly reduced the SSE value and the largest estimation error value, as shown in Table 4.1 (in the last page) and Figure 6(e), respectively.

In order to have the estimated image-plane projections close to the measurements, the identified principal point should be displaced to the left and up by means of the translation components t_x and t_y. Therefore, estimation errors in (r_0,c_0) were compensated by errors in the estimation of t_x and t_y. As the estimation of (r_0,c_0) improved with step 2, so did the estimation of t_y and t_y. Estimation of t_y was seen to be quite accurate. The above results obtained from synthetic data indicate that the method is adequate for calibration provided that the 3D measurements are sufficiently accurate.

4.2 Estimation of Confidence Intervals with Synthetic Data

The above results were produced with one realization of the 3D measurement set. Moreover, it occurs often that a useless biased estimate yields a small relative error because the estimation error is small compared to the magnitude of the estimate. By contrast, the magnitude of the ratio between the estimation error and the uncertainty range provides information on whether the estimate is biased to a point which makes it useless for a given purpose. It encouraged further statistical evaluation based on Montecarlo simulation. The calibration method has been utilized with 30 random realizations of the 3D

measurement set. The virtual camera parameters were selected equal to the ground-truth in Section 4.1 except for the translation $\mathbf{T}=[-11\ 8\ 116]^{\mathrm{T}}$ and the focal length. To the latter was assigned the minimum nominal value, 1.1 cm, in order to provide maximum depth of field when calibrating the actual vision system and therefore $(f_u,f_v)=(-1.1s_u,1.1s_v)$ pixels.

The simulated grid was positioned at four distinct stations and the coordinates of the cell vertices randomly generated as multiples of the actual grid cell dimensions. 3D occluded points were pruned from the measurement set. The number of measurements at each station was kept similar to that acquired by the real vision system, thus yielding a total of 160 synthetic 3D measurements. The interval estimator of the uncertainty range of the parameters is proposed as follows (Papoulis,1991). Let $\hat{\eta}$ denote the estimate of the constant parameter η from noisy measurements η_i, i.e., $\eta_i=\eta+\nu_i$, i=1,2,...,n; ν_i white noise with $N(0,\sigma^2)$. It is required to estimate the γ-confidence interval in which η lies with a given probability γ (confidence coefficient) or confidence level $\delta=1-\gamma$. In this case, the γ-confidence interval is $\hat{\eta}\pm\hat{\sigma}t_u(30)/\sqrt{n}$, $u=1-\delta/2$, where $t_u(30)$ is the standard Student-t density percentile for n=30 samples, $\hat{\eta}$ is the mean $(1/n)\sum_{i=1}^{n}\eta_i$ and $\hat{\sigma}^2=(1/(n-1))\sum_{i=1}^{n}(\eta_i-\hat{\eta})^2$ is the unbiased estimate of the noise variance. The following mean estimates and confidence intervals for γ=0.95 ($t_u(30)$=2.05) were obtained:

$$(\hat{r}_0\ \ \hat{c}_0)=(131\pm 2.24\ \ 161\pm 2.63) \tag{4.1a}$$

$$(\hat{f}_u\ \ \hat{f}_v)=(1.09s_u\pm 1.33\ \ 1.09s_v\pm 1.48) \tag{4.1b}$$

$$\hat{\mathbf{R}}=\begin{bmatrix} 1\pm 3\times 10^{-4} & -5\times 10^{-4}\pm 5\times 10^{-4} & 3.07\times 10^{-2}\pm 5.7\times 10^{-3} \\ 2\times 10^{-4}\pm 5\times 10^{-4} & 1\pm 3\times 10^{-4} & -6.5\times 10^{-3}\pm 6.7\times 10^{-3} \\ -1.76\times 10^{-2}\pm 5.5\times 10^{-3} & -2\times 10^{-4}\pm 6.4\times 10^{-3} & 1\pm 1.3\times 10^{-3} \end{bmatrix} \tag{4.1c}$$

$$\hat{\mathbf{T}}=\left[-7.53\pm 0.652\ \ 8\pm 0.767\ \ 116\pm 0.485\right]^{\mathrm{T}} \tag{4.1d}$$

With regard to the size of the central region, the behavior of the NCE index was observed as the side of the central region h_k=image_size/k, k=1;1.5;2;2.5;...;9.5;10; varied. For each central region size, 30 realizations of the 3D measurement set were employed and the average NCE value and its corresponding standard deviation computed. The results remained essentially invariant with an average NCE just under unity and standard deviation 0.034 for k within the interval [1;5.5]. This indicates that the simulated optical distortion had a minor effect. However, small central regions (k>5.5) resulted in too few points remaining for an adequate initial estimation thus yielding a massive increase in the NCE statistics and hence a high sensitivity to variations in the measurement set.

4.3 Calibration of the Actual Vision System

For the calibration of the monocular tracking vision system the servo-actuated lens was set to its minimum nominal focal length of 1.1 cm in order to keep the whole scenario in focus. This is necessary due to the assumption of pin-hole camera optics used by the algorithm, with which the image is focussed regardless of the depth of an object within the field of view. The size of the central region was set to 1/4 of the image size yielding 116 central points. The correspondence between the image-plane projections and 3D measurements was carried out by the operator. The results are shown in Table 4.2. The

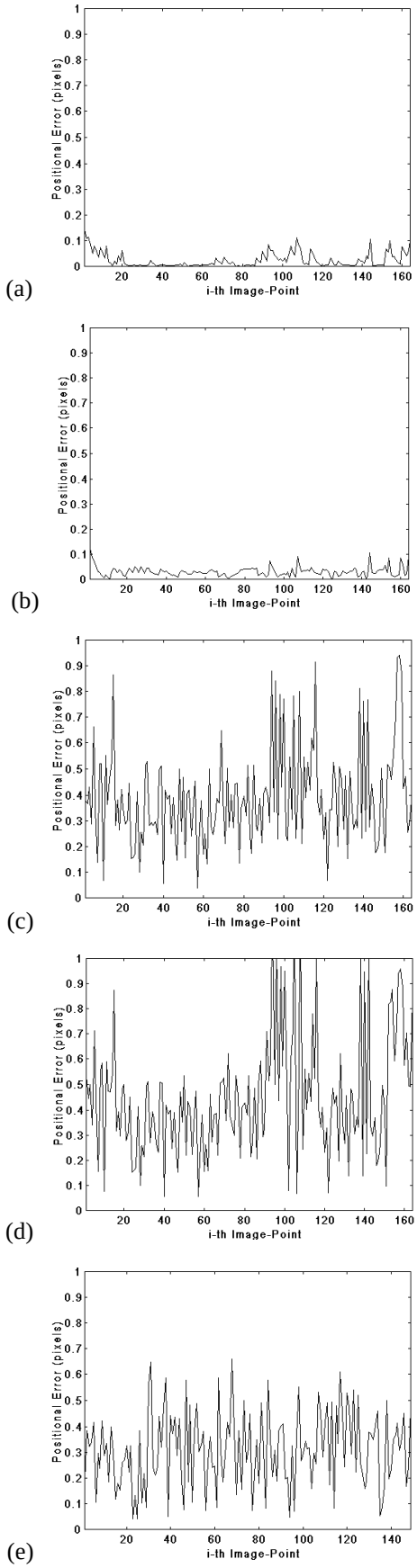


Figure 6: (a) Image-plane projection errors, radial distortion only, step 1 concluded. (b) Idem, step 2 concluded. (c) Image-plane projection errors, radial distortion and digitization, step 1 concluded. (d) Idem, step 2 concluded. (e) Idem with pruned measurement set, step 2 concluded.

resulting image-plane projection error after the initial identification of k_1 (see Section 3.1.2.1) and the twelve discarded measurements due to estimation error above 0.7 pixel are shown in Figure 7(a,b). Such pruning of inconsistent data in the measurement set significantly reduced the SSE, NCE and μ' values as observed in Table 4.2. The resulting SSE, NCE and μ' indicate that adequate calibration was achieved. The estimated focal length was employed in the background motion compensation module of a tracking algorithm with good results (Waldmann and Francisco Jr, 1997).

The statistical evaluation of the confidence intervals from the actual data requires a number of realizations of the measurement set which is not feasible. Thus, one resorts to the confidence intervals obtained for the virtual camera in equation (4.1) and the estimates in Table 4.2. Regarding the principal point, the confidence interval is much smaller than the distance from the estimate to the nominal image center, an occurence which indicates a possible misalignment between the camera and the lens assembly or a significant bias in this estimate. Likewise, the error between the the nominal focal length and its estimate, when compared to the corresponding confidence interval, again indicates the occurrence of either a biased estimate or considerable incorrectness in the datasheet provided by the lens manufacturer. However, the resulting parameter estimates represent quite well the imaging process from the 3D measurement set to the image-plane projections. This claim is confirmed by NCE<1 in Table 4.2 and the results of Figure 7(c,d).

5 CONCLUSIONS

A two-step calibration method has been evaluated in an actual vision system. The first step yielded an initial solution under the assumption of ideal optics whereas the second step refined it by taking the radial distortion into account. The estimated focal length was considered satisfactory for the intended purpose of visual tracking. The estimated focal length was successfully used to carry out the background motion compensation due to camera motion.

The calibration results strongly depended on the quality of the 3D measurements. The pruning of inconsistent 3D measurements was seen to significantly improve the results. Calibration accuracy was seen to be sensitive to a sharp decrease in central region size causing too few points to be used in the estimation of an initial solution. Furthermore, a decrease in the field of view due to zooming had a detrimental effect on calibration accuracy as 3D measurement errors became more significant. Confidence intervals were produced via Montecarlo simulation of a virtual camera and 3D measurements. The resulting confidence intervals together with the NCE value showed that in spite of the estimates for the actual vision system being biased, the calibration algorithm succeeded in the extraction of useful information about the imaging process from the available measurement set.

The calibration method here evaluated will be useful to characterize the relationship between the actual focal length and the distance to a focussed object and their respective commanded values as computed by the lens control algorithm. The characterization of errors in the closed-loop optics control will enable the evaluation of exploration and tracking modes as they compete and cooperate within the framework of an attentive strategy.

REFERENCES

Abidi, M.A. and T.Chandra (1995). A New Efficient and Direct Solution for Pose Estimation Using Quadrangular Targets: Algoritm and Evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.17, nº 05, pp.534-538.

Batista, J.; P.Peixoto, and H.Araújo (1997). Visual Behaviors for Real-Time Control of a Binocular Active Vision System. IFAC AARTC'97 – Algorithms and Architectures for Real-Time Control. Vilamoura, Portugal.

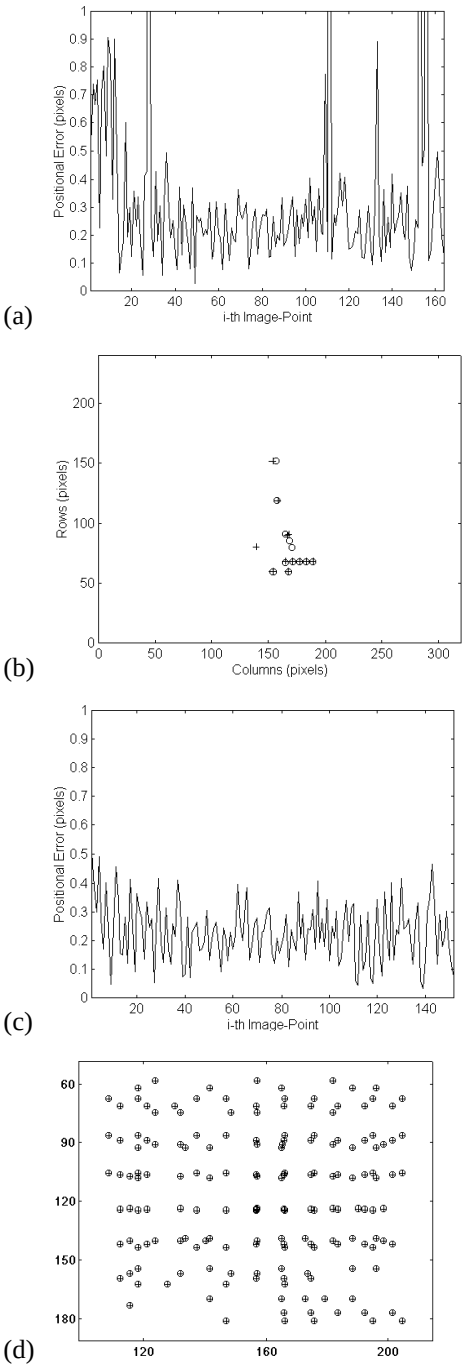


Figure 7: (a) Image-plane projection error after initial identification of k_1. (b) Discarded measurements. "+" indicate actual measurements, "o" indicate the estimated projections. c) Final image-plane projection error of remaining measurements. d) Image-plane measurements and estimated projections for the calibrated vision system

Chaterjee, C. and V.Roychowdhury (1993). Efficient and Robust Methods of Accurate Camera Calibration. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.664-665.

Jackowski, M. and A.Goshtasby (1997). Correcting the Geometry and Color of Digital Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.19, nº 10, pp.1152-1158.

Li, M. and J.-M. Lavest (1996). Some Aspects of Zoom Lens Camera Calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.18, nº 11, pp.1105-1110.

Murray,D. and A.Basu (1994). Motion Tracking with an Active Camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.16, nº 05, pp.449-459.

Papoulis, A. (1991). *Probability, Random Variables and Stochastic Processes*. McGraw-Hill..

Penna, M.A. (1991). Camera Calibration: A Quick and Easy Way to determine the Scale Factor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.13, nº 12, pp.1240-1245.

Tsai, R.Y. (1986). An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.14, nº 10, pp.364-374.

Tsai, R.Y. (1987). A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf Cameras and Lenses. *IEEE Journal of Robotics and Automation*, vol.RA-3, nº 04, pp.323-344.

Tsai, R.Y. (1989). Synopsis of Recent Progress on Camera Calibration for 3D Machine Vision. In O. Khatib, J. J. Craig and T. Lozano-Pérez (Eds). *The Robotics Review*, pp.147-159, The MIT Press.

Tu, X.-W. and B.Dubuisson (1992). CCD Camera Model and its Physical Characteristics Consideration in Calibration Tasks for Robotics. *IEEE International Workshop on Robot and Human Communication*, pp.75-77.

Waldmann, J. and Francisco Jr., A. (1997). Fuzzy Control of Monocular Fixation in a Robotic Vision Head. *Proceedings of the Workshop on Intelligent Robotics, XVII Brazilian National Congress in Computer Science*, pp.56-67.

Wang, L.-L. and W.H-.Tsai (1991). Camera Calibration by Vanhishing Lines for 3D Computer Vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.13, nº 04, pp.370-376.

Weng, J.; P.Cohen and M.Herniou (1992). Camera Calibration with Distortion Models and Accuracy Evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.14, nº 10, pp.965-980.

Table 4.1: **Calibration results for virtual camera with radial distortion and digitalization error.**

Camera parameters	Solution by the end of step 1 (central region only)	Idem, by the end of step 2 (whole image)
Principal point (r_0, c_0) [pixels]	(133,169)	(119,166)
Focal length (f_u, f_v) [pixels]	$(-1.96s_u, 1.98s_v)$	$(-1.98s_u, 1.98s_v)$
Translation **T** [cm]	$[-8.68\ \ 6.36\ \ 128]^T$	$[-11.3\ \ 6.91\ \ 128]^T$
Rotation **R**	$\begin{bmatrix} 1.0022 & -0.0001 & 0.0225 \\ -0.0002 & 1.0019 & -0.0128 \\ 0.0000 & 0.0065 & 0.9993 \end{bmatrix}$	$\begin{bmatrix} 1.0087 & 0.0002 & 0.0138 \\ -0.0006 & 1.0054 & -0.0128 \\ 0.0068 & -0.0053 & 1.0092 \end{bmatrix}$
Radial distortion coefficient k_1	– – –	0.3752
SSE [pixels^2]	30.643	17.482
μ'	7.69×10^{-4}	3.30×10^{-4}

Table 4.2: Calibrated parameters of the actual vision system.

Camera parameters	Solution by the end of step 1 (central region only)	Idem, by the end of step 2 (whole image)
Principal point (r_0, c_0) [pixels]	(100,171)	(100,171)
Focal length (f_u, f_v) [pixels]	$(-1.21s_u, 1.23s_v)$	$(-1.21s_u, 1.23s_v)$
Translation **T** [cm]	$[-21.2\ \ 5.96\ \ 116]^T$	$[-21.2\ \ 5.89\ \ 116]^T$
Rotation **R**	$\begin{bmatrix} 1.0021 & -0.0031 & 0.0403 \\ -0.0036 & 1.0014 & -0.0772 \\ -0.0194 & 0.0673 & 1.0093 \end{bmatrix}$	$\begin{bmatrix} 1.0006 & -0.0014 & 0.0410 \\ 0.0000 & 1.0014 & -0.0767 \\ -0.0534 & 0.0437 & 1.0005 \end{bmatrix}$
Radial distortion coefficient k_1	– – –	6.51×10^{-3}
SSE [pixels^2]	1013	9.282
NCE	6.0899	0.93879
μ'	7.81×10^{-4}	7.07×10^{-4}