Proceedings of the IMProVe 2011
International conference on Innovative Methods in Product Design
June 15th – 17th, 2011, Venice, Italy

On the performance of several stereo calibration methods and models

R. Furferi[(a)], L. Governi[(a)], M. Nunziati [(a)]

[(a)]Dipartimento di Meccanica e Tecnologie Industriali, Università degli Studi di Firenze

Article Information

Keywords:
Camera model,
Computer vision,
Stereo calibration.

Corresponding author:
MatteoNunziati
Tel.:(+39) 0554796394
Fax.:(+39) 0554796400
e-mail:matteo.nunziati@unifi.it
Address: via di Santa Marta, 3, 50139 Firenze

Abstract

Stereo vision is a well-known technique which relies on a pair of cameras in order to reconstruct the shape and position of a generic object, without any additional geometrical and/or parametric information. The central issue in the set-up of a stereo vision system is two-fold: as first, removing the geometrical distortion caused by camera lens from images, then, make cameras aware of their own relative position in space. This paper is aimed to test the influence of different components of a calibration routine. In particular the goal is to compare the performance of several optimization algorithms and a number of alternative implementations of the pin-hole model. A main difference between this work and other tests present in literature is that the calibration performance is evaluated with respect to the measurement accuracy of the system, rather than by means of estimated reconstruction errors. In such a way, we get rid of theoretical errors, which do not represent any real application case, and we evaluate the accuracy on-the-field, facing with real-world issues.
The obtained results show that, even though very complex equations can be used in order to represent cameras, usually, simpler pin-hole models remain competitive and robust, while refinements can be attained by using more powerful operational research algorithms.

1 Introduction

Stereo vision is a well-known technique which relies on a pair of cameras in order to reconstruct the shape and position of a generic object, without any additional geometrical and/or parametric information. Such a technique is used in several industrial applications involving, among others, reverse engineering.

The central issue in the set-up of a stereo vision system is two-fold: as first, removing the geometrical distortion caused by camera lens from images, then, make cameras' aware of their own relative position in space. In order to solve this set of problems, cameras are usually represented by means of so named pin-hole models.

Such models are usually tuned by means of two operational research algorithms: the first one employed to estimate models' optical parameters, the second used to estimate pin-holes (i.e. the cameras) relative position. The goal of the first algorithm is to obtain a parametric representation of lens distortion, which can be employed to undistort camera images. The second algorithm provides a rotation matrix and a translation vector, which act as mapping function between the cameras' reference systems.

This paper is aimed to test the influence of different components of a calibration routine. The goal is to compare the performance of several optimization algorithms and a number of alternative implementations of the pin-hole model. In order to attain such a goal, several pin-hole implementations are calibrated against the same dataset by means of different algorithms. Later, calibrated models are used for the reconstruction of real-life objects' shapesby means of a stereo-triangulation technique. With the aim of validating the accuracy of the calibration models, a comparison between the obtained 3D models and the ones obtained by using 3D commercial acquisition devices is performed.Tested routines have been ranked by means of synthetic descriptors derived from descriptive statistics.

A main difference between this work and other tests present in literature is that the calibration performance is evaluated with respect to the measurement accuracy of the system, rather than by means of estimated reconstruction errors[1]. In such a way, we get rid of theoretical errors, which do not represent any real application case, and we evaluate the accuracy on-the-field, facing with real-world issues.

The study has been limited to state-of-the-art mid-cost equipment and the stereo vision system has been assembled minding about possible constraints derived by industrial needs (as instance the dimension of the vision system itself), with the aim to provide useful hints for other machine vision application in industry.

The remainder of this paper is organized as follow. Section 2 recalls the main calibration equations for both the mono and stereo problems and describes the general optimization procedure required to estimate stereo-vision parameters.

Section 3 describes the different distortion models and optimization algorithms analysed in this paper.

Section 4 concerns with experimental datasets and results.

Finally, section 5 proposes some considerations and conclusions as well as further work.

[1] *It is quite common to calibrate a system with a chessboard image set and validate it by counter distorting another set of chessboard images. This procedure, even though correct, do not produce any direct measurement of on-the-field accuracy. In our experience it is also an optimistic estimation of real life errors induced by different objects and shapes.*

2 The stereo calibration problem

Camera calibration problems can be roughly divided in to two cases: single camera calibration (mono-vision) and multiple camera calibration (multi-vision). In the first case only one camera has to be calibrated, that is, only lens distortions have to be compensated. No 3D reconstruction is possible in mono-vision[2]. In the second case many cameras are employed in order to obtain *N* simultaneous points of views from which derive geometrical information. By using those points of view a 3D reconstruction isconceivable. Stereo-vision is a special case of multi-vision, where *N*=2.

From an algorithmic stand point, stereo-vision calibration (shortly stereo-calibration) is a two-step process: first, cameras are calibrated independently[3] in order to solve the mono-vision problem associated with each of them; later cameras' relative position in space has to be estimated.

Several calibration algorithms have been developed in the past in order to cope with the mono-vision problem. Two main families can be defined: self-calibration methods and photogrammetric methods [1]. The former group tries to estimate camera parameters referring to some unknown items which can be observed from different points of view in space. No metric information about such objects is required. The latter group estimates camera parameters by using a geometrically defined object, commonly referred as pattern. Patterns can be 1D, 2D or 3D objects, whose geometry is fixed. This paper focuses on 2D photogrammetric calibration, providingliterature the evidence that this method is more accurate and dependablewhen high precision measurements are required (as in reverse engineering).

2.1 Pin-hole model with distortions

In order to calibrate cameras a mathematical representation of the hardware is required. Commonly a so called pin-hole model is employed in order to represent cameras, while lens distortions are modelled by means of an auxiliary equation (fig.1).

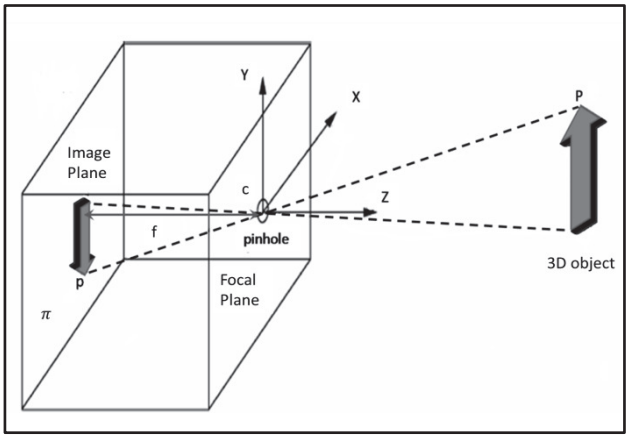

Fig. 1 The pin-hole model.

Therefore solving the mono-vision problem implies the resolution of two simultaneous equations (the pin-hole and the distortion). No effective closed form solution is available for the calibration problem, as a consequence, numerical methods are employed in order to iteratively solve it. In [1], wide room is dedicated to the description and interpretation of the equation set, therefore the authors invite the novel reader to refer to [1] for any further detail.

We shortly recall here the concept of intrinsic and extrinsic parameters and their relation to mono-vision calibration.

Given a 2D object of known geometry, defined by means of a set *P* of points in space (i.e., described by a set of 3D vectors) and a fixed reference system, an homographic transformation $H:\mathbb{R}^3 \rightarrow \mathbb{R}^2$is an invertible function which projects the given pattern into a 2D domain, namely, a plane parallel to an image plane π [1]. *H* comprises a rotation matrix and a translation vector commonly referred as *extrinsic* parameters of a camera.

A pin-hole model $C(f,c):\mathbb{R}^2 \rightarrow \mathbb{R}^2$ is described by the so-named intrinsic parameters: the lens focal length vector[4]*f* and the principal point *c*, that is, the projection of the CCD centre on the image plane π. Both *f* and *c* are 2D vectors. Once an homographic transform has been applied to a pattern*P*, The role of the pin-hole model is to predict the projection *P'* of *P* on π.

The symmetric part of *D(r,t)*evaluates the distortion caused by the applied lens, while the non-symmetric one accounts for tangential distortions induced by possible misalignmentsin the camera-lens coupling. Given *P'*, the role of the distortion equation is to apply a nonlinear mapping in order to predict the exact position *P''* of *P* on the final image.

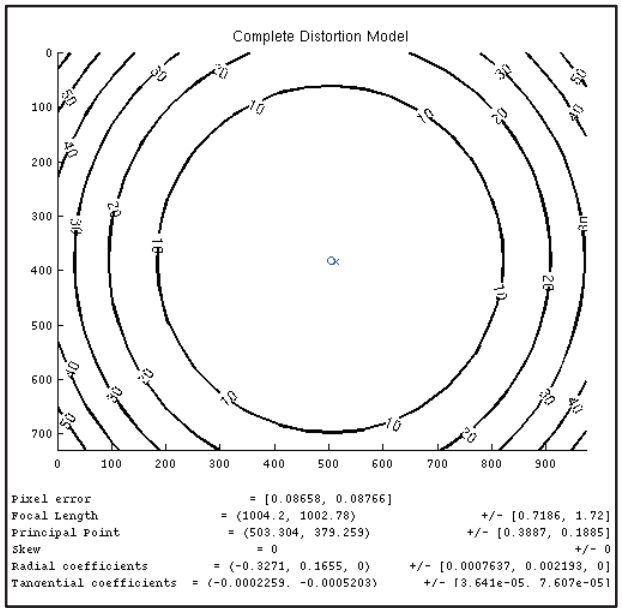

Fig. 2Iso-distortion curves of D(r, t).

The camera is thus representable by a model $M(C(f,c),D(r,t))=M(f,c,r,t):\mathbb{R}^2 \rightarrow \mathbb{R}^2$. Given an estimation of $\widehat{M}$ of *M* and $\widehat{H}$ of*H*, it is therefore possible to approximatethe position*P''*of *P*in a picture, and, theoretically[5], reconstruct the original set *P*given *P''*as

[2] *Actually, 3D reconstruction is possible for an object of known geometry such as a calibration pattern (more in next section).*
[3] *Actually, the same set of simultaneously acquired pictures is commonly employed in practice.*
[4] *Even though lens have a uniquefocal, the pin-hole model represents it as a 2x1 vector, whose values are the estimated focal length expressed in terms of horizontal and vertical CCD cells sizes – which can differ.*
[5] *Actually M is non-invertible, therefore numerical approximations are required.*

$\hat{P} = \hat{H}^{-1}\hat{M}^{-1}(P'')$. Where the *-1* superscript stays for the inverse function.

2.2 Stereo-vision and epipolar geometry

In order to understand stereo-vision it is important to recall that the re-projection in space of a single point *v'* $\epsilon\, \pi$ generates a line connecting the principal point *c* and *v'*. Such a line can intercept more than one vertex in space. In other terms, a set $e_\perp$ of vertexes adequately oriented in space can collapse into a single projected point. This implies that the projection point itself is not enough to reconstruct the actual position of a vertex $v \epsilon e_\perp$. Therefore, a single mapping function $\widehat{M}$ can reconstruct a point in space only if the cardinality of $e_\perp$ is 1 (and if an associated homography is present).

In case an unknown object *O* is projected on an image plane, it is not possible to predict if each projected point Pis associated to a set with cardinality 1 and, also, there is no homography to map the point back in space (fig.3).

As a consequence, as widely known, two or more points of view are required.

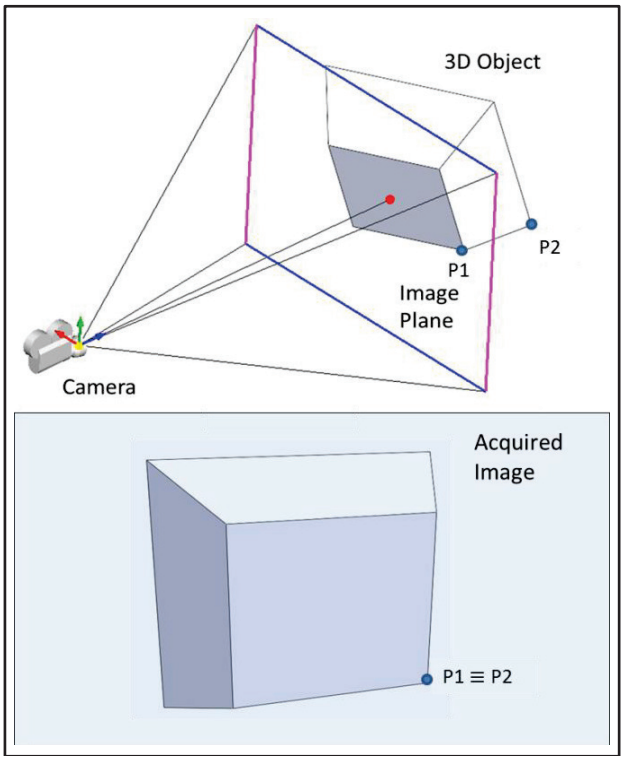

Fig. 3 projection of aligned orthogonal edges.

Given an estimation of camera models $\widehat{M_i}$,with *i=1,2* in the stereo-vision case, and the relative position of $\widehat{M_2}$ w.r.t. $\widehat{M_1}$, it is possible to reconstruct a vertex$v \epsilon e_\perp$, from itsdistorted projections v_i''on images planes π_i.The relative position of a camera w.r.t. the other is described by the augmented matrix $R^a=[R|T]$, where *R* is a 3x3 rotation matrix and *T* a 3D translation vector. R^a can be computed as the composed homography$R^a=H_1H_2^{-1}$which moves from one camera plane to the pattern plane (inverse homography H_2^{-1}), and, hence, projects back to the second camera plane (homography H_1).

An estimation$\widehat{R^a}$ of R^a is required as camerasare not aware of their own position in an external inertial reference system, as a consequence, object projections cannot be referenced one to each other.Let $S_U$be the orthonormal matrix describing the inertial system with coordinates *(x,y,z)*, the cameras lying in the space mapped by S_U. If the cameras position is unknown, that is, if the equations of $\pi_1$and $\pi_2$areunknown in S_U, it is not possible to reference the projection planes and, therefore, to understand on which planes *O* has been projected.

A straightforward definition of $S_U$can beobtained by fixingthe origin of the system on a camera principal point[6], settingz coincident to the optical axis and *(x,y)*parallel to the image plane and oriented according to camera CCD. In other terms, fixing $\widehat{M_1}$ as the reference camera for the sake of simplicity, c_1=(0,0,0) in S_U and $\pi_1 = ax + by + cz$.In such a way it is possible to switch from a camera point of view to the other by applying$\widehat{R^a}$.

Indeed,being the pattern known, it is possible to: estimate the homographic transforms of both cameras w.r.t. the pattern, move from one camera plane to the pattern plane, and, hence, re-project to the second camera.

As instance, it is possible to map O_2'' in $\widehat{M_1}$ by applying the following chain of operators:

$$\hat{O}_{2,1}'' = \widehat{M_1}\left(\widehat{R^a}\,\widehat{M_2}^{-1}(O_2'')\right) \qquad (1)$$

where $\hat{O}_{2,1}''$ is the estimated projection of *O* on image plane $\pi_1$obtained by: re-projecting O_2'' in space; roto-traslating it in S_U and projecting itagain onπ_1.$\widehat{R^a}$is a visual shorthand for the roto-translation operator.

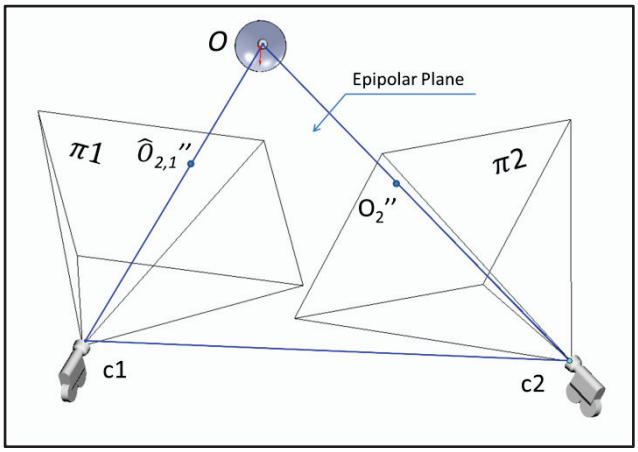

Fig. 4Epipolar geometry.

Let v_1'' and v_2'' be respectively the distorted projections of a vertex $v \epsilon e_\perp$on π_1 and π_2, the following steps have to be followed in order to reconstruct the position of *v* in space. As first, it is necessary to apply the mapping $\widehat{M_i}^{-1}$ to v_i'': in such a way the distorted i-*th* projection is undistorted and re-projected in space. Each re-projection describes a line r_i. The mapping $\widehat{R^a}$is applied to r_2 leading to $\hat{r}_{2,Su}=\widehat{R^a}r_2$. According to the common geometry of a stereo vision system, $\hat{r}_{2,Su}$ and r_1 intersect in space in a unique point: such point is and estimation $\hat{v}$ of *v*.

In real applications, more than a projected point is present on image planes. Therefore, in order to reconstruct a shape,it is important to detect which projected point $v_{k,1}''$ is associatedto which point $v_{k,2}''$. In order to cope to this problem the epipolar line [2] can be computed.

[6] *We consider only the case of non-moving vision systems.*

2.3 The estimation process

The goal of a stereo calibration routine is to iteratively solve both the mono- and stereo- problems by means of optimization algorithms. As stated above, a first iterative round is performed in order to solve the mono-vision problem, that is, in order to obtain an optimal estimation of the parameters vector $m = [f, c, r, t]$, which minimizes, in the least square sense, the root square error:

$$RSE_{mono} = \sqrt{\sum_{j=1}^{n}(v_j'' - \hat{v}_j''(k))^2} \quad (2)$$

Where: *j*is the number of vertexes present on a 2D calibration pattern;v_j'' is the projection of the *j*-th vertex on the distorted picture and $\hat{v}_j''(k)$ is an estimation ofv_j'', obtained during the *k*-th optimization step, by means of a tentative vector (and associated tentative homography).

Once an optimal estimation $\widehat{m_i}^{opt}$ has been obtained for both camera 1 and 2, the second iterative round can be performed in order to obtain and optimal estimation$\widehat{R^a}^{opt}$ of R^a. Again, such an estimation is optimal in the least square sense and is obtained by minimizing aroot square error,a widely[7] used formulationfor itbeing:

$$RSE_{stereo} = \sqrt{\sum_{i=1}^{2}\sum_{j=1}^{n}(v_{i,j}'' - \hat{v}_{i,1,j}''(k))^2} \quad (3)$$

Where: *i* is the camera number and all other items preserve the meaning of eq. (2). It is worth the trouble to point out that, accordingly to the previous notation, $\hat{v}_{i,1,j}''(k)$ represents the distorted projection of the *j*-th vertex on π_1.In case *i=1*, it simply indicates the projected point of the *j*-th vertex on the camera plane, otherwise is it the estimated position obtainedby applying eq. (1)during the *k*-th step, by using a tentative augmented matrix. An alternative version of eq.(3) can be written by omitting camera 1, that is, accounting only for camera 2 projected points.

3 Investigated models and methods

According to the underlying theory recalled in section 2, the following items have been investigated in the present work:

1. a number of optimization routines for the mono and stereo problems;
2. an alternative formulation of the error functions;
3. camera vectors *m* with simplified representations coupled with a number of formulations of *D(r,t)*.

The following paragraphs dedicate room to the description of the selected choices.

3.1 Optimization routines

The iterative procedures commonly employed for the resolution of eq. (2) and (3) belongs to the family of so-called non-linear least square solvers [3]. Among the many algorithms available for this kind of problems, we have limited our investigations to those implemented in the Matlab® programming environment, being it a dependable and commontool in research and providing it a wide set of state of the art solvers. Namely three methods have been analysed here: the gradient descent (as baseline), the Levenberg-Marquardt algorithm [5][6] and the Trust Region method [7].

The gradient descent (GD) is one of the most known methods for function minimization. It dates back to Gauss and searches for local minima[8] by updating an initial vector at each step *k*, according to the formula:

$$x_{k+1} = x_k + \delta_k \; with \; \delta_k = -\alpha_k J(x_k) \quad (4)$$

Where *x* is the vector of free variables, $J(x_k)$ is the gradient of the error function (as instance of eq. (2) or (3)) and $\alpha \epsilon (0,1)$ is a scaling factor. Usually, the scaling factor is updated at each step *k* in order to face with the nature of the gradient function. At the beginning a small value of *α* is used in order to cope with possible irregularities along the gradient direction, while increasing values are used when the research goes closer to the local point. This assures for a faster convergence when the search method impacts on a local minimum, which, commonly, is located in a relatively flat region.

The Levenberg-Marquardt algorithm (LMA) is one of the mostly employed methods for minimization. It is preferred over (4) due to its robustness against local minima, granting an higher probability of convergence in a global point. According to LMA, the free variables are update by an increment δ_k obtained by solving the following equation:

$$[J(x_k)^T J(x_k) + \alpha_k diag(J(x_k)^T J(x_k))]\delta_k = \\ = -J(x_k)^T e(x_k)$$

Where $e(x_k)$ is the error function and all other items preserve the meaning of eq. (4).

The Trusted Region method (TRM) involves a two-step procedure at each iteration *k*: as first the objective function *e(x)* is approximated by a simpler one (usually by an x_k-centred 2nd order Taylor series of the function), then such a function is minimized into a neighbourhood *D* of x_k. Let be *se(x)* the simplified objective function, that is:

$$se(x) = x^T S(x) x + J(x)^T x$$

Where *S(x)* is the hessian matrix of the error function. The goal is to find a point $x_k+\delta_k$in *D*such that: $se(x_k+\delta_k)$ is minimized and $e(x_k+\delta_k)<e(x_k)$. The most challenging part of the problem is the definition of the size of the neighbourhood as well as the find out of a fast way for solving the trust region minimization. For a primer on the topic we remind to [7].

3.2 Error functions

Among the different error formulations, an empirical alternative has been investigated with the aim to increase the impact of errors on the optimization routine and attain, if possible, a globally lowered error. In order to increase the impact of each error in eq. (2) and (3), the following modification has been investigated:

$$LE_{mono} = \sum_{j=1}^{n} -ln\left(1 - \left|\frac{v_j'' - \hat{v}_j''(k)}{v_j'' - \hat{v}_j''(0)}\right|\right) \quad (5)$$

[7] *This is basically the formulation implemented in the well-known Camera Calibration Toolbox for Matlab by Jean-Yves Bouguet.*

[8] *None of the proposed methods can assure for global convergence.*

And the associated stereo version:

$$LE_{stereo} = \sum_{i=1}^{2}\sum_{j=1}^{n} -ln\left(1 - \left|\frac{v''_{i,j} - \hat{v}''_{i,1,j}(k)}{v''_{i,j} - \hat{v}''_{i,1,j}(0)}\right|\right) \quad (6)$$

Where the root square error has been replaced by a logarithmic cost. The adoption of such a cost is motivated by the fact that a *log-cost* of this kind presents a steepest curve if compared to the classical squared error. Due to this, local errors should increase their relative weight in the cost function. In such a way, even a single error on one patter point, should acquire more relevance in the overall optimization routine, leading to more strict results.

3.3 Camera vectors

Among the different camera vectors *m*, two simplified alternatives have been investigated. The first one, named $m_{f=1}$, is motivated by the fact that almost all CCD have squared cell sizes, this means that the focal vector *f* could be forced to collapse into one simple scalar.

The second one, named $m_{c=fix}$, is related to the fact that mid-quality industrial camera assemblies (camera body + lens) usually show a quite good alignment between the optical axis of the lens and the versor of the CCD plane[9]. In other terms the principal point is expected to rely on a very small neighbourhood of the image centre. This leads to a simplified camera representation which do not include the principal point as a free variable, being it fixed to *c=(w/2,h/2)*, where *w* and *h* define the camera resolution in pixels.

For what concerns the distortion model, a number of different polynomial functions have been investigated.Being radial distortion usually the most relevant aberration component, we have focused our investigation on it. One of the most common models for lens distortion is the fourth order radial distortion model with even powers, defined as:

$$D_{4,even}(\rho) = r_1\rho^4 + r_2\rho^2 + 1$$

Where ρ is the radial distance of a projected point, computed w.r.t to *c*. In order to investigate the fitting capability of different polynomial formulas, we have varied the order of polynomials introducing both even/odd and full forms. For the sake of briefness, we identify such formulas by the term $D_{i,type}$, where *i* is the order of the polynomial and *type* can be one among *even, odd or full*depending on the employed powers.

4 Experimental datasets and results

The experimental data set has been retrieved by using a prototype developed for the measurement of car wheel alignment. Such a prototype concerns the development of an innovative artificial vision system capable to perform the measurement of camber and toe angles of vehicle wheels. The vision system acquires the 4 wheels of the vehicle in order to carry out a three-dimensional reconstruction of the scene, thus allowing a non-intrusive and real-time measurement of the toe and the camber angles [9]. The output of a measurement is, therefore, a pair of values representing toe and camber in centesimal degrees.

4.1 Equipment and data

The acquisition system is composed by 2 grey scale industrial cameras (c-mount) with a resolution of 1024x768. 3.5' lens have been mounted on both cameras. The angle between cameras is 20° and images have been acquired by using ultraviolet light sources and filtered by a band pass filter mounted in front of the camera lens. Lens type, system angle and illuminators have been constrained by the application design, thus, they represent possible limitations induced by real cases.

Calibration images have been kept by using a chessboard as 2D calibration pattern. The squares of the chessboard measure a size of 3 cm and the chessboard is composed by 29 rows of 19 squares. 16 images have been simultaneously acquired with the two cameras and processed for both mono and stereo calibration.

50 validation images have been acquired by using the equipment in a real test, thus, neither known geometry is available nor constrains about shape and position of objects can be defined. During the test, the car wheels have been oriented in different positions, causing the system to face with different configurations of the observed objects. The specificity of the validation set is that the observed objects cover the 90% of the image plane, thus they also rely on peripheral areas of the images, being subject to really high distortions.

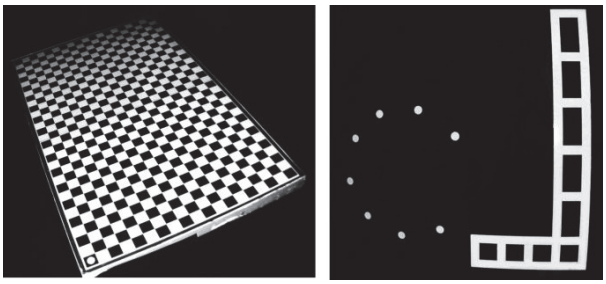

Fig. 5examples of calibration and validation images.

The calibration software has been obtained by hacking the well-known Camera Calibration Toolbox for Matlab by Jean-Yves Bouguet [10]. Therefore some of the algorithms come from this well-tested application, while specific modifications have been plugged in on demand for this paper.

4.2 Methods

As stated above, theforemost difference between the present work and other testsdescribed in literature is that the calibration performance is evaluated with respect to the measurement accuracy of the system, rather than by means of estimated reconstruction errors. As a consequencewe get rid of theoretical errors, which do not represent any real application case, and we evaluate the accuracy on-the-field.

In order to evaluate the accuracy of obtained measures, the output of the system is compared with a state of the art measurement system in use in the automotive industry. Such system being certified for accuracies of ±0.03°.

In order to rank each calibration, some statistics have been computed on validation images. Being the real value

[9] *This is confirmed, also, by the commonly small values assumed by the tangential distortion.*

of both camber and toe known, for each validation images an error has been computed as follows:

$$e = |x_{vision} - x_{ref}| \quad (7)$$

Where $x_{vision}$is the measure obtained by the vision system, while x_{ref} is the one obtained by the reference system.

In order to evaluate the accuracy of the system, the expected error along with its confidence interval has to be computed. The assumption is that the error distribution can be modeled by a normal distribution $N(\mu,\sigma)$. Being both μ and σ unknown, the expected value of the error and its related confidence interval can be computed by means of:

$$e_{99\%} = \hat{\mu} \pm t_{(M-1,0.99)} \frac{\hat{\sigma}}{\sqrt{M}} \quad (8)$$

Where $\hat{\mu}$ and $\hat{\sigma}$ are the unbiased estimators of the mean and standard deviation of error *e*, *M* is the number of error measurements, $t_{(M-1,0.99)}$ is the t-student distribution for the mean estimation at aconfidence level of 99%.

Being eq. (8) fixed for each calibration and varying only the unbiased estimators, a simplified scoring rule has been defined as:

$$s = \hat{\mu} + \hat{\sigma} \quad (9)$$

It can be easily shown that when eq. (9) increases, eq. (8) provides an higher uncertainty and/or an higher average error, thus, eq. (9) can be considered as a synthetic estimator of system bias.

Every measurement set provides two scores: one for toe estimation s_{toe}, and one for camber s_{camber}. In order to get a single index of calibrationcorrectness, a total score s_{tot} has also been computed as sum of toe and camber scores.

Each investigated system will be identified by a set of four items$(m_x;D_{i,type};\ \{RSE,LE\};\ \{GD,MLA,TRM\})$ which uniquely identifies the kind of camera vector and the related radial distortion model along with the objective function (eq. (1) and (2)*or*(5) and (6)) and the employed optimization method.

4.3 Results and discussion

As first, a baseline system $(m_{stock};D_{4,even};RSE;GD)$has been evaluated, where m_{stock} is the classical vector described in section 2. Later, the systems reported in table 1 have been analyzed. Table 1 does not report all the possible combinations of vectors, distortion equations and optimization objectives/methods. The set of reported systems has been obtained by selecting those results which appeared to provide any relevant evidence in support or in contrast to a certain hypothesis. In bold the structural differences w.r.t the baseline system.

Name	*m*	*D*	*obj*	*Method*	*Score*	*%w.r.t base*
Baseline	m_{stock}	$D_{4,even}$	RSE	GD	3.40	0.00
Sys_1	m_{stock}	$D_{4,even}$	RSE	**LMA**	3.24	-4.70
Sys_2	m_{stock}	$D_{4,even}$	RSE	**TRM**	3.10	-8.82
Sys_3	m_{stock}	$D_{4,even}$	**LE**	**TRM**	3.38	-0.58
Sys_4	$\mathbf{m_{c=fix}}$	$D_{4,even}$	RSE	**TRM**	5.15	51.47
Sys_5	$\mathbf{m_{f=1}}$	$D_{4,even}$	RSE	**TRM**	*2.92*	*-14.11*
Sys_6	$\mathbf{m_{f=1}}$	$\mathbf{D_{2,even}}$	RSE	**TRM**	10.69	214.41
Sys_7	$\mathbf{m_{f=1}}$	$\mathbf{D_{6,even}}$	RSE	**TRM**	5.92	74.11
Sys_8	$\mathbf{m_{f=1}}$	$\mathbf{D_{2,full}}$	RSE	**TRM**	3.64	7.05
Sys_9	$\mathbf{m_{f=1}}$	$\mathbf{D_{3,full}}$	RSE	**TRM**	4.26	25.29
Sys_{10}	$\mathbf{m_{f=1}}$	$\mathbf{D_{3,odd}}$	RSE	**TRM**	*2.94*	*-13.52*

Tab. 1analysed systems along with resulted scores.

The very first consideration is that both sys_5 and sys_{10} seems to be the winner in this comparison.More in depth, by comparing the first four systems it is pointed out that the selection of an adequate optimization routine can completely redefine the accuracy of a numerical model. Indeed, a reduction of the inaccuracy score of almost a 9% is obtained without modifying any internal component of the baseline system.

By looking at sys_4 and sys_5 it is pointed out that even if a really accurate coupling can be attained with mid-quality industrial equipment, the principal point estimation is mandatory, nonetheless, a simplified focal length representation can boost the final accuracy up to -14% (with an additional improvement w.r.t to sys_2 of 56%!).

Finally, among the different shapes of the radial distortion equation it seems that a relatively low order polynomial can attain a very good estimation of the actual aberration, confirming what obtained in [8]. What seems new and unpredicted is that even an *odd* formulation can adequately estimate the lens distortion, even though classical photogrammetric models historically relies only on *even*powers.

For what concerns the objective function, it seems that the maximum likelihood estimation obtained by means of the RSE still holds w.r.t to other alternatives, even though the LE alone still doesn't full fit the needs of an exhaustive analysis.

5 Conclusion

The obtained results show that, even though very complex equations can be used in order to represent cameras, usually, simpler pin-hole models remain competitive and robust, while refinements can be attained by using more powerful operational research algorithms.

Namely, the level of modern mid-quality industrial equipment seems to not require any specific computation of 2D focal vectors. At the same time, actual assemblies seems to be affected by misalignments which induces neglectable tangential distortions but still affect the centering of the image plan in a relevant manner.

For what concerns lens distortion, it is known that the number of experimental data grows with the complexity of models in a more that linear manner, nonetheless there is no evidence that the provided dataset was too limited in order to tune complex polynomials.

This leads to consider the fact that, actually, small polynomials have modeled relevant levels of distortion without any issue, overcoming more sophisticated solutions. Therefore, it can be derived that, even if applied lens show a relevant aberration, quite simple models seems good enough to cope with the problem in industrial applications.

Eventually, it is worth the trouble to mind about the numerical methods employed for the calibration, being algorithms responsible of relevant levels of refinement. As far as common methods are involved, the TRM is recommended here.

Wide room has to be dedicated to objective functions, instead. The bare minimum test conducted in this study do not support the idea that a different objective function can induce benefits to the calibration routines, nonetheless, other alternative cost functions should be elaborated in order to cope with the problem. An additional direction of study is the analysis of non-deterministic methods such as genetic algorithms or simulated annealing procedures. Eventually, further work should be dedicated to non-polynomial families of aberration functions, in order to assess the benefit induced industrial applications by these unconventional models.

References

[1] Zhang, Z., “A flexible new technique for camera calibration”, *technical report MRS-TR-98-71*, 1998

[2] Linda G. Shapiro and George C. Stockman (2001). *Computer Vision*. Prentice Hall. pp. 395–403. ISBN 0-13-030796-3.

[3] Dennis, J.E., jr., “Nonlinear Least-squares”, *State of the art in numerical analysis*, ed. Jacobs, Accademic Press, pp.269-312, 1977.

[4] Marquardt, D., "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," *SIAM J. Appl. Math*. Vol. 11, pp 431-441, 1963.

[5] Levenberg, K., "A Method for the Solution of Certain Problems in Least Squares," *Quart. Appl. Math*. Vol. 2, pp 164-168, 1944.

[6] Moré, J.J. and D.C. Sorensen, "Computing a Trust Region Step," *SIAM Journal on Scientific and Statistical Computing*, Vol. 3, pp 553-572, 1983.

[7] The Matlab help manual, distributed with the IDE

[8] Weng, J., Cohen, P., Herniou, M., “Camera Calibration with Distortion Models and Accuracy Evaluation”, *IEEE transactions on pattern analysis and machine intelligence*, Vol.14, pp.965-981, 1992.

[9] M.Carfagni, R.Furferi, L. Governi. Progetto e realizzazione prototipale di un nuovo sistema basato su visione artificiale per la misura tridimensionale dell’assetto ruote di autoveicoli, Congresso Internazionale Congiunto XVI ADM – XIX INGEGRAF, Perugia, 6 – 8 Giugno 2007

[10] Camera Calibration Toolbox website, http://www.vision.caltech.edu/bouguetj/calib_doc.