High-accuracy and robust localisation of large control markers for geometric camera calibration

Damien Douxchamps, *Member, IEEE,* and Kunihiro Chihara, *Senior Member, IEEE*

Abstract—Accurate measurement of the position of features in an image is subject to a fundamental compromise: the features must be both small, to limit the effect of non-linear distortions, and large, to limit the effect of noise and discretization. This constrains both the accuracy and the robustness of image measurements, which play an important role in geometric camera calibration as well as in all subsequent measurements based on that calibration. In this paper, we present a new geometric camera calibration technique that exploits the complete camera model during the localization of control markers, thereby abolishing the marker size compromise. Large markers allow a dense pattern to be used instead of a simple disc, resulting in a significant increase in accuracy and robustness. When highly planar markers are used, geometric camera calibration based on synthetic images leads to true errors of 0.002 pixels, even in the presence of artifacts such as noise, illumination gradients, compression, blurring and limited dynamic range. The camera parameters are also accurately recovered, even for complex camera models.

Index Terms—camera calibration, imaging geometry, image measurement, high resolution, noise, ray tracing, sub-pixel

1 Introduction

Machine vision is frequently used to provide quantitative measurements for a wide range of applications, from simple telecentric designs used in factory automation to more complex, multi-camera systems used for 3D shape reconstruction. High accuracy is commonly required, especially for methods based on triangulation such as stereo vision or laser ranging. Indeed, such methods have a limited baseline and must cope with a wide range of object distances, leading to resolution and accuracy requirements well below the pixel level. To provide useful 3D measures, it is thus necessary to determine the relationship between these image measures and usable metric measures with a comparable or higher accuracy. This relationship is formalized in the camera model, and the determination of its parameters, called geometric camera calibration, must therefore be performed with great care.

For these reasons, geometric camera calibration has been an active research field and numerous further developments have been proposed since the method of Tsai [1]. Most offline approaches now achieve less than 0.1 pixels residual using low-cost hardware [2], such as the ubiquitous chess-board pattern [3] [4]. While this calibration is sufficient in most cases, several machine vision applications such as those mentioned above would benefit from a higher accuracy. To this end, Heikkilä proposed a high-accuracy ellipse-fitting technique [5] that uses a first-order elliptic approximation for the circular shape of a calibration marker in the image plane, taking into account the effect of perspective. He later applied this marker localization technique to camera calibration [6] [7] and, using a dihedron as target, obtained high-quality results with a residual error of 0.02 pixels. Developing this technique, Redert et al. added second-order corrections by weighting the pixel intensities with their true area and taking into account image curvature [8]. Residual errors as low as 0.01 pixels were obtained. Although these results are impressive, they have not been tested on a wide range of image imperfections and they fail in a number of cases; for instance, Heikkilä's results were obtained with images containing little noise and Redert's technique fails to maintain its accuracy with noisy or blurred images.

The origin of both the limited accuracy and the noise sensitivity of these techniques is explained by a fundamental compromise that must be made during the localization of calibration markers: the markers must be large, to decrease noise sensitivity and increase localization accuracy, but they also must be small, to minimize the difference between the marker model (such as ellipses or lines) and the more complex real shape of the marker in the image, a shape which is affected by non-linear optical distortions.

The approach we present in this paper overcomes the compromise on the marker size by applying the full camera model to build synthetic images that are then compared with the real images during optimization, thereby avoiding unnecessary approximations such as a marker model in the image plane or an inverse calibration model [7]. This type of calibration was pioneered by Robert [9]. Using larger markers opens new perspectives, such as the use of markers made of a dense pattern rather than a single boundary. This novelty increases the effective interface length of the marker, and enhances both the robustness and the accuracy of the localization process.

Besides the localization of markers in the image plane, the accuracy of geometric camera calibration depends on the camera model, the geometric accuracy of the calibration target, the optimization algorithms used to converge on the camera parameters and the operational conditions of the experiments. These aspects are addressed respectively by using a complete camera model, by compensating for manufacturing errors of the calibration target and by using a robust matching criterion. Together with the use of large markers and high-resolution images, taking all these aspects into account allows us to reach a true error of under 0.002 pixels. This accuracy holds even in the simultaneous presence of several strong image artifacts. The markers, however, must remain very accurate and planar.

2 Camera model and imaging geometry

In this section, the classic geometry and camera model are briefly discussed. To this effect, let us review the projection of a point P (Fig. 1, top) in three-dimensional space onto its image p' on the image plane. This projection involves four transformations (Fig. 1, bottom).

The authors are with the Nara Institute of Science and Technology, Japan.
Manuscript received Month Day, Year; revised Month Day, Year.

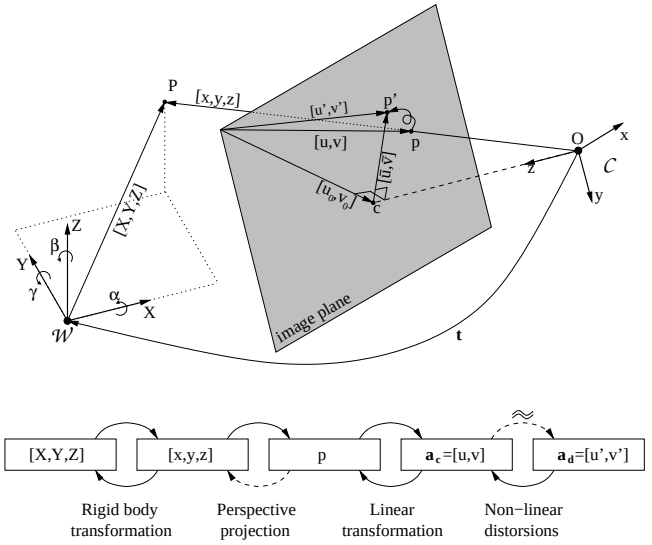

Fig. 1. Geometric transformations modeled by the camera calibration and the corresponding transformations. Inverting the perspective projection is not possible in itself but can be performed with additional information such as a second view (stereo-vision), a scene model or a laser plane. There is no analytic solution to the transformation from ideal image coordinates $\mathbf{a_c}$ to distorted image coordinates $\mathbf{a_d}$ [7]. The rigid body transformation is based on three rotation angles α, β and γ and the translation vector $\mathbf{t}$.

In the first transformation, the position $[X, Y, Z]^T$ of P in the world referential $\mathcal{W}$ is expressed in the camera referential $\mathcal{C}$. The center of projection O of the camera is located at the origin of $\mathcal{C}$. The camera referential has its z axis perpendicular to the image plane and its x and y axes parallel to the image lines and columns, respectively. This rigid body transformation can be written as

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \tag{1}$$

where $[x, y, z]^T$ are the coordinates of P in $\mathcal{C}$, $\mathbf{R}$ is a rotation matrix built on three Euler angles (α, β, γ) and $\mathbf{t}$ is the translation vector between the two referentials.

The second transformation is a perspective projection of P onto its image $[u, v]^T$ in the image plane using the pinhole camera model [10]. The third step is a linear transformation of the image that includes a change in origin and a rescaling. The second and third transformations can be written together as a single matrix operation [11] [12]:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \propto \begin{bmatrix} \lambda u \\ \lambda v \\ \lambda \end{bmatrix} = \begin{bmatrix} f_x & k & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \tag{2}$$

where f_x and f_y are the horizontal and vertical focal lengths expressed in pixel units, k is the skew of the image plane axes and $[u_0, v_0]^T$ are the coordinates of the principal point c where the z axis of $\mathcal{C}$ intersects with the image plane.

The linear camera model obtained by combining (1) and (2) is convenient because it is reversible and its parameters can be estimated directly [11]. However, non-linear optical distortions are not included, and it is therefore insufficient for accurate camera calibration. In reality, only the geometrically distorted image coordinates $\mathbf{a_d}$=$[u', v']^T$ can be measured in the image plane. These are linked to the ideal, or corrected, image coordinates $\mathbf{a_c}$=$[u, v]^T$ with the following fourth transformation [7]:

$$\mathbf{a_c} = \mathbf{a_d} - \mathcal{F}(\delta, \mathbf{a_d}), \tag{3}$$

where $\mathcal{F}$ is a non-linear distortion function of parameters δ. The classic model for the non-linear distortion function is the sum of radial distortions $\mathbf{D_r}$ and tangential distortions $\mathbf{D_t}$ [10], to which we also add the prism distortions $\mathbf{D_p}$ [13]. The center of the distortions $\mathbf{c}_d = (u_d, v_d)$ also does not necessarily coincide with the principal point $\mathbf{c} = (u_0, v_0)$ [10]. This results in the following model for the non-linear distortions:

$$\mathcal{F}(\delta, \mathbf{a_d}) = \mathbf{D_t} + \mathbf{D_r} + \mathbf{D_p}, \tag{4}$$

$$\delta = (a_0, a_1, a_2, a_3, a_4, p_0, p_1, p_2, p_3, s_0, s_1, s_2, s_3, \mathbf{c}_d). \tag{5}$$

Each of these three distortions is defined as a function of different distortion parameters:

$$\mathbf{D_r} = (a_0 r^2 + a_1 r^4 + a_2 r^6 + a_3 r^8 + a_4 r^{10}) \begin{bmatrix} \bar{u} \\ \bar{v} \end{bmatrix}, \tag{6a}$$

$$\mathbf{D_t} = \begin{bmatrix} (p_0 + r^2 p_2)(r^2 + 2\bar{u}^2) + 2(p_1 + r^2 p_3)\bar{u}\bar{v} \\ (p_1 + r^2 p_3)(r^2 + 2\bar{v}^2) + 2(p_0 + r^2 p_2)\bar{u}\bar{v} \end{bmatrix}, \tag{6b}$$

$$\mathbf{D_p} = \begin{bmatrix} s_0 \\ s_1 \end{bmatrix} r^2 + \begin{bmatrix} s_2 \\ s_3 \end{bmatrix} r^4, \tag{6c}$$

$$\bar{u} = u' - u_d, \quad \bar{v} = v' - v_d, \quad r^2 = \bar{u}^2 + \bar{v}^2. \tag{6d}$$

In this form, the parameters of the distortions can spread over more than 20 orders of magnitude, which can lead to instabilities in the optimization. This problem is solved by dividing each parameter a_i, p_i and s_i by f_x elevated to a power equal to the power of their multiplier r, e.g., $a_3 \rightarrow a_3/f_x^8$, $p_2 \rightarrow p_2/f_x^4$ [14].

Note that the inverse of (3) has no analytical solution [10], so that the following distortion model is only approximate [7] [15]:

$$\mathbf{a_d} \approx \mathbf{a_c} + \mathcal{F}(\delta, \mathbf{a_c}), \tag{7}$$

even though it is often considered an equality when high accuracy is not required [16] [17]. Therefore, while it is possible to obtain a corrected image coordinate from a distorted one, some approximations are necessary to obtain a distorted coordinate from an ideal image coordinate. Heikkilä presented an approximate inverse model which is able to fit observations with an average error of 0.005 pixels [7]. This additional error remains significant if a very high accuracy is sought, as it is here, so that only the forward model (3) will be used in what follows. If needed, the parameters of the inverse model can easily be obtained once the parameters of the forward model have been determined.

3 Estimation of the Image Position of the Calibration Markers

Accurate location of 3D features in the image plane is one of the most important steps in camera calibration. For reasons of accuracy and stability, the use of larger, typically circular markers is desirable. As the markers encompass increasingly larger image areas, however, non-linear distortions significantly affect the markers' image representation to the point at which an elliptic approximation for their outline (such as [5]) becomes insufficient. The approach chosen here is to use the full camera

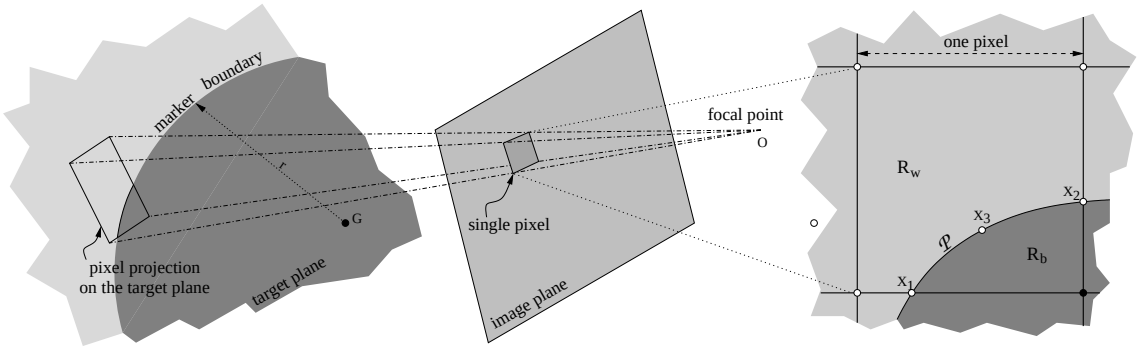

Fig. 2. The model of the image intensity at the pixel level. A single pixel of the image plane (center) is projected onto the target plane (left) where it intersects with the boundary of a marker of center G and radius r. The resulting black area and white area encompassed by this pixel projection are estimated by calculating their approximative boundary $\mathcal{P}$ (right) in the image plane. The ratio of these two areas can then be used to estimate the pixel intensity.

model, with all its non-linear distortion terms, to match the 2D discrete image data directly with the 3D target and marker models. Optimizing the match between synthetic and observed images of the marker provides the location of the marker. In addition, the camera model, the target model and the marker model do not depend on each other, leading to a more flexible approach.

3.1 Image modeling

Ray tracing can be used to build an image of the target from its model providing that the camera is already calibrated. However, the camera cannot be considered calibrated since the intent is to use ray tracing during the calibration phase itself. This quandary will be solved, as described in Section 4, by proceeding iteratively. Here, meanwhile, the camera is assumed to be already calibrated. The models of the calibration markers are also known, both in 3D shape and in intensity pattern. For the sake of simplicity, the marker model is chosen here as a black disc on a white background resting on a planar surface.

Each pixel of the image sensor collects light arriving within a pyramid delimited by the four outbound projection rays from each corner of the pixel (Fig 2). If non-linear distortions can be considered negligible at the pixel level, the intersection of this pyramid with a planar marker is a quadrilateral which consists of a white area, a black area or both. Furthermore, suppose that the target surface is lambertian and the lighting is uniform at the pixel level. Defining R_w as the white area of the pixel on the image plane and R_b as the black area, the modeled intensity $I_m(\varphi)$ of a pixel φ is then proportional to the ratio between the white area R_w and the total area $R_w + R_b$ of the pixel:

$$I_m(\varphi) = K \frac{R_w(\varphi)}{R_b(\varphi) + R_w(\varphi)}, \tag{8}$$

where K is a proportionality factor (for example, K=255 to model 8-bit images). To determine the intensity of each pixel, the relative amount of white area encompassed by each image pixel φ must thus be calculated. This process, depicted in Fig. 2, is based on one elementary ray tracing operation $\mathcal{R}$ which uses the complete camera model (1)-(6) to associate a binary marker intensity with every image coordinate $\mathbf{p}$ and center $\mathbf{g}$ of the circular marker. Applied to $\mathbf{p}$ and $\mathbf{g}$, these four transformations yield two points $\mathbf{P}$ and $\mathbf{G}$ of the target plane, from which one can readily infer the intensity of $\mathbf{P}$.

The ray tracing operation $\mathcal{R}$ is used multiple times within the same pixel to determine three points X_i of the boundary between the black and white regions. A parabolic approximation $\mathcal{P}$ of the boundary is then calculated and integrated to find R_w and R_b.

3.2 Matching the synthetic and observed images

The observed intensity $I_o(\varphi)$ also depends on the ratio of white and black areas, but with a transformation $\mathcal{L}$ that is not known and may vary within an image region, for example due to uneven lighting. To bypass the lack of knowledge about $\mathcal{L}$, two quantities B_w and B_b are calculated:

$$B_w = \frac{\sum_{\varphi \in \mathcal{U}} I_o(\varphi) I_m(\varphi)}{\sum_{\varphi \in \mathcal{U}} I_m(\varphi)}, \tag{9a}$$

$$B_b = \frac{\sum_{\varphi \in \mathcal{U}} I_o(\varphi) \left(K - I_m(\varphi)\right)}{\sum_{\varphi \in \mathcal{U}} \left(K - I_m(\varphi)\right)}, \tag{9b}$$

where $\mathcal{U}$ is the image region where the marker has been detected. B_w is a measure of the match between the bright areas of the observed image I_o and the synthetic image I_m; it will be maximal when the bright areas of the synthetic image are aligned with the bright areas of the observed image. Similarly, B_b measures the match of the dark areas. The difference $B_w - B_b$ is thus a measure of the match between the synthetic and observed images of the marker, and can be used to estimate the optimal position $\hat{\mathbf{g}}$ of the observed marker's center $\mathbf{g}$ in the image plane:

$$\hat{\mathbf{g}} = \underset{\mathbf{g}}{\operatorname{argmax}} \left(B_w - B_b\right). \tag{10}$$

To start the iterative process, an initial estimate of the observed marker center $\mathbf{g_0}$ is obtained by an intensity-weighted measure of its center of gravity (COG). The marker center $\hat{\mathbf{g}}$ is then updated to maximize the correlation between the synthetic and observed images of the markers using a Nelder-Mead optimization scheme [18].

This marker localization technique takes the full camera model into account, since the latter is at the heart of the image modeling in $\mathcal{R}$, thereby avoiding approximate camera models [7] or marker models in the image plane [5]. This approach also avoids direct comparison between an observed luminance and a theoretical luminance, which would be difficult to calculate in the presence of image artifacts such as illumination gradients. Finally, it is not based on differential measurements, and is

therefore expected to be more robust to noise. Localization accuracies of below 0.0005 pixels and having negligible bias are typically obtained using this technique with large markers and known camera model parameters, which is close to the theoretical estimate of 0.00033 pixels [19].

4 Geometric camera calibration

The non-linear geometric camera calibration itself is a well-documented process that iteratively optimizes the intrinsic and extrinsic parameters of the camera model, given a set of matched image coordinates $[x_i, y_i]^T$ and world coordinates $[X_i, Y_i, Z_i]^T$ obtained from known markers (see, for example, [1] [3] [7] [20]). However, the procedure described in Section 3 requires the camera parameters (the calibration output) to estimate the image location of the markers (the calibration input). This quandary can be solved by nesting the iterative non-linear camera calibration process within an iterative estimation of the image position of the markers. The complete calibration process then consists of three steps: 1) the determination of initial estimates; 2) the camera calibration itself; and 3) the accurate localization of the markers in the image plane. Steps 2 and 3 are repeated until convergence is reached.

In what follows, multiple views of a planar target are used for calibration [3]; other cases, such as a single view of a trihedral target, can be easily derived. When multiple views are used, the last view can yield useful extrinsic parameters since the camera and target will not subsequently be moved. However, the calibration process will provide estimations for the extrinsic parameters of each view.

4.1 Initial estimates

Five initial estimations are necessary before the calibration can start. First, the world coordinates of the calibration markers are obtained from the known target model. Second, initial estimates of the image positions of the calibration markers are calculated by analyzing each calibration image. Third, a match between the world and image coordinates in each view is determined using the Vogel approximation method (VAM) [21]. Fourth, the internal camera parameters are set to

$$f_x = f_y = f/d_x, \quad u_0 = s_x/2, \quad v_0 = s_y/2, \quad k = 0, \qquad (11)$$

where f is the assumed focal length, $[s_x, s_y]$ is the image size and d_x is the pixel pitch. The non-linear distortions are all set to zero. Fifth and last, initial estimates for the camera orientation and position with respect to the target (i.e., the extrinsic parameters) are obtained for each view of the target using Dementhon's algorithm [22] on the initial image and world coordinates of the markers.

4.2 Optimization

Once initial estimates are obtained, an off-the-shelf Levenberg-Marquart optimization scheme [23] on the intrinsic and extrinsic camera parameters is used to minimize the difference between observed and synthetic image coordinates simultaneously, for all views of the calibration target. The synthetic image coordinates are usually obtained by projecting the corresponding world coordinates onto the image plane and applying distortions to them. However, this requires an approximate inverse model, which cannot be used here for reasons of accuracy. As Fig. 1 shows, this difference can be computed without approximations in the space of corrected image coordinates $\mathbf{a_c}$, thereby splitting the application of the camera model in two: 1) the observed image coordinates are corrected using the non-linear distortion parameters; and 2) the synthetic image coordinates are obtained by applying (1) and (2) to the 3D positions of the target markers. The difference between these two image coordinates can then be used as the criterion for the optimization.

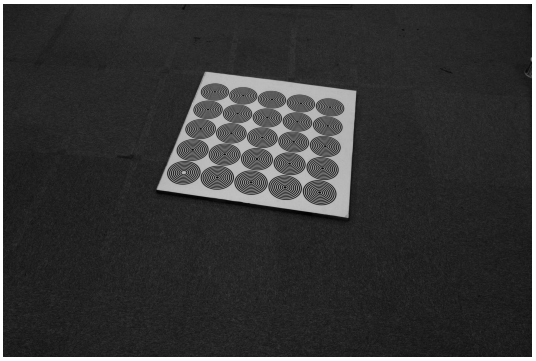

Fig. 3. A real calibration image with a 25-marker target in its center.

A refinement is added to the process when the optimization has reasonably converged: the 3D positions of the calibration markers are allowed to vary and thus become part of the set of unknown variables. This was proposed by Lavest [24] and can reduce the manufacturing cost of accurate targets. Errors in the marker's position of up to 10% in their spacing can be compensated for in this fashion.

After convergence, a more accurate position of the markers can be obtained using the latest camera parameter set in the relocation process described in Section 3. A new optimization of the calibration parameters can then be recomputed using the latest marker positions, and reiteration of these two processes leads to two interleaved optimizations. The convergence of the inner camera calibration is usually fast, requiring less than 50 iterations, and the convergence of the whole outer optimization, including image measurements, is completed in three to five cycles.

5 Experiments

Experiments were conducted using a white planar calibration target made of 25 calibration black markers in a 5×5 matrix arrangement. Each calibration marker is 106 mm in diameter and their spacing is 110 mm. A pattern of concentric circles is used instead of a simple disc (Fig. 3), thereby allowing the total length of the interface between the black and white regions of a single marker to grow to over 3 meters. Each of the eight circles has a thickness of 3 mm. To avoid orientation ambiguity, the reference marker of the target does not possess a central dot (Fig. 3, lower-left marker).

A set of 69 calibration pictures were taken with a Nikon D2h digital camera equipped with a Nikon AFS 12-24 mm DX zoom lens set to 24 mm and f/16. The resolution of this camera is 2482×1648 pixels for a sensor size of 24×16 mm. The target and the camera were placed at different relative locations to cover both the whole image (Fig. 4) and the viewed volume. The distances between the target and the camera range from ≈1m to ≈4m. The marker diameters in the image plane range from 76 to 230 pixels (average 150 pixels). The uncompressed raw images had their exposure optimized and were demosaiced

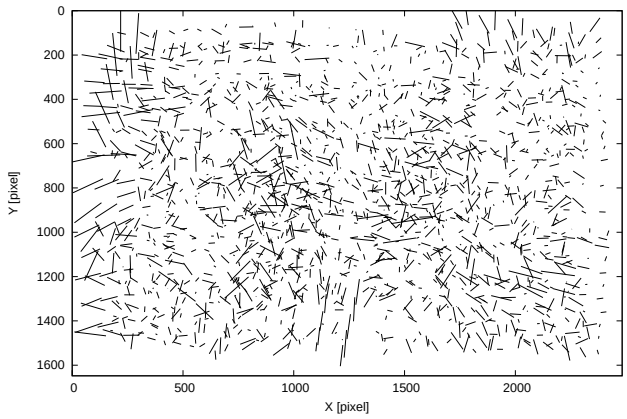

Fig. 4. True errors for the positions of the 1725 image points across the image plane (enlarged 20000 times)

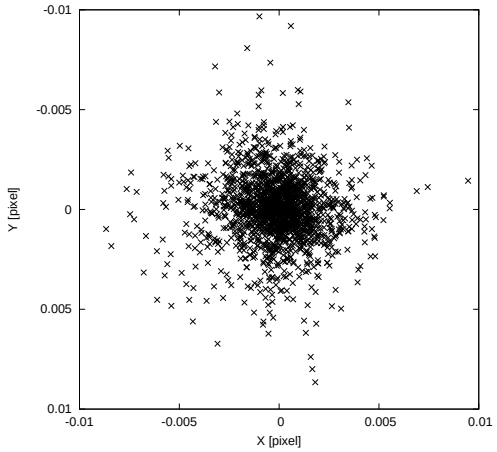

Fig. 5. Isotropic character of the true error for the position of the 1725 image points

using the AHD technique [25]. The number of unknowns is 495 and the number of equations 3450.

The global accuracy of the calibration can be estimated from the residual of the optimization, defined as the root mean square (RMS) of the differences between the projection of each marker's 3D center onto the image plane and the measured image coordinates of the center of the same marker. However, when using synthetic images, all internal and external parameters of the calibration are known and a better estimate, the true pixel error (TPE), can then be computed similarly but with true image coordinates instead of measured ones.

5.1 Synthetic images

The accuracy of the calibration process is estimated by applying it to a set of synthetic calibration images built with known marker positions, intrinsic parameters and extrinsic parameters. A first calibration of the real image set was performed, resulting in a residual of 0.045 pixels. The estimated parameters were then rounded and used to create this set of synthetic calibration images with the sub-pixel intensity estimation detailed in Section 3. The image was then blurred by a 3-by-3 Gaussian filter and additive noise was added (σ=2).

The calibration of the synthetic image set returns a residual of 0.0065 pixels and a TPE of 0.0021 pixels. We have observed

TABLE 1
Retrieved calibration parameters and their errors

	True value	Average bias	Theoretical std. dev.	Observed std. dev.
f_x	2475.0	0.0328	0.00286	0.00340
f_y	2475.0	0.00521	0.00217	0.00245
u_0	1240.0	-0.733	0.0525	0.0631
v_0	790.0	-0.300	0.0493	0.0506
u_d	1180.0	0.351	0.0546	0.0699
v_d	836.0	-0.219	0.0450	0.0500
k	0.0001	0.00000563	0.000000517	0.000000578
a_0	0.01	0.0000978	0.00000614	0.00000688
a_1	0.9	-0.000468	0.000110	0.000134
a_2	-7.0	-0.00202	0.000879	0.00110
a_3	23.0	0.0230	0.0317	0.00406
a_4	-28.0	-0.0463	0.0424	0.00549
p_0	26.0	-0.342	0.0263	0.0312
p_1	-12.0	-0.181	0.0222	0.0254
p_2	-0.1	-0.00223	0.00400	0.00406
p_3	-2.0	0.114	0.00394	0.00481
s_0	-26.0	0.361	0.0264	0.0315
s_1	14.0	0.191	0.0222	0.0252
s_2	0.5	0.0128	0.00643	0.00607
s_3	2.0	-0.165	0.00506	0.00557
t_x	-244.676	0.449	0.0297	0.0335
t_y	354.410	0.165	0.0247	0.0292
t_z	1349.413	0.0290	0.00713	0.00797
α	178.029	0.0242	0.00176	0.00243
β	135.270	-0.00976	0.00101	0.0311
γ	176.598	0.0174	0.00127	0.00224

this 3-fold difference between these two figures in all successful calibrations. Systematic errors along the x and y axes are negligible at -0.00015 and -0.00007 pixels, respectively. The standard deviation of the TPEs is 0.0011. Fig. 4 shows that the TPEs for the 1725 points used in this calibration are uniformly distributed across the image plane, and Fig. 5 shows the isotropic character of the distribution of the TPEs.

The bias and standard deviation of the estimated camera model parameters were obtained from the calibration of 100 image sets and are presented in Table 1. The biases are small and the theoretical errors, estimated as described in [7], are very close to the observed errors. All extrinsic parameters cannot be shown due to their large number (69×6); one set is presented for a typical calibration image.

Compared to the reference TPE of 0.0021 obtained with large markers (eight circles for a total interface length of 3098 mm), smaller markers increase the TPE: 0.0035 (five circles, 1382 mm), 0.0066 (two circles, 346 mm) and 0.0102 (one circle, 151 mm). Our approach shows scalability, achieving results that are on a par with Heikkilä's [7] when small markers are used. The global performance of our calibration technique remains stable for a smaller number of images, with TPEs of 0.0022 (50 images), 0.0024 (30 images), 0.0029 (10 images) and 0.0035 (six images). The accuracies of the camera model parameters are more affected by a limitation of the number of calibration images, and are roughly inversely proportional to the number of images. Accuracies are unpredictable when fewer than 10 images are used.

The importance of using a full non-linear camera model during the localisation of calibration markers can be emphasised by attempting this localisation with a reduced set of non-linear

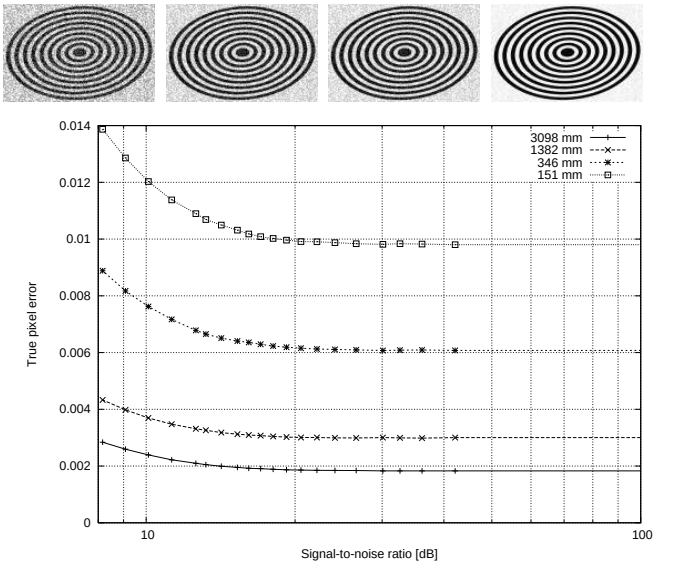

Fig. 6. Evolution of true pixel error (TPE) with increase in uniform noise. Miniatures (left to right) show a target marker for SNRs of 8.2 dB (A_n=50), 11.3 dB (A_n=35), 16.1 dB (A_n=20) and 22.1 dB (A_n=10).

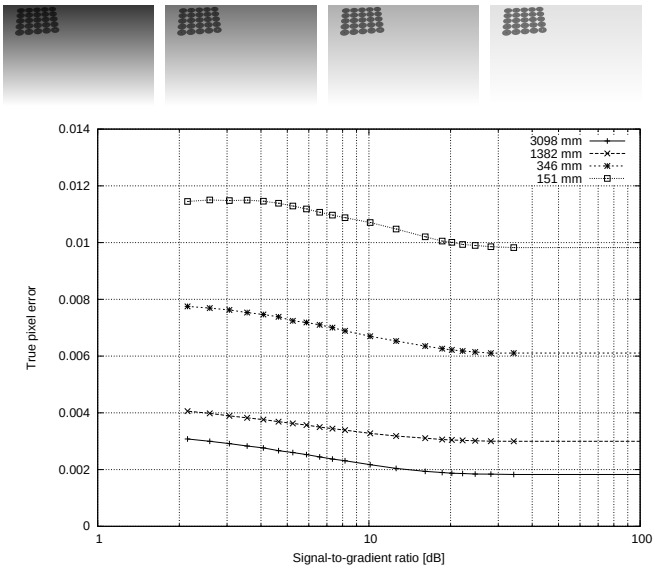

Fig. 7. Influence of a luminance gradient. Miniatures (left to right) show a calibration image for gradients of 2.1 dB, 5.2 dB, 10.1 dB and 18.6 dB.

parameters. Omitting s_2 and s_3 leads to a TPE of 0.0029, a small but measurable loss in accuracy. Further omitting p_2 and p_3 leads to a TPE of 0.0034 pixels and further omitting a_3 and a_4 leads to a TPE of 0.0465 pixels. Omitting all non-linear distortion terms results in a large TPE of 0.446 pixels.

5.2 Noise tests

Various levels of different imperfections were applied to the image to assess the robustness of the large-marker approach: uniform noise, an illumination gradient, Gaussian low-pass filtering, limitation of the dynamic range and JPEG compression. In addition, two target imperfections were tested: mis-location of the calibration markers, and the non-planarity of the target's surface. The number of concentric circles in each target marker was also varied to investigate the importance of using large patterns with a long interface.

5.2.1 Uniform noise

Robustness to noise is important for high-resolution images, which often display more noise due to their smaller photosites. The effect of an additive uniform noise on calibration accuracy is shown in Fig. 6. The image SNR is defined in dB as $20\log(A_s/A_n)$, where A_s is the signal amplitude (here, 128) and A_n is the noise amplitude [10]. The accuracy remains stable over most of the noise range and increases steadily with the interface perimeter of the calibration marker. For large calibration markers, calibration accuracy remains below 0.003 pixels even with a very low SNR of 8.2 dB, clearly outperforming other calibration techniques [3] [8]. This robustness is attributed to the averaging of the noise across the whole effective perimeter of the calibration marker. The accuracy obtained for the smallest calibration markers matches the results of Heikkilä [7].

5.2.2 Illumination gradient

Illumination gradients are difficult to avoid in real-life situations. In this test (Fig. 7), a vertical luminance gradient was applied to the whole image. The gradient level is defined in dB as $20\log(2A_s/(G_M - G_m))$, where G_M and G_m are respectively the maximal and minimal luminance of the image. This gradient is not constant with respect to the target, but with respect to the camera. Therefore, errors in marker localization due to the gradient were expected to be more pronounced on internal parameters such as the focal lengths and the principal point. However, no significant change was observed in these parameters even with a steep gradient. The effect of the gradient is in fact minimal and almost linear over the wide test range.

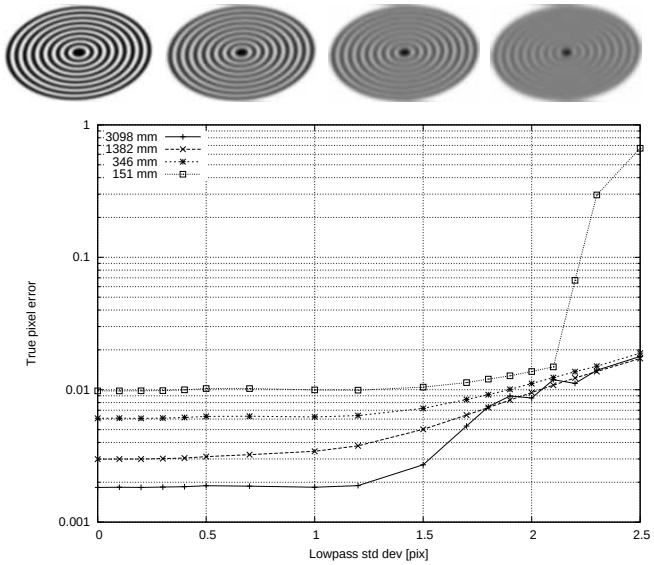

Fig. 8. Evolution of the TPE with Gaussian low-pass filtering. Miniatures (left to right) show a target marker for low-pass filter widths of σ=1.0, 1.5, 2.0 and 2.5 pixels.

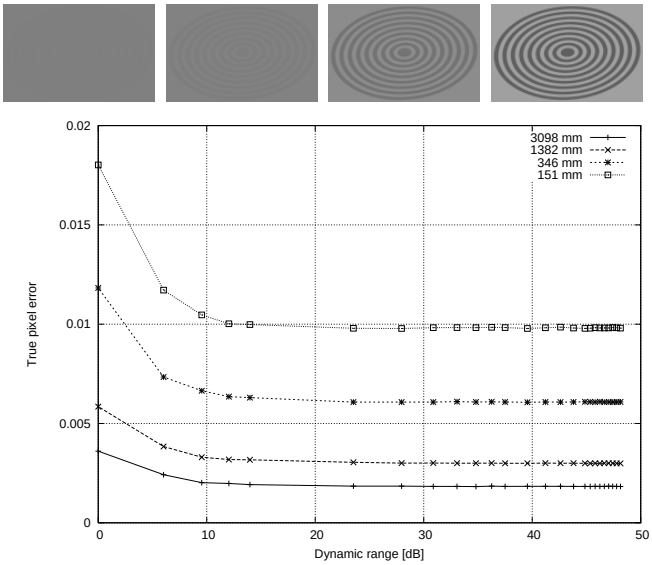

Fig. 9. Evolution of the TPE with dynamic range. Miniatures (left to right) show a target marker for dynamic ranges of 0 dB (binary image), 9.5 dB (three levels), 23.5 dB (15 levels) and 34.8 dB (55 levels).

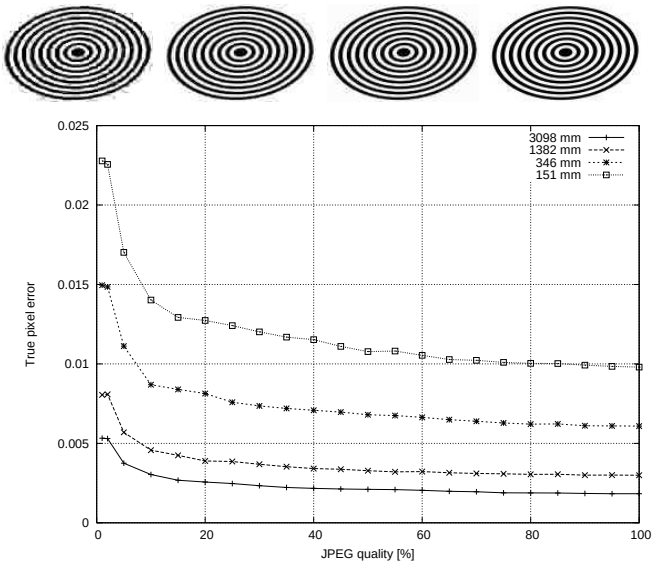

Fig. 10. Evolution of the TPE with JPEG compression. Miniatures (left to right) show a target marker for JPEG qualities of 2%, 10%, 20% and 80%.

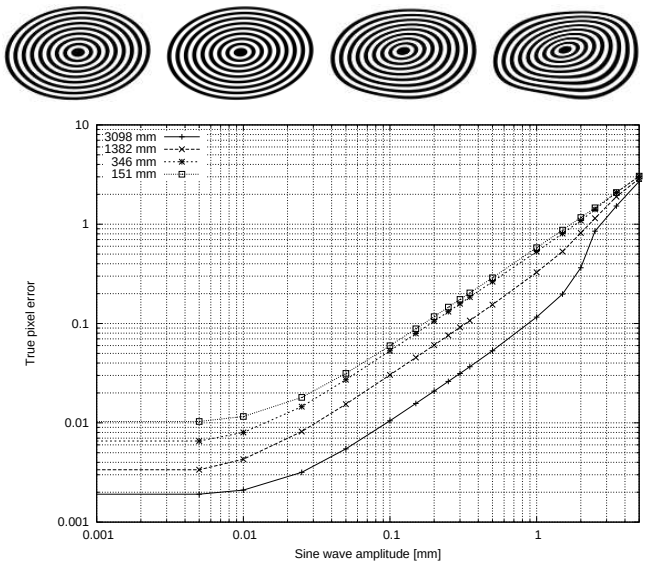

Fig. 11. Evolution of the TPE with target non-planarity. Miniatures (left to right) show a target marker for surface amplitudes of 0.025 mm, 0.25 mm, 2.5 mm and 5.0 mm.

5.2.3 Gaussian low-pass filtering

Another common defect of real images is low-pass filtering, which can be due to optics or post-processing (e.g., in the case of Bayer filters). Our tests (Fig. 8) show much better results than those achieved by Redert et al., and are comparable to those of Heikkilä [7], having good performance up to a standard deviation of σ=1.5. The maximal acceptable standard deviation also depends on the image distance between two circles of the marker: when σ approaches this distance, the advantage of using large markers is quickly lost due to the strong overlapping between the concentric circles. The advantage of multiple circles being lost, results converge towards the same curve for broader low-pass filters. For the smallest markers (151 mm perimeter), the localisation fails to converge for broad low-pass filtering. In such cases, the poor contrast over a small area does not allow the localisation process to lock on the faint marker's edges.

5.2.4 Dynamic range

For N_i luminosity levels, the dynamic range is defined in dB as $20 \log N_i$. Our tests (Fig. 9) show that the algorithm is almost completely insensitive to reductions of the dynamic range. The error starts to climb steadily from a limited range of 20 dB, but even at the minimal range of 0 dB (a binary image) the effect of the dynamic range is limited to a doubling of the error. From 20 dB to 48 dB (255 levels of gray) the error is inversely proportional to the effective perimeter of the marker. This increasing error is due to discretization. For instance, in a binary image with a marker perimeter of 151 mm and a pixel size of 1 mm, only about 151 discrete levels are available for the optimization function (10), which seems to be not enough to converge smoothly to a global optimum.

5.2.5 JPEG compression

Although this is less important than other tests, an algorithm that is robust to image compression is advantageous in the case of high-resolution images and videos and would, for instance, facilitate remote calibration of cameras. JPEG compression was applied at different levels of quality to the calibration images prior to calibration (Fig. 10). The errors grow steadily but remain below 130% of the minimum error even when the quality reaches 20%. For a quality of 1%, the worst calibration accuracy remains under 0.025 pixels.

5.2.6 Target non-planarity

The assumption that the target is planar was made in Section 3, but in practice perfect planarity cannot be achieved. To evaluate the effect of target non-planarity, sets of synthetic calibration pictures were created using a bi-dimensional sine-wave ripple, rather than a plane, as the support surface. To obtain a different distortion for each marker, the period of the

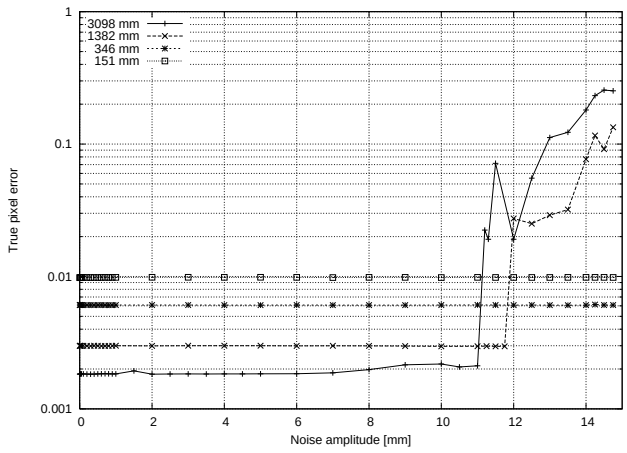

Fig. 12. Evolution of the TPE with mis-location of target markers.

ripple was set to 90 mm, a non-multiple of the marker interval, and the center of the ripple was offset by 20 mm along the x axis. The amplitude of the ripple was varied from zero (a plane surface) to 5 mm. The effect of non-planarity is significant (Fig. 11), with a TPE of 0.5 pixels for an amplitude of 2 mm, as was also observed by Zhang [3]. In fact, an amplitude as low as 0.02 mm produces a measurable decrease in accuracy. The curves converge when non-planarity becomes very large, because matching between the distorted shape and the circular model can only be performed close to the target center, which remains relatively elliptic due to its small size relative to the sine wave period. Striving for a planar target is therefore of great importance.

5.2.7 Mis-location of calibration markers

Random noise was added to the initial estimates of the world coordinates of the markers to verify that their optimization was effective (Fig. 12). The calibration remains stable even when noise of up to 10% of the marker distance (10 mm) is applied, which shows the effectiveness of the approach of Lavest [24]. For larger markers, large errors in their location lead to poor initial estimations of their position during the first localization step due to incorrect estimations of the perspective distortions. These errors are difficult to compensate for in the later stages, and the algorithm fails to converge. However, this has few practical consequences since the threshold for this phenomenon is very high (Fig. 12).

5.3 Real images

Noise tests show that our method is robust to many sources of imperfection, even when such imperfections greatly affect the calibration images. However, synthetic images may still depart from real images. To investigate this, a number of image imperfections were added to the synthetic image set of Section 5.1 to approach the real image set: blurring (σ=0.7), limitation of the dynamic range to 45 dB, application of a luminance gradient (18.6 dB) and addition of noise (SNR of 33 dB). The amplitudes of these artifacts were determined by visual inspection of the synthetic and real images. Finally, a target non-planarity of 0.1 mm was applied as described above. The initial measurements of the positions of the markers were also perturbed with uniform noise (amplitude of 0.1 mm). The calibration of this set of realistic but synthetic images yields a residual of 0.031 pixels. The greater part of this residual (0.025 pixels) is due to target non-planarity. The calibration of the corresponding real image set has a slightly higher residual of 0.045 pixels, which comes from artifacts that were not taken into account such as optical diffraction, lens aberration and thermal effects. This difference between the residual obtained from synthetic images (here, 0.0065 pixels) and real images (here, 0.045 pixels) was previously observed by Heikkilä [7] (synthetic: 0.020 pixels; real: 0.061 pixels) and highlights the importance of using controlled experimental conditions and high-quality hardware to achieve accurate calibration.

6 Conclusions

We have presented a geometric camera calibration technique that uses large markers to provide simultaneously very robust and very accurate results. The proposed approach has a number of advantages over existing techniques. First, it involves separated models for the camera, the target and the markers. Each model can therefore be adjusted independently, for instance by adding higher-order terms to the camera model, by using different marker patterns or by extending the ray tracing to non-lambertian surfaces. Second, the exact world positions of the markers do not need to be known exactly; only the marker's pattern must be very accurate and planar. This both lowers the cost and enhances the durability of accurate calibration targets. Third, the approach is scalable to lower-resolution images. However, as the resolution increases, so does the pixel accuracy and the robustness of the calibration, because larger marker patterns can be used. This combined increase in both resolution and accuracy makes high-resolution imaging particularly attractive. Our experiments show that a precision of 0.6 to 4 ppm can be obtained on an image line of 2500 pixels if the calibration is performed in very controlled conditions. Finally, the procedure for the localization of markers can be used for measurements after calibration has been performed. Real-time processing can be achieved, with a synthetic marker image being generated in less than 0.003 seconds on standard hardware. Our approach is thus not limited to calibration but offers a complete solution for high-accuracy measurements for machine vision.

References

[1] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE Trans. Robot. Autom.*, vol. 3, pp. 323–344, Aug. 1987.

[2] J. Salvi, X. Armangu, and J. Batlle, "A comparative review of camera calibrating methods with accuracy evaluation," *Pattern Recognition 35 (7) (2002) 1617–1635.*, vol. 35, no. 7, pp. 1617–1635, 2002.

[3] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.

[4] H. Zollner and R. Sablatnig, "Comparison of methods for geometric camera calibration using planar calibration targets," in *Proc. of the 28th Workshop of the Austrian Association of Pattern Recognition (AAPR'04)*, 2004, pp. 234–244.

[5] J. Heikkilä, "Moment and curvature preserving technique for accurate ellipse boundary detection," in *Intl. Conf. on Pattern Recognition*, vol. 1, Brisbane, Australia, Aug. 1998, pp. 734–737.

[6] J. Heikkilä and O. Silvén, "A four-step camera calibration procedure with implicit imagecorrection," in *Intl. Conf. on Computer Vision and Pattern Recognition (CVPR'97)*, San Juan, Puerto Rico, June 97, pp. 1106–1112.

[7] J. Heikkilä, "Geometric camera calibration using circular control points," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1066–1077, Oct. 2000.

[8] A. Redert, E. Hendriks, and J. Biemond, "Accurate and robust marker localization algorithm for camera calibration," in *Proc. of the First Intl. Symposium on 3D Data Visualization and Transmission (3DPVT'02)*, Padova, Italy, June 2002, pp. 522–525.
[9] L. Robert, "Camera calibration without feature extraction," *Computer Vision and Image Understanding*, vol. 63, no. 2, pp. 314–325, Feb. 1996.
[10] C. McGlone, E. Mikhail, and J. Bethel, Eds., *Manual of Photogrammetry*, 5th ed. Bethesda, MD, USA: Am. Soc. of Photogrammetry and Rem. Sens., 2004.
[11] R. Jain, R. Kasturi, and B. G. Schunck, *Machine Vision*. New York, USA: McGraw-Hill, 1995.
[12] O. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press, Cambridge, Massachusetts, 1993.
[13] J. Weng, P. Cohen, and M. Herniou, "Camera calibration with distorsion models and accuracy evaluation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 10, pp. 965–980, Oct. 1992.
[14] J.-M. Lavest, G. Rives, and J. Lapresté, "Underwater camera calibration," in *Proc. European Conf. on Computer Vision (ECCV'2000)*, vol. 2. Dublin, Ireland: Springer-Verlag, June 2000, pp. 654–668.
[15] R. Klette, K. Schlüns, and A. Koschan, *Computer Vision*. Singapore: Springer, May 1998.
[16] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge University Press, 2003.
[17] O. Faugeras and Q.-T. Luong, *The geometry of multiple images*. Cambridge, MS, USA: MIT Press, 2001.
[18] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer Journal*, vol. 7, pp. 308–313, 1963.
[19] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*. Addison-Wesley, 1993, vol. 2.
[20] J.-M. Lavest and M. Dhome, "Short focal length camera calibration," in *RFIA2000 Conference*, vol. 3, Paris, France, Feb. 2000, pp. 81–90.
[21] N. V. Reinfeld and W. R. Vogel, *Mathematical Programming*. Englewood Cliffs, New Jersey, USA: Prentice Hall, 1958.
[22] D. DeMenthon and L. S. Davis, "Model-based object pose in 25 lines of code," *Intl. J. of Computer Vision*, vol. 15, no. 1-2, pp. 123–141, June 1995.
[23] M. Galassi, J. Davies, J. Theiler, G. Jungman, M. Booth, and F. Rossi, *GNU Scientific Library Reference Manual*, 2nd ed., B. Gough, Ed. Bristol, United Kingdom: Network Theory Ltd., Aug. 2006.
[24] J.-M. Lavest, M. Viala, and M. Dhome, "Do we really need an accurate calibration pattern to achieve a reliable camera calibration?" in *Proc. 5th European Conf. on Computer Vision (ECCV'98)*, vol. 1. Freiburg, Germany: Springer-Verlag, June 1998, pp. 158–174.
[25] K. Hirakawa and T. W. Parks, "Adaptive homogeneity-directed demosaicing algorithm," *IEEE Trans. Image Process.*, vol. 14, no. 3, pp. 360–369, Mar. 2005.

Damien Douxchamps received his PhD in Electrical Engineering from the Université catholique de Louvain, Belgium, in 2004. He was invited scientist at the National Taiwan University and the Advanced Telecommunications Research Institute. He is currently assistant professor at the Nara Institute of Science and Technology. His research interests include ocean engineering, robotics and 3D computer vision. He is a member of the IEEE and the Oceanic Engineering Society.

Kunihiro Chihara received the BS degree from Osaka University, Japan, in 1968 and the MS and PhD degrees in engineering science from Osaka University in 970 and 1973, respectively. He is currently a dean and professor at the Graduate School of Information Science, Nara Institute of Science and Technology, Japan. His research interests include medical ultrasonic imaging, signal processing, ubiquitous computing, and the application of multimedia. He is a member of ISCIE, JSMBE, JSUM, and the IEEE.