Camera Calibration: Active versus Passive Targets

Christoph Schmalz,[1,2] *Frank Forster*[1] *, Elli Angelopoulou*[2]

[1]Siemens AG, CT T HW2

Otto-Hahn-Ring 6, 81739 Munich, Germany

[2]University of Erlangen-Nuremberg, Pattern Recognition Lab

Martensstrasse 3, 91074 Erlangen, Germany

Abstract

Traditionally, most camera calibrations rely on a planar target with well-known marks. However, the localization error of the marks in the image is a source of inaccuracy. We propose the use of high-resolution digital displays as *active* calibration targets to obtain more accurate calibration results for all types of cameras. The display shows a series of coded patterns to generate correspondences between world points and image points. This has several advantages. No special calibration hardware is necessary since suitable displays are practically ubiquitious. The method is fully automatic, no identification of marks is necessary. For a coding scheme based on phase shifting, the localization accuracy is approximately independent of the camera focus settings. Most importantly, higher accuracy can be achieved compared to *passive* targets like printed checkerboards. A rigorous evaluation is performed to substantiate this claim. Our active target method is compared to standard calibrations using a checkerboard target. We performed camera calibrations with different combinations of displays, cameras and lenses, as well as with simulated images and find markedly lower reprojection errors when using active targets. For example, in a stereo reconstruction task the accuracy of a system calibrated with an active target is five times better.

1

1 Introduction

A calibrated camera makes it possible to relate measurements in images to metric quantities in the world. The calibration process is therefore very basic and essential to every computer vision task that involves image based measurements. Many different ways of calibrating cameras have been developed [1]. So-called self calibration methods do not assume any knowledge about the scene [2]. They are for example very useful for autonomous navigation. Metrology applications aim for highest accuracy and typically use dedicated calibration targets with well-known marks. Some methods use three-dimensional targets [3] but planar targets are more common as they are easier to build and handle. We aim to offer an alternative to the latter. Digital displays that can be found on everybody's desk can be used as *active* calibration targets. This has many advantages. There is no need to manufacture and validate a special target. The displays are produced lithographically to a very high standard of accuracy. The marks on the target do not have to be laboriously identified in an error-prone manual process, as is often the case. Instead of marks, we propose the use of Structured Light coding scheme that is tolerant against defocusing, so the target does not have to be in focus for the calibration. This makes it possible to position the camera close to the target and easily cover the whole field of view. The displays typically have a tilt/swivel base so that different poses can be set up very comfortably. Example code to generate and decode the necessary images will be made publicly available at [4]. Lastly and most importantly the achievable accuracy (as measured in the RMS reprojection error) is comparable to the best published calibration results and much better than the typical values reached with passive targets.

Self-identification of the calibration marks can also be achieved in other ways, for example by using ARTags [5], but it comes for free with the active calibration. The idea of using a Structured Light coding scheme for camera calibration has also been proposed by [6], where it was used to undistort the images of a wide-angle camera in a model-free manner. In contrast, we perform a full camera calibration and recover the camera parameters, as they are needed for many tasks, for example 3D reconstruc-

tion. A similar active calibration approach is also briefly mentioned in [7] and [8] in the context of calibrating a catadioptric wide-angle camera. All these works focus on calibration for wide-angle imaging and do not include thorough quantitative performance comparisons with other calibration methods. We are of the opinion that active camera calibration has advantages for any camera, not only extreme wide-angle cameras where the traditional pinhole model breaks down. We substantiate that claim with experiments demonstrating a clear improvement in the reprojection errors.

The contributions of this work are:

- We generate virtual calibration marks from the correspondence maps which are backward compatible with any classic calibration algorithm.

- We introduce a correction for the refraction effects caused by the glass plate covering the display.

- We perform an extensive comparative evaluation and show that the active calibration technique yields more accurate calibration results compared to a passive calibration with a checkerboard target.

2   Prior Work

A camera calibration consists of two parts. The external calibration refers to the camera pose relative to fixed world coordinate system. It maps the camera coordinate system, defined by the camera image plane, to the world coordinate system, often defined by the calibration target (see figure 2.1). It can be described by 6 parameters - 3 for rotation and 3 for translation. Additionally, there are intrinsic parameters, which depend on the camera model that is employed. For an ideal pinhole camera, this is the principal point $[u_0, v_0]$ and the scale factor $[f_x, f_y]$. In homogenous coordinates the perspective projection mapping the point $[X_w, Y_w, Z_w, 1]^T$ in the world coordinate system to the point $[u, v, 1]$ in the image coordinate system (see figure 2.2) can be written as

3

$$
\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = AP_0 \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{2.1}
$$

$$
A = \begin{bmatrix} f \cdot d_x & 0 & u_0 \\ 0 & f \cdot d_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \qquad P_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}
$$

where A is the camera matrix and P_0 the projection matrix. The coordinate transformation from world to camera coordinates is given by rotation and translation parameters $[R|t]$. The parameter f is the focal length, d_x and d_y are the pixel pitch in the x and y directions, and u_0 and v_0 are the pixel coordinates of the principal point of the image.
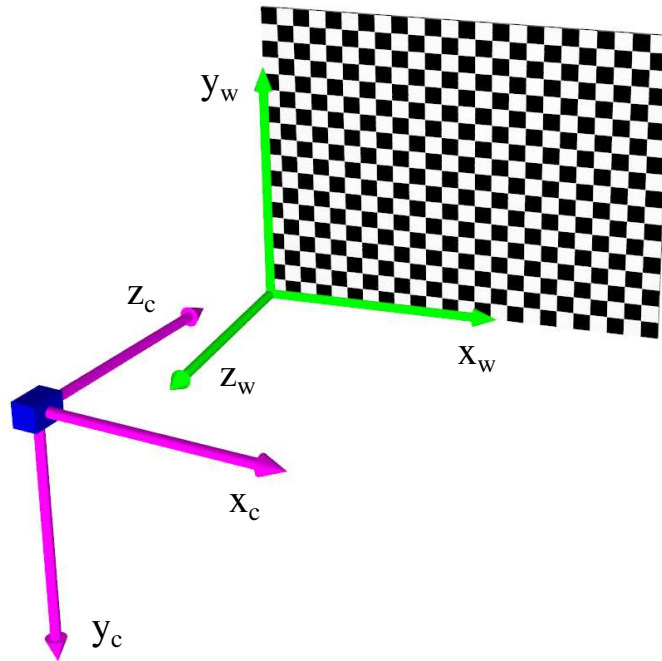

Fig. 2.1: The world coordinate system (top) is often defined by the calibration target. The camera coordinate system (bottom) is defined by the camera.
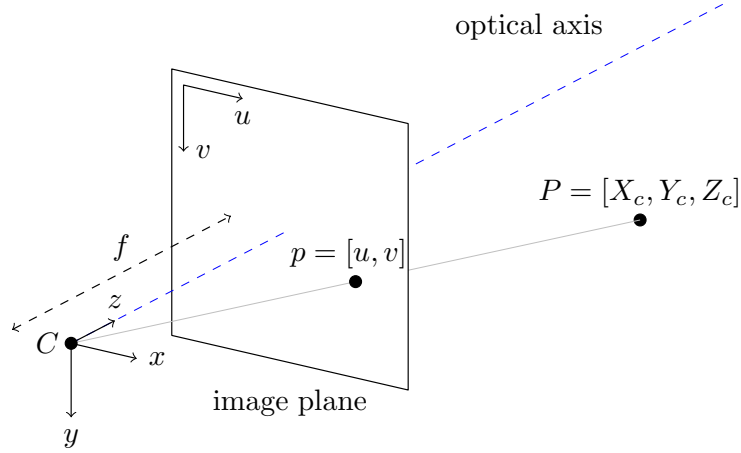
Fig. 2.2: Perspective projection of a point P in space to a point p on the image plane.

2.1 Camera Models

Real cameras use lenses and always exhibit some degree of image distortion, especially for wide-angle lenses. This also has to be modelled. The most common camera model is the pinhole model augmented by parameters for radial and tangential distortion. Tsai presented a relatively simple variant [9] with only two parameters for radial distortion. Heikkilä [10] used two parameters for radial distortion and two for tangential distortion. The model (and calibration algorithm) proposed by Zhang [11] is very popular, especially because it is available in the widely used OpenCV library [12] and as a Matlab toolbox [13]. These implementations support up to five parameters, three (k_1, k_2, k_3) for radial and two (p_1, p_2) for tangential distortion. Zhang's model maps undistorted image coordinates $[u_u, v_u]$ to distorted image coordinates $[u_d, v_d]$ via

$$\bar{u_d} = \bar{u_u}\left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6\right) + 2p_1 \bar{u_u}\bar{v_u} + p_2\left(r^2 + 2\bar{u_u}^2\right) \qquad (2.2)$$

$$\bar{v_d} = \bar{v_u}\left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6\right) + 2p_2 \bar{u_u}\bar{v_u} + p_1\left(r^2 + 2\bar{v_u}^2\right) \qquad (2.3)$$

where $u_{u,d}^- = u_{u,d} - u_0$, $v_{u,d}^- = v_{u.d} - v_0$ and $r^2 = \bar{u_u}^2 + \bar{v_u}^2$.

All these models assume that the principal point is also the center of distortion and that all rays pass through the pinhole. This generally holds for typical applications with

5

limited distortion, but not always [14, 15]. There are also fully generic camera models that work for any type of camera, like extreme wide angle or even non-single-viewpoint catadioptric cameras [16, 17]. However, such generic models deliver lower accuracy for narrow-angle cameras [18].

The actual input data to perform a calibration are lists of correspondences between points on the calibration target (in world coordinates) and their image coordinates. For the usual planar calibration targets, several different poses of the camera relative to the target are required. From these lists a calibration algorithm Tsai [9], Heikkilä and Silven [10], Zhang [11] calculates the internal camera parameters and the coordinate transformation for each pose.

2.2 Feature-based Calibration

The usual planar calibration patterns are checkerboards. The corners of the checkers are the fiducial marks. They are typically localized by intersecting lines fitted to the sides of the checkers or by looking for a saddle point in the gradient (as is implemented in OpenCV). An alternative target type consists of an array of circular dots. The centers of the dots are commonly computed via centroid methods, ellipse fitting to the contours, or deformable templates. However, for oblique viewing directions, the detected ellipse center is not the projection of the original circle's center and has to be corrected [10]. Image distortion also reduces the localization accuracy for dots [19].

There is little comprehensive information in the literature about achievable feature localization accuracy. Shortis et al. [20] tested different algorithms for circular marks. He reported errors in the range of a few hundreths of a pixel, but did not include noise in his analysis. Heikkilä [3] shows lighting-dependent shifts of up to 0.5 pixels in the location of circular marks. Mohr [21] found errors of around 0.1 pixels in corner local- ization. White and Schowengerdt [22] examine the effect of the point spread function on edge localization accuracy and find errors of up to 0.2 pixels. Mallon and Whelan [19] show errors around 0.1 pixels for circular marks (without distortion bias) and up to 0.03 pixels for a checkerboard target. Chen and Zhang [23] give errors of about 0.05 pixels for checkerboard corner localization.

The final calibration errors are within the same range. Heikkilä [3] claims that an accuracy of 0.02 pixels is a realistic goal. He achieves it for synthetic images and reports 0.061 pixels on real images. Douxchamps and Chihara [24] even reach 0.0065 pixels on synthetic images and 0.045 on real images. However, in his widely known paper [11], Zhang gives an RMS reprojection error of about 0.3 pixels. Albarelli et al. [25] achieve an initial error of 0.23 pixels but reduce it to 0.089 by additional bundle adjustment [26]. Fiala and Shu [5] also reach values of around 0.2 pixels. The differences between these figures might be due to outlier removal steps, differences in image and target quality, or simply different pixel sizes. In conclusion, an RMS reprojection error of 0.05 pixels seems to be a lower bound for a very careful calibration in an optimal environment, while errors up to 0.3 pixels are acceptable in day-to-day calibrations.

2.3 Active Targets

We compare the calibration results achieved with a traditional *passive* target to *active* digital displays as calibration targets. One practical advantage of the latter is the self-identifying nature of the patterns that can be shown on the display. Tedious manual mark identification therefore becomes unnecessary. Digital displays are suitable for calibration tasks as they are manufactured to very high precision using lithographic techniques. The pixel pitch is well-known, therefore pixel coordinates can be converted to metric 2D coordinates. One could simply show a checkerboard on the display and use that for calibration. However, such a method would still be subject to the noise-prone corner localization step. Instead, we propose the use of a series of coded patterns which can uniquely identify each individual pixel.

Such patterns are widely researched in the subfield of Structured Light. Many coding schemes are possible [27]. Phase shifting offers very high precision and dense coding. This is because it does not involve any differentiation or binarization steps but works directly with the measured image intensities in each pixel. We use two four-bucket phase shift sequences, one horizontal and one vertical, to determine the x and y components of the pixel coordinates. The recovered phase is ambiguous, however. To

obtain a unique phase value we have to "unwrap" the phase values. There are various ways to achieve this. In our case it is known that the target is flat, so naive unwrapping would work. But since calibration is not a time-critical task, we use additional Gray Code sequences. Details of this standard Structured Light coding scheme can be found for example in [28] or [29]. It has the additional advantage that the decoding is very simple. To facilitate the use of this method, code to generate and decode the necessary images will be made publicly available at [4]. All in all, a full pattern sequence consists of 4 images for the phase shift and 8 for the Gray Code (depending on display resolution). Some of the resulting camera images are shown in figure 2.3. Nonlinear display brightness is a concern for a high-quality phase shift, as the sinusoidal intensity pattern is distorted. However, the four-bucket phaseshift is robust against such errors [30], so a precise gamma calibration is not necessary.
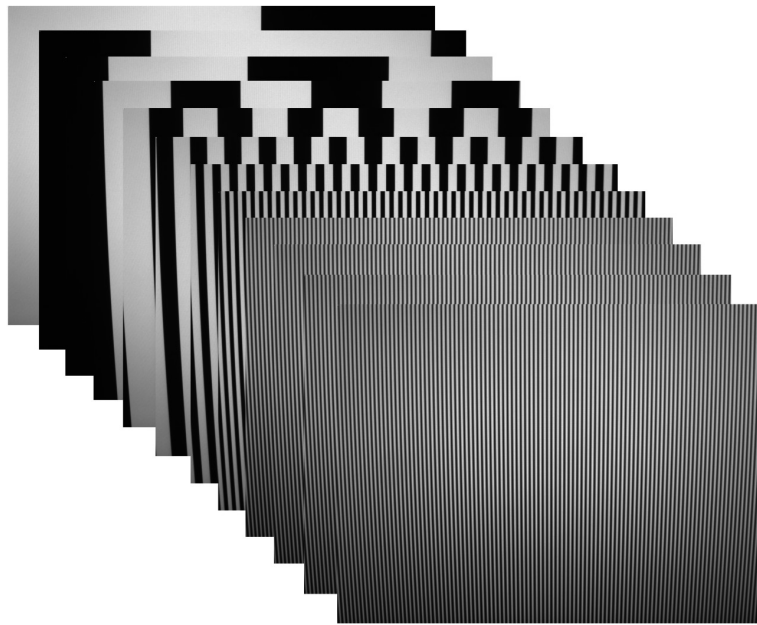

Fig. 2.3: A pattern sequence to uniquely identify all pixels of the display. Only the vertical component is shown. The four images in the front are used to compute ambiguous phase values. The images in the back form the Gray Code used to unwrap the phase.

Examples of the final unwrapped phase maps can be seen in figure 2.4. Using the
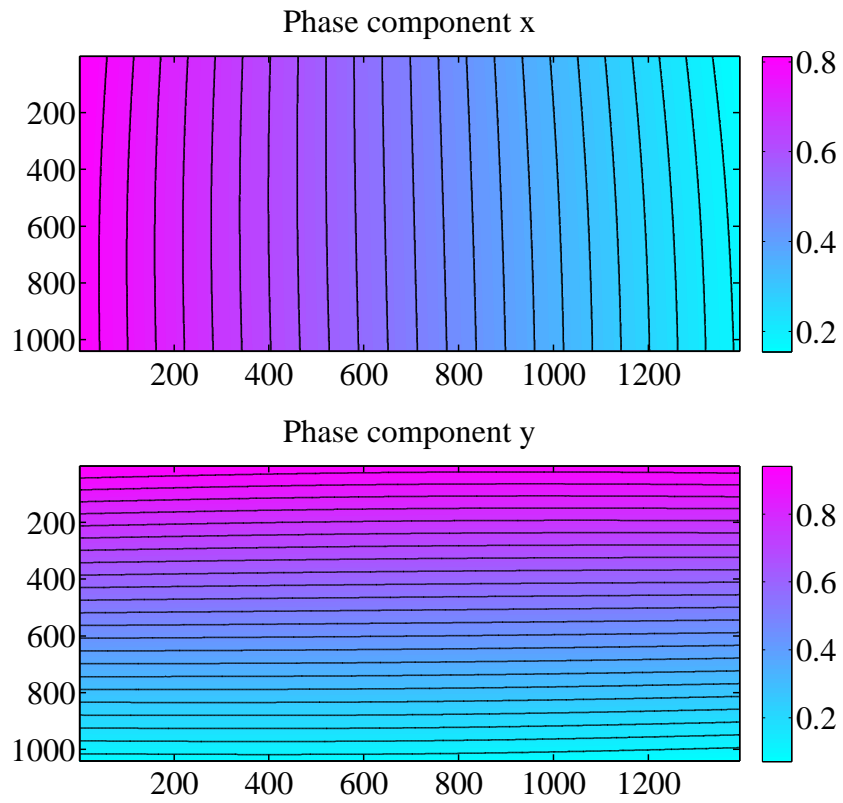
Fig. 2.4: Phase coords φ_x and φ_y components with contour lines as seen by the camera. The values are normalized to $[0; 1]$. In this particular view the camera was rotated by approximately 180 degrees relative to the display.

phase maps φ_x and φ_y we can find correspondences of world coordinates with image coordinates. These can then be used as input for the camera calibration just as before. The actual lookup of the subpixel image coordinates (x_i, y_i) for given phase coordinates (x_p, y_p) is done by the "reverse" bilinear interpolation described in algorithm 1 (see also figure 2.5).

Algorithm 1 Subpixel Phase Lookup

---

1. Find a block of four neighboring pixels $\{p_k\}$ in the phase maps where both $min(\varphi_x(p_k)) \leq x_p < max(\varphi_x(p_k))$ and $min(\varphi_y(p_k)) \leq y_p < max(\varphi_y(p_k))$.

2. Fit a plane P_x to the values of φ_x in $\{p_k\}$. Fit a plane P_y to the values of φ_y in $\{p_k\}$. The set $\{p_k\}$ can be augmented by additional neighbors.

3. Intersect P_x with the plane $\varphi_x = x_p$ and P_y with the plane $\varphi_y = y_p$. This gives two lines.

4. Set the phase-component of the lines to zero and calculate the intersection point (x_i, y_i)

With algorithm 1 we can generate "virtual" marks from the phase maps with arbitrary density. As the apparent display brightness changes with the viewing angle, it can happen that some areas of the display appear very dark even when other parts of the image have optimal brightness. Because of quantization effects the accuracy of the phase map suffers if the local dynamic range is close to zero. In that case a High Dynamic Range approach with multiple different exposure times can be employed. The plane fitting in step 2 of algorithm 1 also provides us with the standard deviation of the measured phase values from the fitted plane. Good phase maps are very smooth, so typical values of the standard deviation are around 10^{-6}. If the phase map is noisy, the standard deviation is higher and those marks can be discarded.

2.3.1 Ray offsets

A further improvement can be achieved by modelling the refraction caused by the glass plate that covers the pixels of the display. The protective glass plate refracts the emitted light and causes a shift in the pixels' apparent position. To correct for this effect we use a three-step algorithm. We first calibrate with the point correspondences we found as
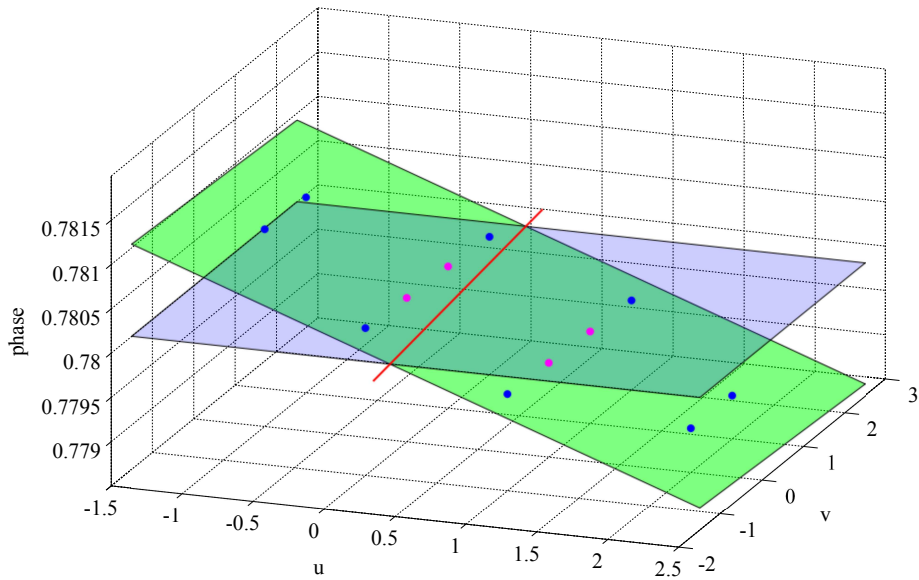
Fig. 2.5: Phase coordinate lookup for one component. The dots are the measured phase values. Magenta indicates the original block of four pixels. The blue dots are additional neighbors used in the plane fit. The green plane is the linear local approximation of the phase (P_x). The blue plane ($\varphi = x$) represents the sought-after phase value. The intersection of the two planes is marked by the red line. The second phase component yields another line (not shown here). The intersection of both lines gives the pixel location of the phase coordinate of interest.

if there were no glass plate. We obtain approximations of the camera poses relative to the display. In the second step we compute the offsets introduced by oblique viewing angles through the glass. The height offset is

$$h = d \cdot \left(1 - \frac{\tan \alpha_2}{\tan \alpha_1}\right) \qquad (2.4)$$

where α_1 and α_2 are related by Snell's law (figure 2.6). Finally we calibrate again with the corrected coordinates.

The thickness of the glass layer and its index of refraction are only approximately known. For our experiments we assumed a refractive index $n = 1.5$, which is typical for glass and glass-like substances. We estimated the thickness of the coating as $d = 1mm$. In the example plot of the height offsets shown in figure 2.7 the difference in the height offset between a perpendicular view in the center and an oblique view at the edges is only 0.04mm, while the pixel size is 0.272mm. The lateral offsets introduced are thus below 0.1 pixels.
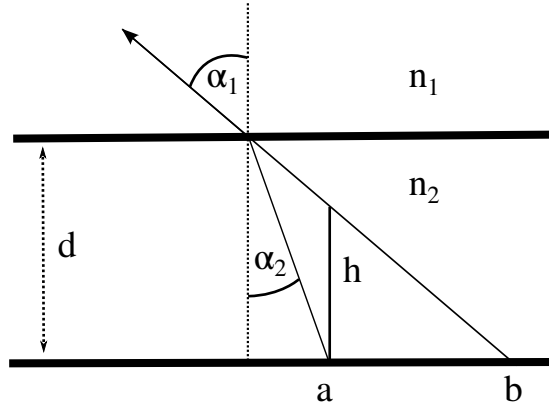

Fig. 2.6: The glass plate refracts the ray coming from pixel (a) so that its apparent position is (b). Adding the offset h corrects the error.

3 Experimental evaluation

While the use of active calibration certainly has practical advantages, the most important factor is the calibration quality that can be achieved in comparison to alternative methods. Other authors like Sagawa et al. [6], Tardif et al. [7], Grossberg and Nayar
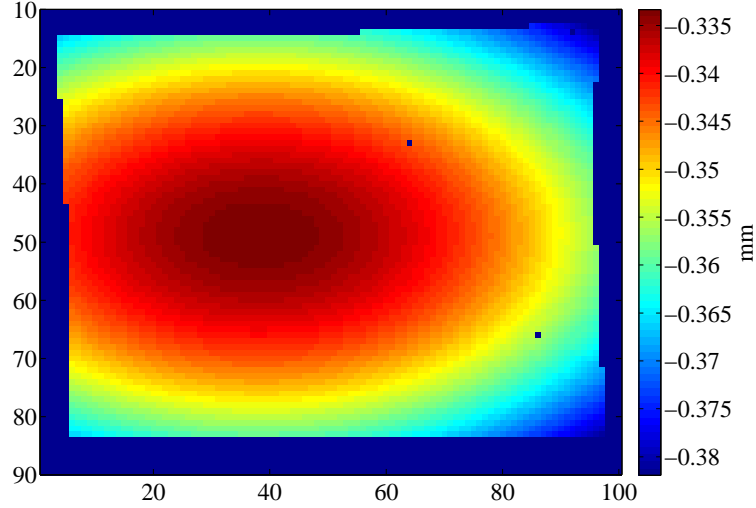
Fig. 2.7: Offsets introduced by the glass plate of display 2 for the high-resolution camera with an 8.5mm lens. Arbitrary units in x and y direction.

[8] have used similar approaches, but only for extreme wide-angle and catadioptric cameras, and did not provide systematic accuracy evaluations. As they used different cameras and different encoding schemes, we cannot compare our results to theirs. Instead, the proposed calibration method was evaluated in several other ways. We used simulated images where the ground truth camera parameters are known. We tested various real-world setups with different combinations of cameras, lenses and displays. In each test, the calibration with an active target is compared to a calibration using a checkerboard pattern. Finally, we compared the stereo triangulation accuracy of the two calibration methods.

The standard targets in our lab are checkerboard targets with isolated squares. Their advantage is that the unoccupied space in between the markers can be used to perform projector calibrations. On a regular dense checkerboard the projected marks are much harder to detect. The targets have been examined with a coordinate-measuring machine, so the mark locations are known with very high precision. The corners of the checkers are localized in the camera image either with the Saddle Point method (SP)

[31] or with the Line Intersection technique (LI) [32]. We use the standard SP implementation provided by OpenCV and a self-implemented LI variant. Since we use a Phase Shift coding for the active target, our proposed method is abbreviated as PS in the subsequent sections.

All calibrations use the camera model and optimization algorithm proposed by Zhang, as implemented in the OpenCV library. The error metric used is the undistorted RMS reprojection error between the observed and undistorted mark coordinates and the projected mark coordinates in the images. This is a standard metric and should be comparable with results presented in different publications. The projected mark locations $[\widetilde{u}_i, \widetilde{v}_i]$ are computed from the known world coordinates of the mark using equation 2.1. The tilde indicates that these coordinates are calculated by pure perspective projection without image distortion. Another way to obtain these "ideal" coordinates is to correct the distortion in the observed coordinates. The undistorted mark coordinates $[\hat{u}_i, \hat{v}_i]$ are denoted with a hat. They are computed from the observed distorted mark positions by inverting equations 2.2 and 2.3. The reprojection error is then

$$e = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{u}_i - \widetilde{u}_i)^2 + (\hat{v}_i - \widetilde{v}_i)^2} \qquad (3.1)$$

Here N is the total number of points used for the calibration. An individual calibration mark can occur several times in the images from different camera poses.

3.1   Simulated images

Calibration images in five camera poses were rendered with a simulated resolution of 800x600 pixels. No noise was added to the images. We took care to minimize aliasing artifacts by avoiding poses aligned with the camera axes. The five poses used can be seen in figure 3.1. Table 1 shows the resulting internal camera parameters and reprojection error for all three calibration methods. While LI gives a slightly higher reprojection error than SP, the recovered parameters are closer to the truth. The parameters obtained with the proposed method are much closer to the real values than any of the other two methods. The remaining error is probably due to quantization noise.
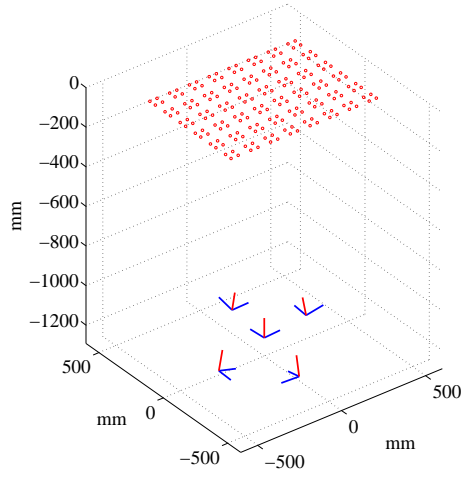
14

Fig. 3.1: Simulated Camera Poses. The red circles are marks, the red lines the camera z-axes. The blue line pairs indicate the camera image planes.

	ideal	PS	LI	SP
$f\ [mm]$	12	12.0004	11.9973	12.0059
$k_1\ [\frac{10^{-4}}{mm^2}]$	0	-1.03	-4.67	-15.06
$k_2\ [\frac{10^{-4}}{mm^4}]$	0	4.74	19.47	373.38
$k_3\ [\frac{10^{-4}}{mm^6}]$	0	-6.87	-115.42	-2205.61
$p_1\ [\frac{10^{-6}}{mm^2}]$	0	0.442	-14.05	-5.46
$p_2\ [\frac{10^{-6}}{mm^2}]$	0	-6.54	-6.81	147.51
$u_0\ [px]$	399.5	399.491	399.421	399.948
$v_0\ [px]$	299.5	299.508	299.514	299.342
$RMSE\ [px]$	0	0.01424	0.07156	0.06434

Tab. 1: Calibration results for simulated images. The parameters recovered with the proposed PS method are closest to the ground truth. While the reprojection error of the SP method is better than for the LI method, the actual parameter values are worse.

3.2 Real images

In our real-world experiments we used combinations of different displays, cameras and lenses. The details of the displays and cameras are collected in table 2. The lenses had focal lengths of 12.5mm, 8.5mm, 6.0mm and 4.8mm. We used an f-number of 8 and an object distance of 0.5m in our image acquisition.

Name	Type	Resolution	Pixel Size
D1	ScenicView A24W	1920x1200	0.270mm
D2	SyncMaster 2433LW	1920x1080	0.272mm

(a) Displays

Name	Type	Resolution	Pixel Size
HR	Basler Scout 1390m	1392x1040	4.65µm
LR	Basler A312fc	780x580	8.3µm

(b) Cameras

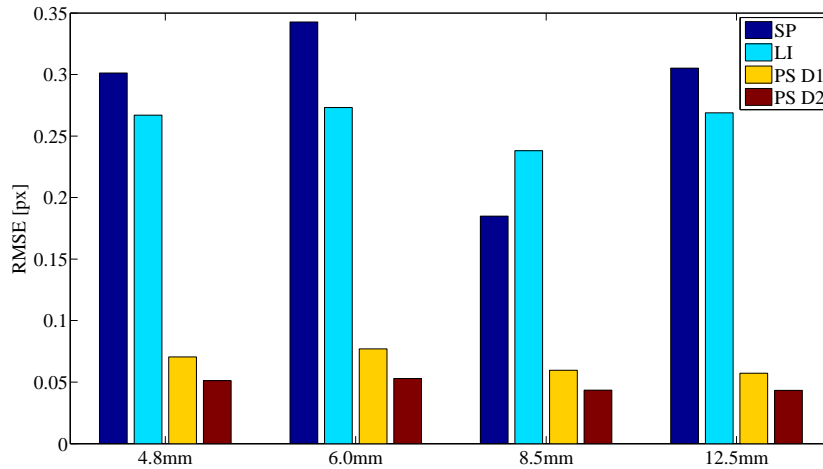Tab. 2: Hardware used for experiments

Fig. 3.2: Results for the low-resolution camera. The LI method is slightly better than SP. The errors of the proposed method are much lower.

The results are shown in figures 3.2 and 3.3. The main conclusion is that PS is the best method for the low-resolution camera by a large margin and for the high-resolution camera by a smaller margin. Compared to the LI method, the reprojection error is between a factor four and five better in the low-resolution case and up to a factor two better in the high-resolution case. Please note that perfect reproducibility of the poses
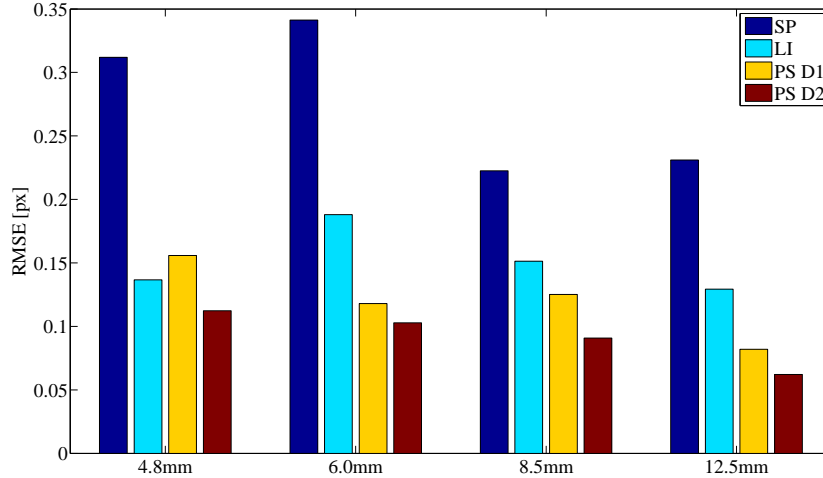
Fig. 3.3: Results for the high-resolution camera. The errors of the proposed method are slightly lower than the LI method. The SP method performs worst.

used for the calibrations can only be guaranteed with a robotic setup. Lacking that, the poses in our experiments were arranged manually. Hence, they are not perfectly identical in the different experiments. We therefore regard differences of a hundredth of a pixel or lower as not significant. Still, display 2 consistently gave slightly better results than display 1. The existance of display-specific systematic errors, especially non-perfect planarity, is a topic for further study.

The SP method performs worst. This comes as a surprise, since it is the standard method for users of OpenCV. The residuals of the SP calibration show a systematic error as the corner positions are mostly shifted towards the center of the calibration squares, compared to the LI positions (figure 3.4). Fiala and Shu Fiala and Shu [5] have identified this effect as related to lighting; it seems to arise for defocusing as well. In a true checkerboard the shift should cancel out between the two touching corners. The high-quality targets at our lab have isolated squares, so we use only the LI method in the following experiments. In table 1, LI on a 'sparse' checkerboard was compared to SP on a traditional 'dense' checkerboard pattern. It performed comparably with respect to the reprojection error and better with respect to the ground truth camera parameters, so LI can be used as a reference calibration method.

The difference in the PS residual error between the HR and LR cameras is approx-

17

Fig. 3.4: Typical corner detection images, enlarged by a factor 20. The green cross marks the LI corner position, the red cross the SP corner position. Left: High resolution image. Right: Low resolution image.

imately in line with the difference in pixel size (table 3 and 4). This is consistent with

a constant size focus spot on the sensor that depends on the employed lens.

D1	4.8mm	6mm	8.5mm	12.5mm
LR RMS [px]	0.0704	0.0770	0.0596	0.0572
HR RMS [px]	0.1557	0.1180	0.1250	0.0819
LR RMS [mm]	0.5845	0.6392	0.4948	0.4754
HR RMS [mm]	0.7242	0.5488	0.5815	0.3812

Tab. 3: The reprojection error for different lenses with display 1. Expressed in micrometers the values are similar between the low resolution and the high resolution cameras.

D2	4.8mm	6mm	8.5mm	12.5mm
LR RMS [px]	0.0512	0.0528	0.0435	0.0432
HR RMS [px]	0.1123	0.1027	0.0908	0.0620
LR RMS [mm]	0.4250	0.4388	0.3611	0.3585
HR RMS [mm]	0.5223	0.4777	0.4224	0.2886

Tab. 4: The reprojection error for different lenses with display 2. Expressed in micrometers the values are similar between the low resolution and the high resolution cameras.

3.2.1 Poses

The choice of camera poses is of course a major factor in the quality of a calibration. We tried to use comparable poses, that is with similar angles to the target. They are shown in figure 3.5 and figure 3.5b. Table 5 shows that the reprojection error barely changes, whether 3, 5 or all 7 poses are used for the calibration. However, the resulting internal camera parameters do change. There is no ground truth to compare against,

18

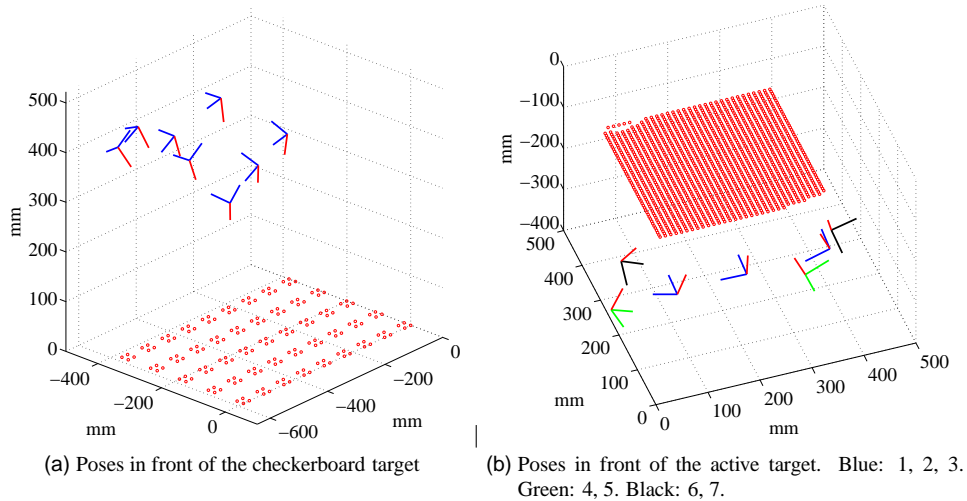but it seems reasonable to have higher confidence in a calibration result if it is based on more poses.

(a) Poses in front of the checkerboard target

(b) Poses in front of the active target. Blue: 1, 2, 3. Green: 4, 5. Black: 6, 7.

Fig. 3.5: Example camera poses.

poses	RMSE [px]	u_0 [px]	v_0 [px]	f [mm]
7	0.1481	702.9860	526.0807	6.1911
5	0.1474	701.9275	526.4855	6.1961
3	0.1469	701.3100	526.6469	6.1979

Tab. 5: RMS errors for different number of poses and some of the resulting internal parameters. High-resolution camera with 6.0 mm lens.

3.2.2 Mark Density

The density of marks generated with the PS approach has little influence (table 6). However, as already stated in the previous section, when in doubt there is no reason not to use as many marks as possible. Also, for the calibration of a fully generic non-parametric camera model, dense correspondences are important and can be easily generated with PS.

3.2.3 Defocusing

For the classic checkerboard targets, high feature localization accuracy depends on a sharp image. This can be a problem, for example when depth-of-field is limited. PS

HR	8.5mm	D2
number of marks	poses	RMSE [px]
23238	4	0.09105
2583	4	0.09085
466	4	0.09449
224	4	0.08456
LR	4.8mm	D1
number of marks	poses	RMSE [px]
10876	4	0.07173
1739	4	0.07042
435	4	0.06890
106	4	0.06187

Tab. 6: Influence of mark density. For comparison, a typical view of a checkerboard yields around 100 marks.

results are robust against defocusing (table 7). The measured phase at a given pixel does not change when the image is blurred, only the contrast is reduced. In fact, PS even profits from a moderate amount of defocusing as aliasing between the display pixel grid and the camera pixel grid is reduced. Therefore it is possible to move the camera close to the display during calibration so that the entire field of view is covered.

f-stop	RMSE LI [px]	RMSE PS [px]
5.6	0.1834	0.1312
11	0.1367	0.1400

Tab. 7: Robustness against defocusing. High-resolution camera with 4.8mm lens.

3.2.4 Glass Plate Offsets

As mentioned in section 2.3.1, the protective glass plate in front of the display pixels introduces a shift in the apparent mark coordinates. As can be seen in table 8, modelling this refraction does indeed result in an improvement of the reprojection error. However, the effect is relatively minor. It is on the order of a few thousands of a pixel only, while the mark offsets are up to 0.1 pixels in the lateral direction. This is because the shifts can be partially compensated by the camera distortion parameters.

	4.8mm	6.0mm	8.5mm	12.5mm
D1+LR	0.0755	0.0825	0.0692	0.0598
D1+LR+glass	0.0704	0.0770	0.0596	0.0572
D1+HR	0.1592	0.1168	0.1286	0.0846
D1+HR+glass	0.1557	0.1180	0.1250	0.0819
D2+LR	0.0523	0.0548	0.0449	0.0435
D2+LR+glass	0.0512	0.0528	0.0435	0.0432
D2+HR	0.1161	0.1048	0.0925	0.0625
D2+HR+glass	0.1123	0.1027	0.0908	0.0620

Tab. 8: Improvements by modelling the ray offsets introduced by the glass cover of the display.

3.3 Repeatability

Another important test for the proposed calibration method is repeatability. We performed ten external calibrations of a pre-calibrated Basler A312fc camera with a 12.5mm lens. Purely external calibration has the advantage that a single view of the calibration target suffices, so no parts of the setup had to be moved. A classic feature-based target and an active target were used. The resulting poses are plotted in figure 3.6. The standard deviations of the translational parameters are shown in table 9. The mean offsets from the mean position were 5.3μm for the active calibration target and 22.1μm for the classic target. The absolute distances to the respective calibration targets were practically equal at 258mm and 256mm.

	σ_x [mm]	σ_y [mm]	σ_z [mm]
classic	0.0146	0.0173	0.0138
active	0.0014	0.0023	0.0060

Tab. 9: Standard deviations of the translation parameters for a classic and an active target.

3.4 Stereo calibration

As seen in table 1 and also noted by [25], a lower reprojection error does not automatically imply a more correct calibration. Therefore, we tested the proposed method further. We performed a stereo calibration and subsequently triangulated the positions of the calibration marks. We then compared the known positions of the marks to the triangulation results. Since the display was positioned closer to the camera than the

21

(a) Poses recovered from the classic target   (b) Poses recovered from the classic target
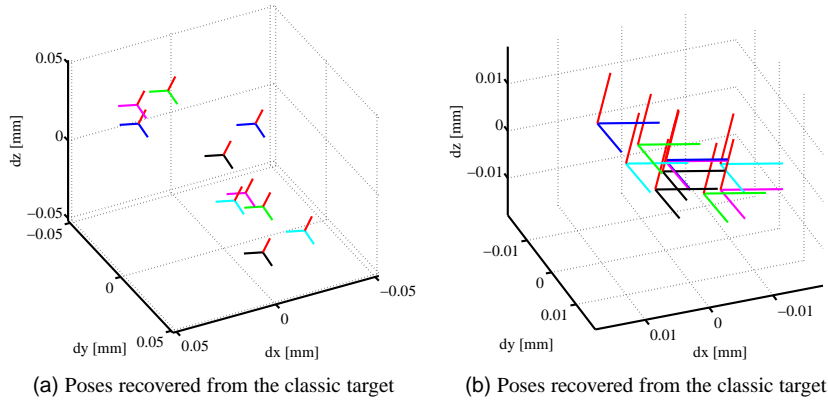
Fig. 3.6: Repeatability of external calibration. The camera indicators have the same size, but the scales are different. The red lines indicate the camera z-axis, the different colors for the x and y-axes are for better visual differentiation.

checkerboard target, we also normalized the errors. Our stereo rig consisted of two Basler A312fc cameras (the low resolution model in the previous experiments) with 8.5mm lenses and a baseline of approximately 150mm. Four poses were used. The checkerboard target yielded 169 marks visible in both cameras, the active target yielded 1219 marks to triangulate. As can be seen in table 10, the errors are much lower for the proposed Phase Shift calibration. The accuracy is improved approximately by a factor of five, which is consistent with the results of the monocular calibration (figure 3.2).

	error [mm]		normalized error [mm/m]	
	mean	sigma	mean	sigma
PS	0.0299	0.0175	0.1153	0.0643
LI	0.3028	0.2286	0.5528	0.3947

Tab. 10: Stereo Triangulation Results. The error for the proposed PS technique is approximately one fifth of the error resulting from the LI method.

4 Conclusions

There are many variables that influence the quality of a calibration, from the choice of camera poses to the tuning of algorithm parameters. Additionally, errors are often compounded, so the source of problems is not always obvious. The calibration method with active targets has several advantages. It is fully automatic, no user interaction

to identify marks is necessary, and no labeling errors can occur. Digital displays are highly accurate targets, so there is no need for costly target validation. The only input parameters are the display resolution and the pixel size. The method is robust against defocusing and easy to set up. Lastly and most importantly, the achievable accuracy is very high. One possible disadvantage is that the calibration requires multiple images per pose and cannot be performed with a hand-held camera. However, we provided a thorough evaluation and found a marked increase of the calibration quality. We are therefore of the opinion that the additional accuracy over a classic feature-based calibration is worth the effort for tasks like precise 3D reconstruction.

References

[1] F. Remondino and C. Fraser. Digital camera calibration methods: considerations and comparisons. In Isprs, editor, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences,*, volume XXXVI, Dresden, Germany, 2006.

[2] E. Hemayed. A survey of camera Self-Calibration. *Advanced Video and Signal Based Surveillance, IEEE Conference on*, page 351, 2003.

[3] J. Heikkilä. Geometric camera calibration using circular control points. *Pattern Analysis and Machine Intelligence (PAMI), IEEE Transactions on*, 22:1066–1077, October 2000. ISSN 0162-8828.

[4] ActiveCalibCode. http://www5.informatik.uni-erlangen.de/en/research/software/.

[5] M. Fiala and C. Shu. Fully automatic camera calibration using self-identifying calibration targets. Technical report, NRC Institute for Information Technology, National Research Council Canada, 2010.

[6] R. Sagawa, M. Takatsuji, T. Echigo, and Y. Yagi. Calibration of lens distortion by structured-light scanning. In *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*, page 832–837, 2005. ISBN 0780389123.

[7] J. Tardif, P. Sturm, M. Trudeau, and S. Roy. Calibration of cameras with radially symmetric distortion. *Pattern Analysis and Machine Intelligence (PAMI), IEEE Transactions on*, 31 (9):1552–1566, 2009. ISSN 0162-8828.

[8] M. D Grossberg and S. K Nayar. The raxel imaging model and ray-based calibration. *International Journal of Computer Vision*, 61(2):119–137, 2005. ISSN 0920-5691.

[9] R. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. In Lawrence B Wolff, Steven A Shafer, and Glenn Healey, editors, *Radiometry*, page 221–244. Jones and Bartlett Publishers, Inc., , USA, 1992. ISBN 0-86720-294-7.

[10] J. Heikkilä and O. Silven. A four-step camera calibration procedure with implicit image correction. In *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, page 1106–, Washington, DC, USA, 1997. IEEE Computer Society. ISBN 0-8186-7822-4.

[11] Z. Zhang. A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence (PAMI), IEEE Transactions on*, page 22, 2000.

[12] G. Bradski and A. Kaehler. *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly Media, 1st edition, September 2008. ISBN 0596516134.

[13] J. Y. Bouguet. *Camera calibration toolbox for Matlab*. 2008. URL `http://www.vision.caltech.edu/bouguetj/calib_doc/`.

[14] R. Willson and S. Shafer. What is the center of the image? Technical report, Carnegie Mellon University, Pittsburgh, PA, USA, 1993.

[15] R. Hartley and S. Kang. Parameter-free radial distortion correction with center of distortion estimation. *Pattern Analysis and Machine Intelligence (PAMI), IEEE Transactions on*, page 1309–1321, 2007. ISSN 0162-8828.

[16] M. Grossberg and S. Nayar. A general imaging model and a method for finding its parameters. *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, 2:108, 2001.

[17] Srikumar Ramalingam, Peter Sturm, and Suresh K Lodha. Theory and experiments towards complete generic calibration. Technical report, INRIA – CNRS, 2005.

[18] A. Dunne, J. Mallon, and P. Whelan. A comparison of new generic camera calibration with the standard parametric approach. In *Proceedings of the IAPR Conference on Machine Vision Applications, Tokyo, Japan*, volume 1, page 114–117, 2007.

[19] J. Mallon and P. Whelan. Which pattern? biasing aspects of planar calibration patterns and detection methods. *Pattern Recognition Letters*, 28:921–930, June 2007. ISSN 0167-8655.

[20] M. R. Shortis, T. A. Clarke, and T. Short. Comparison of some techniques for the subpixel location of discrete target images. In S. F. El-Hakim, editor, *Proc. SPIE, Videometrics III*, volume 2350, page 239–250, October 1994.

[21] P. Mohr and P. Brand. Accuracy in image measure. *Proc. SPIE, Videometrics III*, 2350: 218–228, 1994.

[22] R. G. White and R. A. Schowengerdt. Effect of point-spread functions on precision edge measurement. *Journal of the Optical Society of America A*, 11(10):2593–2603, 1994. ISSN 1084-7529.

[23] D. Chen and G. Zhang. A new sub-pixel detector for x-corners in camera calibration targets. In *The 13th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, page 97–100, 2005.

[24] D. Douxchamps and K. Chihara. High-accuracy and robust localization of large control markers for geometric camera calibration. *Pattern Analysis and Machine Intelligence (PAMI), IEEE Transactions on*, 31(2):376–383, 2008. ISSN 0162-8828.

[25] A. Albarelli, E. Rodolà, and A. Torsello. Robust camera calibration using inaccurate targets. *Pattern Analysis and Machine Intelligence (PAMI), IEEE Transactions on*, 31(2): 376–383, 2009.

[26] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. Bundle adjustment—a modern synthesis. *Vision algorithms: theory and practice*, page 153–177, 2000.

[27] J. Salvi, S. Fernandez, T. Pribanic, and X. Llado. A state of the art in structured light patterns for surface profilometry. *Pattern Recognition*, In Press, Corrected Proof:–, 2010. ISSN 0031-3203.

[28] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, volume 1, 2003. ISBN 0769519008.

[29] G. Sansoni, M. Carocci, and R. Rodella. Three-Dimensional vision based on a combination of Gray-Code and Phase-Shift light projection: Analysis and compensation of the systematic errors. *Applied Optics*, 38(31):6565–6573, 1999.

[30] K. Liu, Y. Wang, D. L Lau, Q. Hao, and L. G Hassebrook. Gamma model and its analysis for phase measuring profilometry. *Journal of the Optical Society of America A*, 27(3): 553–562, 2010. ISSN 1084-7529.

[31] L. Lucchese and S. K Mitra. Using saddle points for subpixel feature detection in camera calibration targets. In *Asia-Pacific Conference on Circuits and Systems (APCCAS)*, volume 2, page 191–195, 2002. ISBN 0780376900.

[32] C. Stock, U. Mühlmann, M. K Chandraker, and A. Pinz. Subpixel corner detection for tracking applications using cmos camera technology. In *26th Workshop of the AAPR/ÖAGM*, volume 160, page 191–199, 2002.

Christoph Schmalz received his diploma in Physics from the University of Erlangen-Nuremberg, Germany, in 2005. Currently he is working on a PhD thesis about Single Shot Structured Light scanning in a joint project of Siemens AG and the Pattern Recognition Lab in Erlangen.

Dr. Frank Forster received his PhD in Computer Science from the Technical University of Munich, Germany, in 2005. He is the program manager for Machine Vision technologies in at Siemens CT HW2 in Munich. His research interests include camera calibration, range imaging and 2D and 3D image processing.

Dr. Elli Angelopoulou received her PhD in Computer Science from the Johns Hopkins University in 1997. She did her postdoc at the GRASP Laboratory at the University of Pennsylvania. Her research is focused on multispectral imaging, reflectance analysis, image forensics and shape reconstruction. She has served on the program committees of ICCV, CVPR and ECCV and is an associate editor of MVA and JISR. She is a member of the OSA and the IEEE Computer Society Technical Committee on Pattern Analysis and Machine Intelligence.