Simultaneous self-calibration of a projector and a camera using structured light

Shuntaro Yamazaki[†] Masaaki Mochimaru[†]
[†]National Institute of Advanced Industrial Science and Technology
2-3-26 Aomi, Koto-ku, Tokyo 135-0064, JAPAN
{shun-yamazaki,m-mochimaru}@aist.go.jp

Takeo Kanade[†‡]
[‡]Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213, USA
tk@cs.cmu.edu

Abstract

We propose a method for geometric calibration of an active vision system, composed of a projector and a camera, using structured light projection. Unlike existing methods of self-calibration for projector-camera systems, our method estimates the intrinsic parameters of both the projector and the camera as well as extrinsic parameters except a global scale without any calibration apparatus such as a checker-pattern board. Our method is based on the decomposition of a radial fundamental matrix into intrinsic and extrinsic parameters. Dense and accurate correspondences are obtained utilizing structured light patterns consisting of Gray code and phase-shifting sinusoidal code. To alleviate the sensitivity issue in estimating and decomposing the radial fundamental matrix, we propose an optimization approach that guarantees the possible solution using a prior for the principal points. We demonstrate the stability of our method using several examples and evaluate the system quantitatively and qualitatively.

1. Introduction

Active vision setups composed of a projector and a camera, have been widely used for a variety of practical applications as the acquisition setup can be easily adjusted depending on the shape of objects and the acquisition environment. While the system components can be flexibly reconfigured, calibrating the geometric parameters of both projector and camera is tedious and time-consuming. This is a practical drawback of projector-camera systems, particularly for mobile applications where the components are to be adjusted frequently in position, orientation, and field-of-view of both projector and camera.

A number of methods have been proposed for camera calibration using either a calibration object [24] or correspondences between multiple view images [15, 21, 4]. There also exists nifty software that conducts camera calibration with little effort [5, 19]. A projector has, in theory, the same optical structure as a camera, and therefore can be calibrated using the existing methods and software for camera calibration. In fact, there are some techniques of projector calibration based on the algorithm for camera calibration [7, 1].

In practice, however, the projector calibration algorithms based on camera calibration cannot achieve as accurate results as camera calibration. First, these methods typically require a pre-calibrated camera to calibrate a projector. From a theoretical point of view, the two-step calibration leads to error accumulation. Moreover, the two-step process is not as convenient as camera calibration methods which can be conducted by using a single calibration object and/or a single algorithm. Second, they require an object whose shape is known, such as a checker-pattern board or a textured box. Thus, in 3D shape acquisition, we have to prepare a calibration object of nearly the same size as the the object which we want to capture.

In this paper, we propose a method for calibrating the geometric parameters of both projector and camera simultaneously using structured light. We estimate the focal lengths, principal points, and lens distortions of projector and camera, as well as their relative pose, up to a global scale. Our calibration algorithm is based on the decomposition of the fundamental matrix. Several researchers point out that the decomposition of fundamental matrix is unstable, and often results in large errors in camera parameters [11, 4, 13]. This is true when the fundamental matrix is estimated from sparse correspondences between image feature points. We show that the decomposition of the fundamental matrix is

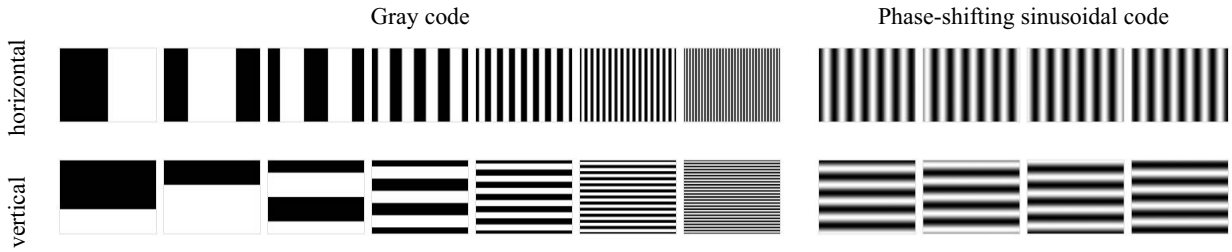


Figure 1. Structured light patterns used for calibration

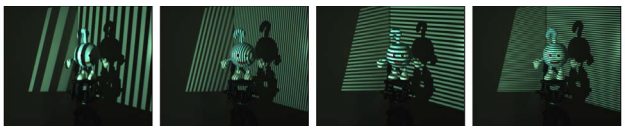


Figure 2. Example of acquired images

quite stable when the correspondences are obtained densely and accurate in active vision systems.

The major contributions of this work are twofold:

- We derive the degenerated correlation between the corresponding points in projector and camera images with lens distortion. This is the generalized form of the radial fundamental matrix proposed by Barreto and Daniilidis [2]. We show that the radial fundamental matrix can be factorized into a fundamental matrix and two lens distortion matrices, up to a homography which can be determined by making a common assumption that the centers of lens distortion coincides with the principal points.
- We demonstrate that the self-calibration of two-view geometry is of practical use when the correspondences are estimated densely and accurately using structured light. The basic algorithm of our calibration is borrowed from prior work on camera self-calibration which has been regarded as rather impractical due to numerical instability.

2. Related work

2.1. Structure light projection

Our calibration method uses structured light to obtain the dense point correspondences between the images of a projector and a camera. Each pixel in the projected patterns encodes the corresponding projector image coordinates. The pattern on some object surface is observed by a camera and decoded to obtain the correspondence. The strategies of structured light patterns are comprehensively surveyed and systematized by Salvi et al. [17] In this section we review some of the related codification schemes.

Time-multiplexing code is a set of different patterns successively projected onto the measuring surface. Each projector pixel can encode its projector coordinate uniquely, which enables robust decoding. Among various patterns in this category, Gray code is the most commonly used for its efficiency and robustness [18, 9]. Since the binary code can not achieve sub-pixel accuracy, interpolation using observed image intensity [18] and combination of phase shift patterns [3] are proposed to achieve sub-pixel accuracy in the correspondence.

On the other hand, spatial code encodes the projector image coordinates into a single pattern. This approach enables to establish the correspondence by only a single image acquisition. A pseudo-random pattern is used to distinguish the coordinate within a pattern as uniquely as possible. In general, however, the ambiguity of a correspondence is left either by design of the pattern, or because of a discontinuity on the object surface. The ambiguity is resolved by non-linear optimization methods, such as dynamic programming [23] or belief propagation [22], and/or some additional assumptions such as coplanarity in the projected patterns [10].

Our calibration relies on the fundamental matrix estimated from the correspondences between projector and camera images. Thus, our method is independent of the choice of structured light, and flexible enough to choose the patterns suitable for each application. In this paper, we focus on the accuracy of calibration, and propose an approach that uses Gray code and phase-shifting sinusoidal patterns as shown in Figure 1.

2.2. Projector self-calibration

Projector-camera systems have been used for a variety of applications, including active vision, large format display, augmented reality, and distortion-correcting projector. Each of these systems has its own algorithm of calibration. Furukawa and Kawasaki proposed a technique of projector calibration using structured light projection [9]. They also use Gray code patterns to obtain the correspondences, and the intrinsic and extrinsic parameters of the projector are estimated using the epipolar constraints. The calibration relies on the non-linear optimization of an objective function, which requires good initial values of both intrinsic and extrinsic parameters. Okatani and Deguchi address the calibration of projectors using multiple patterns projected onto

1. Adjust a projector and a camera.
2. Acquire the images of structured light projection.
3. Generate dense point correspondences.
4. Estimate fundamental matrix and lens distortion.
5. Estimate focal lengths.
6. Estimate relative pose.
7. Reconstruct a depth map. (optional)

Figure 3. Overview of calibration process

planar surfaces [14]. Raskar and Berdsley propose an algorithm for calibrating the focal length and the principal points of a projector by observing a planar surface using a pre-calibrated camera [16].

These calibration methods assume that the camera is calibrated in advance. In practical scenarios of shape acquisition, camera parameters such as focal length have to be adjusted for each acquisition according to the shape of the object, lighting condition, and acquisition setup. Our method is capable of calibrating all parameters that may be changed by users during the acquisition. We also estimate lens distortion which has been ignored in previous methods.

3. Overview

In this paper we consider a minimal active vision system composed of a projector and a camera. Our goal is to calibrate the projector-camera system for acquiring the depth maps of object surfaces using structured light. The calibration is conducted as follows. After adjusting the field of views of a projector and a camera, we first project a set of structured light patterns onto the object surface, and acquire the images. The images are then decoded and build the correspondence from camera image coordinates to projector image coordinates (Section 4). Then, we estimate the radial fundamental matrix (Section 5.1), which is decomposed into the fundamental matrix and lens distortion parameters (Section 5.2). The fundamental matrix is further decomposed into the focal length of the projector and the camera, and their relative pose is also recovered (Section 5.3). Once the calibration is finished, the depth map can be reconstructed from the correspondences used for calibration or captured in an additional acquisition. The process is summarized in Figure 3.

4. Accurate correspondence using structured light

The correspondences between the images of a projector and a camera are obtained by projecting and decoding structured light patterns encoded using Gray code and phase-shifting sinusoidal patterns [3]. In order to encode the two-dimensional coordinates uniquely, we project the patterns in horizontal and vertical directions. The patterns described in this section are designed to achieve the best accuracy and robustness in correspondence. When the number of image acquisitions is limited, we may decrease the number of patterns and solve the ambiguity in correspondence assuming the smoothness of the object surface and incorporating spatial coding schemes [10].

Gray code allows to encode N codewords into $\lceil \log N \rceil$ pattens in black and white. In order to detect the binary patterns robustly, we also project the complementary patterns with black and white inverted, and detect the change of sign in their differences. The pixels whose difference is less than a certain threshold are discarded and unused in the following calibration.

The correspondences obtained by the binary code have quantization errors (See plot (a) in Figure 4). To improve the correspondences to sub-pixel accuracy, we also use phase-shifting sinusoidal patterns, each of which is represented as 8-bit gray-value images. When we project a sinusoidal pattern in phase δ, the intensity I of the acquired image is represented as

$$I = A\sin(\theta + \delta) + m \quad (1)$$
$$= (\sin\delta, \cos\delta, 1) \cdot (A\cos\theta, A\sin\theta, m)^T. \quad (2)$$

The amplitude of the observed sinusoidal pattern A, phase difference θ, and DC component m are the unknowns, which can be determined as a linear solution with at least 3 Is for different δs. In Equation (2), we assume that the intensity of outgoing light at a surface point is linearly dependent on the incoming light. However, the response of the projector and camera device is not necessarily linear, which leads to systematic drift of correspondences as shown on plot (b) in Figure 4. Representing the the non-linear response by exponential functions with indices α and β respectively, we have

$$I = \left(\tilde{A}\sin^{\alpha}(\tilde{\theta} + \delta) + \tilde{m}\right)^{\beta}. \quad (3)$$

The unknowns in Equation (3) are the parameter of the sinusoidal curve $\tilde{A}, \tilde{\theta}$, and $\tilde{m}$ as well as the indices α and β. Since α and β are considered as constant for each image, we have in total $3n + 2$ parameters with the number of valid pixels n. We can solve the unknowns by two step algorithm. We first fix $\alpha = \beta = 1$ and determine $\tilde{A}, \tilde{\theta}$, and $\tilde{m}$ in a linear solution, then fit all parameters by the Levenberg-Marquardt method using the linear solution as initial estimate. We have tried various initial values for α and β within 0.5 and 2.0, and do not find any significant difference in convergence. Figure 4(c) shows the correspondence obtained considering the non-linear response functions.

Number of patterns: When the resolution of the projector image is $M \times N$, the number of required Gray code pat-

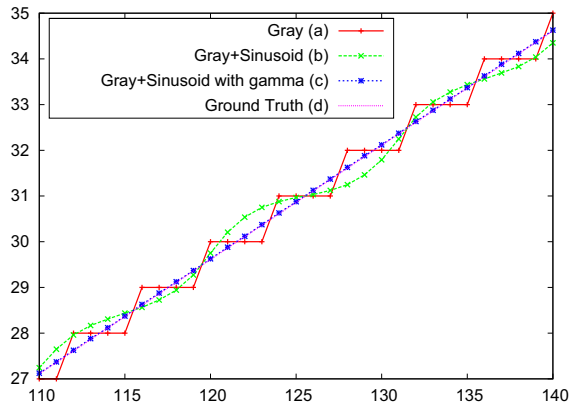


Figure 4. Comparison of point correspondences reconstructed from simulated structured light images. The horizontal and vertical axes indicate the image coodinates. (a) Only Gray code. (b) Gray code and sinusoidal code modulated with $\beta = 2.0$. (c) Gray code and response-corrected sinusoidal code. (d) Ground truth.

terns is $\lceil \log M \rceil + \lceil \log N \rceil$. We project 2 sets of complementary patterns. We also use at least 4 sinusoidal patterns for each of the horizontal and vertical directions. Since the phase-shifting code can determine the coordinate uniquely within a period, the results of the decoding can be redundant when the period is more than one pixel size. That is, for instance, when the period of the sinusoidal pattern is 2^n, we can omit projecting n least significant bits in Gray code and still reconstruct the coordinate uniquely. In our system, we use the redundancy to correct the error in Gray code instead of reducing the number of patterns. When the difference between the results from Gray code and sinusoidal code is within $2^{(n-2)}$, the Gray code is corrected using the sinusoidal patterns. If the difference is more than the value, the pixels is regarded as unreliable and discarded from the following calibration.

5. Simultaneous self-calibration of projector and camera

We assume the projection of the projector and camera follows the pinhole model. We also assume that the aspect ratio is 1, and skew is 0. That is, the intrinsic parameters can be represented by 1 DOF of focal length f and 2 DOFs of principal point $\boldsymbol{p} = (p, q)$ in a matrix form as follows:

$$\boldsymbol{K}_{f,\boldsymbol{p}} = \begin{pmatrix} f & 0 & p \\ 0 & f & q \\ 0 & 0 & 1 \end{pmatrix}. \tag{4}$$

5.1. Epipolar constraint with lens distortion

Suppose a projector image coordinate (u', v') corresponds to a camera image coordinate (x', y'), then the epipolar constraint is, at least in theory, represented by a

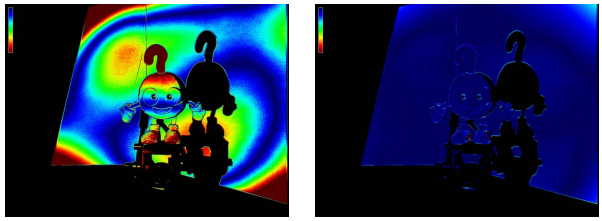


Figure 5. Reprojection error of fundamental matrix. (left) 3×3 Fundamental Matrix $\boldsymbol{F}$. (right) 4×4 Radial Fundamental Matrix $\boldsymbol{R}$. The distance to corresponding epipolar lines in camera image coordinate is encoded in color. Blue indicates no error, and red pixels have more than 1 pixel error.

3×3 fundamental matrix $\boldsymbol{F}$ as follows:

$$(u', v', 1)\boldsymbol{F}(x', y', 1)^T = 0. \tag{5}$$

In practice, Equation (5) does not hold because the lenses of projector and camera are radially distorted, and the image coordinates are non-linearly transformed (See Figure 5 left). Suppose the center of the projector and camera distortion is defined respectively as (a_p, b_p) and (a_c, b_c) in the image coordinates, then the lens distortion can be represented by a division model [8, 20] as

$$(u', v') = \frac{1}{1 + d_p|\boldsymbol{r}_p|^2}\boldsymbol{r}_p + (a_p, b_p) \quad \text{and} \tag{6}$$

$$(x', y') = \frac{1}{1 + d_c|\boldsymbol{r}_c|^2}\boldsymbol{r}_c + (a_c, b_c), \tag{7}$$

where $\boldsymbol{r}_p = (u, v) - (a_p, b_p)$, $\boldsymbol{r}_c = (x, y) - (a_c, b_c)$, and d_p and d_c are respectively the distortion parameters of projector and camera. By representing the projector image coordinate (u, v) in a lifted coordinate as $(u^2 + v^2, u, v)$, the lens distortion is represented in a linear form as

$$(u', v', 1)^T \simeq \boldsymbol{D}_{d_p,a_p,b_p}(u^2 + v^2, u, v, 1)^T \tag{8}$$

where

$$\boldsymbol{D}_{d,a,b} = \begin{pmatrix} da & 1 - 2da^2 & -2dab & da(a^2 + b^2) \\ db & -2da & 1 - 2db^2 & db(a^2 + b^2) \\ d & -2da & -2db & 1 + d(a^2 + b^2) \end{pmatrix}. \tag{9}$$

The symbol $\simeq$ represents equality up to scale. The same equation holds between (x, y) and (x', y') for camera image coordinates.

From Equation (5) and (8), we have

$$(u^2 + v^2, u, v, 1)\boldsymbol{R}(x^2 + y^2, x, y, 1)^T = 0 \tag{10}$$

where

$$\boldsymbol{R} = \boldsymbol{D}^T_{d_p,a_p,b_p}\boldsymbol{F}\boldsymbol{D}_{d_c,a_c,b_c}. \tag{11}$$

The 4×4 matrix $\boldsymbol{R}$ is the radial fundamental matrix which describes the epipolar constraints between the images with lens distortion. The elements in $\boldsymbol{R}$ can be determined by a linear method using 15 correspondences like as

for the fundamental matrix [12]. Figure 5 shows the comparison between the conventional fundamental matrix and the radial fundamental matrix. Note that Equation (10) is the general form of the radial fundamental matrix proposed by Barreto and Daniilidis [2].

In Equation (11), $\boldsymbol{F}$ is rank-2, and $\boldsymbol{D}$ is rank-3. Hence, $\boldsymbol{R}$ is rank-2. That is, $\boldsymbol{R}$ has a two-dimensional null space. On the other hand, a vector $\aleph = (a^2 + b^2 - d^{-1}, a, b, 1)^T$ is a right null vector of $\boldsymbol{D}_{d,a,b}$. Thus, if the center of lens distortion (a, b) is known, we can determine d uniquely in both right and left null spaces of $\boldsymbol{R}$, which are the lens distortion parameters of the projector and camera. Once the lens distortion is estimated, we can obtain the fundamental matrix $\boldsymbol{F}$ by multiplying the pseudo-inverse of $\boldsymbol{D}$ to both sides of $\boldsymbol{R}$.

Unfortunately, the decomposition of $\boldsymbol{R}$ into d_p, d_c, and $\boldsymbol{F}$ is not unique, unless the center of distortion is known (Appendix A). We solve this ambiguity by assuming the center of lens distortion coincides with the principal point, and estimating the principal point from the point correspondences. We will describe the details in the next section.

5.2. Estimating radial fundamental matrix

Suppose the principal points of projector and camera are given as $\boldsymbol{p}_p$ and $\boldsymbol{p}_c$. Then, the focal length f_p and f_c of respectively projector and camera can be estimated from the fundamental matrix $\boldsymbol{F}$ as follows [4]:

$$\begin{aligned} f_p^2 &= -\frac{\boldsymbol{p}_c^T[\boldsymbol{e}_c]_\times \hat{\boldsymbol{I}}_3 \boldsymbol{F}^T \boldsymbol{p}_p \boldsymbol{p}_c^T \boldsymbol{F} \boldsymbol{p}_c}{\boldsymbol{p}_c^T[\boldsymbol{e}_c]_\times \hat{\boldsymbol{I}}_3 \boldsymbol{F}^T \hat{\boldsymbol{I}}_3 \boldsymbol{F} \boldsymbol{p}_c} \\ f_c^2 &= -\frac{\boldsymbol{p}_p^T[\boldsymbol{e}_p]_\times \hat{\boldsymbol{I}}_3 \boldsymbol{F} \boldsymbol{p}_c \boldsymbol{p}_p^T \boldsymbol{F}^T \boldsymbol{p}_p}{\boldsymbol{p}_p^T[\boldsymbol{e}_p]_\times \hat{\boldsymbol{I}}_3 \boldsymbol{F} \hat{\boldsymbol{I}}_3 \boldsymbol{F}^T \boldsymbol{p}_p}, \end{aligned} \tag{12}$$

where $\boldsymbol{e}_p, \boldsymbol{e}_c$ are left and right null vectors of F, $[\cdot]_\times$ is a skew-symmetric matrix generating a vector product, and $\hat{\boldsymbol{I}}_3 = \mathrm{diag}(1, 1, 0)$. The vectors are represented in homogeneous coordinates.

In practice, however, solving Equation (12) is sensitive to even small errors in the fundamental matrix, and therefore the result is unreliable. The right side of Equation (12) can become negative, which makes the solution impossible. This situation can occur even when the correspondence has little error [11]. In order to estimate the focal lengths from the fundamental matrix in a stable manner, Kanatani et al. propose the approach of selecting only inliers that generate possible solutions [13]. Hartley and Silpa-Anan propose a method which they call as a-priori algorithm, where an objective function is introduced to penalize impossible solutions [11]. In the a-priori method, the principal points of the images can also be estimated as well as the fundamental matrix using the priors. As mentioned in Section 5.1, it is necessary to estimate the center of lens distortion in order to decompose the radial fundamental matrix uniquely. Thus, we assume the center of lens distortion coincides with the principal point, and estimate the principal point during the estimation of the radial fundamental matrix.

We formulate the estimation of radial fundamental matrix as the minimization of the following objective function, and solve it with respect to $\boldsymbol{F}, \boldsymbol{p}_p, \boldsymbol{p}_c, d_p$, and d_c by Levenberg-Marquardt method:

$$C_R(\boldsymbol{F}, \boldsymbol{p}_p, \boldsymbol{p}_c, d_p, d_c) + C_f(\boldsymbol{F}, \boldsymbol{p}_p, \boldsymbol{p}_c) + C_p(\boldsymbol{p}_p, \boldsymbol{p}_c). \tag{13}$$

C_R is the error term that describes how the radial fundamental matrix fits the correspondence. Among several choices for the error function, we use the Sampson approximation of reprojection error [12] and define the error term as

$$C_R(\boldsymbol{F}, \boldsymbol{p}_p, \boldsymbol{p}_c, d_p, d_c) = \sum_i \frac{\boldsymbol{u}_i^T \boldsymbol{R} \boldsymbol{x}_i}{|\hat{\boldsymbol{I}}_4 \boldsymbol{R}^T \boldsymbol{u}_i|^2 + |\hat{\boldsymbol{I}}_4 \boldsymbol{R} \boldsymbol{x}_i|^2}, \tag{14}$$

where $\boldsymbol{u}_i = (u_i^2 + v_i^2, u_i, v_i, 1)^T$ and $\boldsymbol{x}_i = (x_i^2 + y_i^2, x_i, y_i, 1)^T$ are the lifted coordinates of correspondences in homogeneous representation, and $\hat{\boldsymbol{I}}_4 = \mathrm{diag}(1, 1, 1, 0)$. $\boldsymbol{R}$ is computed using Equation (11).

C_f imposes a penalty on the negative values in the right side of Equation (12):

$$C_f(\boldsymbol{F}, \boldsymbol{p}_p, \boldsymbol{p}_c) = w_{fp}(f_{min}^2 - f_p^2) + w_{fc}(f_{min}^2 - f_c^2). \tag{15}$$

Notice that f_p^2 and f_c^2 in the right side are the values computed by Equation (12) and therefore can be negative. f_{min} is the predefined minimum value of the focal length. w_{fp}, w_{fc} are the weights that take positive values when f_p^2, f_c^2 are less than f_{min}^2, and otherwise 0. In our experiments, we used $f_{min}^2 = 0$. w_{fp}, w_{fc} can be very small values.

C_p is the prior term of principal points. We use the equation

$$C_p(\boldsymbol{p}_p, \boldsymbol{p}_c) = w_{pp}|\boldsymbol{p}_p - \bar{\boldsymbol{p}}_p|^2 + w_{pc}|\boldsymbol{p}_c - \bar{\boldsymbol{p}}_c|^2 \tag{16}$$

where $\bar{\boldsymbol{p}}_p$ and $\bar{\boldsymbol{p}}_c$ are the priors for the principal points of projector and camera, and w_{pp} and w_{pc} are the weights, which can be very small again. In our experiments, we used the image center as the prior for the principal point of a camera. The principal point of a projector is usually located below the image center, and often available in the reference manual.

Initialization: The initial parameters for the non-linear optimization of Equation (13) are chosen so that the solution of Equation (12) is at least possible, as follows. First, the initial value of the principal points $\dot{\boldsymbol{p}}_p, \dot{\boldsymbol{p}}_c$ are respectively set to $\bar{\boldsymbol{p}}_p, \bar{\boldsymbol{p}}_c$. The initial focal lengths for a projector and a camera, $\dot{f}_p, \dot{f}_c$, are set to any possible values. In our implementation, we use the diagonal lengths of projector and camera images. Then, a radial fundamental matrix $\ddot{\boldsymbol{R}}$ is

Table 1. Stability of estimating intrinsic parameters. Ground truth parameters are obtained by Audet and Okutomi's method using a patterned calibration board [1]. The ground truth distortion is not presented because they used a different lens distortion model.

	f_p	f_c	d_p	d_c	$\boldsymbol{p}_p$	$\boldsymbol{p}_c$
1	1798.2	2127.3	-8.0×10^{-8}	-1.4×10^{-8}	(511.5, 636.6)	(511.1, 384.1)
2	1808.3	2138.7	-7.9×10^{-8}	-1.2×10^{-8}	(513.2, 631.9)	(512.9, 383.8)
3	1780.9	2056.2	-8.1×10^{-8}	-1.2×10^{-8}	(512.9, 637.4)	(518.3, 384.8)
4	1799.5	2094.9	-8.0×10^{-8}	-1.1×10^{-8}	(516.1, 627.1)	(511.9, 383.9)
Ground truth [1]	1869.3	2196.9	—	—	(512.7, 683.1)	(518.1, 398.8)

estimated by the linear method using 15 reliable correspondences. The reliable corresponding points are determined by RANSAC. $\ddot{\boldsymbol{R}}$ is then decomposed into initial lens distortions $\dot{d}_p, \dot{d}_c$ and fundamental matrix $\ddot{\boldsymbol{F}}$ using the initial principal points. By multiplying the inverses of the initial intrinsic matrices to both sides of $\ddot{\boldsymbol{R}}$, we obtain an initial essential matrix $\ddot{\boldsymbol{E}}$ as

$$\ddot{\boldsymbol{E}} = \boldsymbol{K}^{-T}_{\dot{f}_p,\dot{\boldsymbol{p}}_p} \ddot{\boldsymbol{F}} \boldsymbol{K}^{-1}_{\dot{f}_c,\dot{\boldsymbol{p}}_c}. \tag{17}$$

In general, $\ddot{\boldsymbol{E}}$ does not satisfy the requirements of the essential matrix. By replacing the least singular value of $\ddot{\boldsymbol{E}}$ with 0 and the other two with 1, we obtain a valid initial essential matrix $\dot{\boldsymbol{E}}$. Then, the initial fundamental matrix $\dot{\boldsymbol{F}}$ can be obtained as follows:

$$\dot{\boldsymbol{F}} = \boldsymbol{K}^{T}_{\dot{f}_p,\dot{\boldsymbol{p}}_p} \dot{\boldsymbol{E}} \boldsymbol{K}_{\dot{f}_c,\bar{\boldsymbol{p}}_c}. \tag{18}$$

5.3. Estimating extrinsic parameters

The essential matrix $\boldsymbol{E}$ is obtained by the following formula:

$$\boldsymbol{E} = \boldsymbol{K}^{-T}_{f_p,\boldsymbol{p}_p} \boldsymbol{F} \boldsymbol{K}^{-1}_{f_c,\boldsymbol{p}_c} \tag{19}$$

$$= [\boldsymbol{t}]_\times \boldsymbol{\Theta}. \tag{20}$$

The extinsic parameters $\boldsymbol{t}$ and $\boldsymbol{\Theta}$ are respectively the translation and the rotation of projector's relative motion, which can be determined by applying the singular value decomposition to $\boldsymbol{E}$ [12].

6. Experiment results

We conducted several experiments using real images using two data projectors (EPSON EMP-765 and EB-1735W of respectively 1024×768 and 1280×800 resolution) and a USB camera (Artray ARTCAM 300MI of 2048×1536 resolution). The field of view, position and orientation of the projector and the camera are uncalibrated. The prior of principal point of the projector is obtained from the reference manual. We projected full-resolution complementary Gray code patterns and 4 sinusoidal patterns in horizontal and vertical directions, using 48 images for EMP-765 and 50 images for EB-1735W. Some of the acquired images are shown in Figure 2.

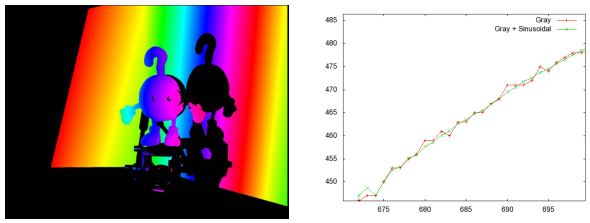


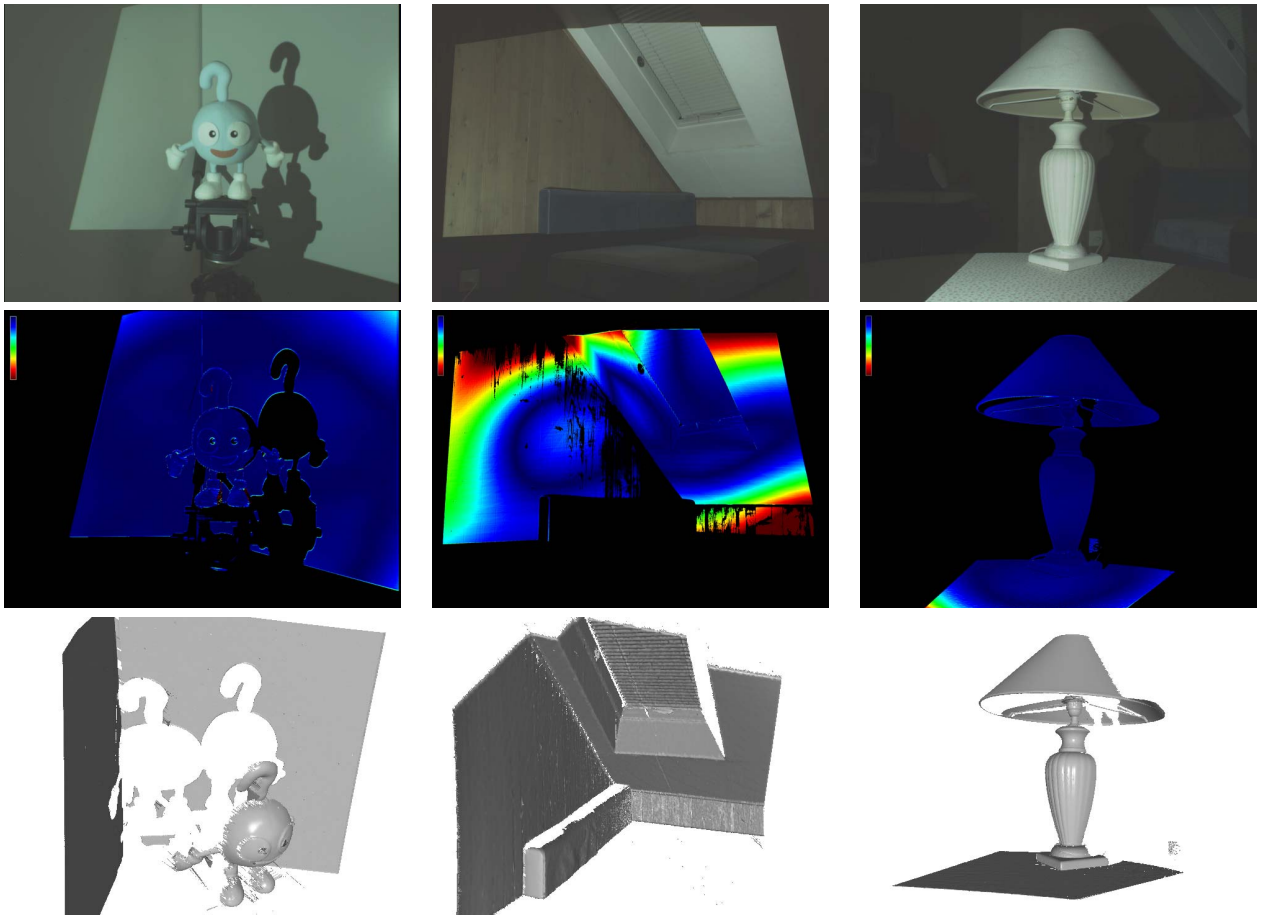
Figure 6. Point correspondences. (left) Correspondences from camera image coordinates to projector image coordinates in horizontal direction are shown in color. (right) Comparison of the correspondences obtained by only Gray code, and by Gray code and sinusoidal code. The horizontal and vertical axes indicate the image coodinates.

We applied our calibration algorithm to the images of 4 different objects acquired using the same geometric parameters. The results for the estimated parameters are summarized in Table 1. We have also compared the results with the parameters estimated using a calibration pattern proposed by Audet and Okutomi [1]. The estimated camera parameters are consistent across experiments and sufficiently close to the values obtained using an additional calibration apparatus. This is contrary to the fact that in passive stereo systems, self-calibration using the decomposition of the fundamental matrix is known to be sensitive and regarded as useless in practice [12]. The major reason is that in our system the number and accuracy of point correspondences is considerably higher than those in passive stereo systems. We have tried similar experiments with different combinations of projectors and cameras, and found the algorithm always converges in consistent manners. Figure 7 summarizes the results of our calibration for various objects.

7. Summary

We have described a method for self-calibration of projector-camera systems. Compared to existing techniques for calibrating projector-camera systems, our method has the following advantages:

First, our method can estimate all geometric parameters that can be changed by a user in acquisition. This makes the acquisition system extremely easy to use. Existing techniques [9] require a calibrated camera, which makes it hard for non-experts to use it. For instance, we can start 3D acquisition using a pair of newly-purchased projector and



$f_p = 1691.786, \quad f_c = 1923.786$ $d_p = -1.334487 \times 10^{-8}$ $d_c = -4.489834 \times 10^{-8}$ $\boldsymbol{p}_p = (512.5, 638.27)$ $\boldsymbol{p}_c = (640.4, 512.5)$	$f_p = 1012.871, \quad f_c = 1437.654$ $d_p = -4.744778 \times 10^{-8}$ $d_c = -5.919533 \times 10^{-8}$ $\boldsymbol{p}_p = (640.4, 696.86)$ $\boldsymbol{p}_c = (1024.6, 768.5)$	$f_p = 1379.434, \quad f_c = 1959.396$ $d_p = -2.395326 \times 10^{-8}$ $d_c = -5.528283 \times 10^{-8}$ $\boldsymbol{p}_p = (640.5, 696.87)$ $\boldsymbol{p}_c = (800.5, 600.5)$

Figure 7. Reconstructed 3D shapes. From top to bottom, each row shows the images of objects under white illumination, the reprojection error of $\boldsymbol{R}$ where red indicates at least one pixel error, the 3D shapes reconstructed using the estimated camera parameters and the correspondences used for the calibration, and the estimated intrinsic parameters.

camera without any calibration.

Second, the accuracy of point correspondence is considerably increased by combining Gray code and phase-shifting code, followed by the estimation of projector and camera response functions. Thanks to the accrate correspondences, the epipolar geometry can be reconstructed stably, and the camera parameters are estimated reliably.

Third, the lens distortions of projector and camera are corrected simultaneously. This is also important to make the system practical since most consumer cameras suffer from severe lens distortion.

Limitations: Our calibration method can estimate the camera geometry only up to a global scale. To determine the remaining degree of freedom, we have to use an additional apparatus, such as an object of known size or calibrated light sources, or additional camera to reconstruct a trifocal tensor [12]. Another issue is the strong coupling of the focal lengths, which leads to either larger or smaller focal length than actual in certain camera configurations as pointed out by Bougnoux [4]. Large lens distortion such as those of fish-eye camera are currently not considered. Integrating general camera model, such as rational function lens distortion model [6], remains to be investigated.

Acknowledgement

We appreciate valuable comments from anonymous reviewers. We thank Andrea Fossati for English manuscript proofreading. This work was supported by JSPS Grant-in-Aid for Young Scientists (B) No.22700190.

References

[1] S. Audet and M. Okutomi. A user-friendly method to geometrically calibrate projector-camera systems. In *Proc. International Workshop on Projector-Camera Systems*, pages 47–54, 2009.

[2] J. P. Barreto and K. Daniilidis. Fundamental matrix for cameras with radial distortion. In *Proc. International Conference on Computer Vision*, pages 625–632, 2005.

[3] D. Bergmann. New approach for automatic surface reconstruction with coded light. In *Proc. Remote Sensing and Reconstruction for Three-Dimensional Objects and Scenes*, volume 2572, pages 2–9, 1995.

[4] S. Bougnoux. From projective to euclidean space under any practical situation. In *Proc. International Conference on Computer Vision*, pages 790–798, 1998.

[5] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, pages 120–126, Nov. 2000.

[6] D. Claus and A. W. Fitzgibbon. A rational function lens distortion model for general cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 213–219, 2005.

[7] G. Falcao, N. Hurtos, J. Massich, and D. Fofi. Projector-Camera Calibration Toolbox, 2009. http://code.google.com/p/procamcalib.

[8] A. W. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *Proc. Computer Vision and Pattern Recognition*, pages 125–132, 2001.

[9] R. Furukawa and H. Kawasaki. Dense 3D reconstruction with an uncalibrated stereo system using coded structured light. In *Proc. International Workshop on Projector-Camera Systems*, pages 107–116, 2005.

[10] R. Furukawa, H. Q. H. Viet, H. Kawasaki, R. Sagawa, and Y. Yagi. One-shot range scanner using coplanarity constraints. In *Proc. International Conference on Image Processing*, pages 1524–1527, 2008.

[11] R. Hartley and C. Silpa-Anan. Reconstruction from two views using approximate calibration. In *Proc. Asian Conference on Computer Vision*, pages 338–343, 2002.

[12] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd edition, 2004.

[13] K. Kanatani, A. Nakatsuji, and Y. Sugaya. Stabilizing the focal length computation for 3-d reconstruction from two uncalibrated views. *International Journal of Computer Vision*, 66(2):109–122, 2006.

[14] T. Okatani and K. Deguchi. Autocalibration of a projector-camera system. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 27:1845–1855, 2005.

[15] M. Pollefeys and L. V. Gool. A stratified approach to metric self-calibration. In *Proc. International Conference on Computer Vision*, pages 407–412, 1997.

[16] R. Raskar and P. A. Beardsley. A self-correcting projector. In *Proc. Computer Vision and Pattern Recognition*, pages 504–508, 2001.

[17] J. Salvi, J. Pages, and J. Batlle. Pattern codification strategies in structured light systems. *Pattern Recognition*, 37(4):827–849, 2004.

[18] K. Sato and S. Inokuchi. Range-imaging system utilizing nematic liquid crystal mask. In *Proc. Int. Conf. on Computer Vision*, pages 657–661, 1987.

[19] N. Snavely, S. M. Seitz, and R. Szeliski. Modeling the world from internet photo collections. *International Journal of Computer Vision*, 80(2):189–210, 2008.

[20] R. M. Steele and C. Jaynes. Overconstrained linear estimation of radial distortion and multi-view geometry. In *Proc. European Conference on Computer Vision*, pages 253–264, 2006.

[21] B. Triggs. Autocalibration and the absolute quadric. In *Proc. Computer Vision and Pattern Recognition*, pages 609–614, 1997.

[22] A. O. Ulusoy, F. Calakli, and G. Taubin. Robust one-shot 3d scanning using loopy belief propagation. In *Proc. CVPR Workshop on Applications of Computer Vision in Archaeology*, pages 15–22, 2010.

[23] L. Zhang, B. Curless, and S. M. Seitz. Rapid shape acquisition using color structured light and multi-pass dynamic programming. In *Proc. the 1st International Symposium on 3D Data Processing, Visualization, and Transmission*, pages 24–36, 2002.

[24] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.

Appendix

A. Ambiguity in decomposing radial fundamental matrix

Given two pairs of lens distortion parameters (d, a, b) and (d', a', b'), consider a 3×3 rank-3 matrix $\boldsymbol{H}$:

$$\boldsymbol{H} = \boldsymbol{I}_3 + d' \begin{pmatrix} a' \\ b' \\ 1 \end{pmatrix} \left(2(a-a'), 2(b-b'), (a'^2 + b'^2) - (a^2 + b^2)\right)$$

where $\boldsymbol{I}_3$ is a 3×3 identity matrix. Then, the lens distortion matrix $\boldsymbol{D}$ in Equation (9) satisfies

$$\boldsymbol{H}\boldsymbol{D}_{d,a,b} = \boldsymbol{D}_{d',a',b'}. \tag{21}$$

Multiplying any fundametal matrix $\boldsymbol{F}$ to both sides from left, we have

$$\boldsymbol{F}\boldsymbol{H}\boldsymbol{D}_{d,a,b} = \boldsymbol{F}\boldsymbol{D}_{d',a',b'}. \tag{22}$$

$\boldsymbol{F}\boldsymbol{H}$ is a fundamental matrix because $\boldsymbol{H}$ is a rank-3 matrix. Equation (22) shows the factorization of $\boldsymbol{F}\boldsymbol{D}$ is ambiguous. This ambiguity exists for both projector and camera in Equation (11). The decomposition is unique when the lens has no distortion ($d = d' = 0$) or the center of lens distortion is known ($a = a'$ and $b = b'$).