S. Vagharshakyan, A. Durmush, O. Suominen, R. Bregović, and A. Gotchev, "Accuracy Evaluation of a Linear Positioning System for Light Field Capture," *7th Asian Conf. on Intelligent Information and Database Systems (ACIIDS)*, Bali, Indonesia, March 2015 – published in LNCS vol. 9012, pp. 388-397.

Accuracy Evaluation of a Linear Positioning System for Light Field Capture

Suren Vagharshakyan, Ahmed Durmush, Olli Suominen, Robert Bregovic, Atanas Gotchev

Tampere University of Technology, Finland
suren.vagharshakyan@tut.fi, ahmed.durmush@tut.fi,
olli.j.suominen@tut.fi, robert.bregovic@tut.fi,
atanas.gotchev@tut.fi

Abstract. In this paper a method has been proposed for estimating the positions of a moving camera attached to a linear positioning system (LPS). By comparing the estimated camera positions with the expected positions, which were calculated based on the LPS specifications, the manufacturer specified accuracy of the system, can be verified. Having this data, one can more accurately model the light field sampling process. The overall approach is illustrated on an in-house assembled LPS.

1 Introduction

In order to properly capture the light field, significant number of densely positioned cameras are needed, e.g. a camera array. For static scenes, such camera array can be replaced by a system in which a single camera is precisely positioned on the camera array plane. A motorized linear positioning system (LPS) which can position a camera on a plane with high precision (sub-pixel resolution) has been built so that it is possible to capture images from different viewpoints. Thus, the spatial resolution of the light field captured by moving a single camera is higher than a light field, which would be captured by an array of similar cameras. In the following section we will describe in detail the built LPS and present manufacturer provided specifications for it. In Section 1.2 we describe the well-known pinhole camera model together with modeling of the lens distortions. In Section 1.3 we explain the mathematical model of the light field capturing process using an LPS. In Section 2 we describe the methodology of estimating model parameters for the LPS, which will allow us to evaluate the accuracy specifications provided by the manufacturer. Experimental results presented in Section 3 show the importance of evaluating the working accuracy of the LPS based on proposed methods in comparison to standard methods.

1.1 LPS Specifications

The in-house assembled LPS is composed of a hard alloy aluminum base, precision lead screw, anti-backlash nut and a stepper motor. The hard alloy aluminum base

provides the structural stability to the system. The precision lead screw with an anti-backlash nut converts rotary motion of the stepper motor to precise linear movements. Backlash describes the loss of motion due to gaps between the mechanical parts and it directly influences the positioning accuracy. Therefore, anti-backlash mechanisms are used in precision required applications. Furthermore, stepper motors used in the LPS have built-in encoders that provide closed loop servo like operation and increase the precision of the system.

The specifications of the LPS based on the characteristics of the mechanical and electrical parts are summarized in Table 1. The accuracy is measured by the deviation of the actual position from the desired position along the whole travel distance whereas repeatability is the measure of a system's consistency to achieve identical results. Straight-line accuracy is the measure of deviation from straight line along the motion axis.

Accuracy	±20μm
Precision (Repeatability)	4μm
Straight Line Accuracy	38μm
Maximum Linear Speed	20 mm/s
Maximum Payload	20 kg
X Axis Travel Distance	1524mm
Y Axis Travel Distance	1016mm

Table 1. Specifications provided by the manufacturer.

1.2 Pinhole camera with lens distortions image formation model

A view captured by a camera is formed by projecting 3D points of the scene on the image plane of the camera according to the transform

$$sm = A[R|t]M$$

$$m = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}, A = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}, t = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}, M = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

where (X, Y, Z) is the 3D position of a scene point expressed through its homogeneous coordinates M, $f = (f_x, f_y)$ and $c = (c_x, c_y)$ are the focal lengths and principal point coordinates measured in pixels, respectively, r_{ij} represents the camera rotation matrix and (t_1, t_2, t_3) are coordinates of the camera's optical center. Moreover, (f, c) and (R, t) are respectively the intrinsic and extrinsic camera parameters. For camera rotation there exists a more compact representation based only on three free parameters, however the matrix form will be kept for convenience. The pair (u, v) is the 2D coordinate of the projected point as a pixel position on the image plane. The result of

the perspective projection is written in homogeneous coordinates m with a scaling factor s.

The pinhole camera model describes the perspective projections for a given camera. For real world cameras, another important problem is to model the distortions produced by camera lenses. Lens radial and tangential distortions at the image plane can be formalized using Brown's distortion model [1]

$$x^* = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_1 xy + p_2(r^2 + 2x^2)$$

$$y^* = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + p_1(r^2 + 2y^2) + 2p_2 xy$$

where $(k_1, k_2, k_3, p_1, p_2)$ are radial and tangential distortion coefficients, (x, y) are undistorted coordinates at image plane $\left(r = \sqrt{x^2 + y^2}\right)$ and (x^*, y^*) are final coordinates taking into account lens distortions. Hereafter, the distortion transform is denoted as D. The model can be extended by adding higher order polynomial terms of r and assuming also fractional terms of r.

The image formation procedure is formalized by Eq. 1. The matrix F combines the extrinsic matrix (R, t), the distortion-modelling transform and the intrinsic matrix (see Fig. 1). For brevity, the set of intrinsic parameters with lens distortion parameters are denoted as $I = (f, c, k_1, k_2, k_3, p_1, p_2)$.

$$\begin{bmatrix} u \\ v \end{bmatrix} = F_{R,t,I}\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) \tag{1}$$

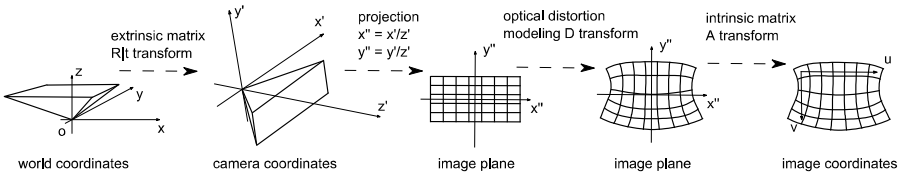


Fig. 1. Mathematical model of image formation

1.3 Model of the motorized LPS

In our model of the motorized LPS, it is assumed that a single specific camera moves in space along a line with fixed step size d and takes photos of a stationary 3D scene. Using the mathematical model of the capturing function presented before, at each movement step, the camera projection can be described as function $F_{R,t_i,I}$, where parameters R, I are fixed during the movement, while the positions t_i are linearly changing, $t_i = l_0 + (id)l_n,\ i = 1, \dots, K$. Here, l_0 is vector to the start of the line over which the camera is moving and l_n is a normalized direction vector of the line as shown in Fig 2.

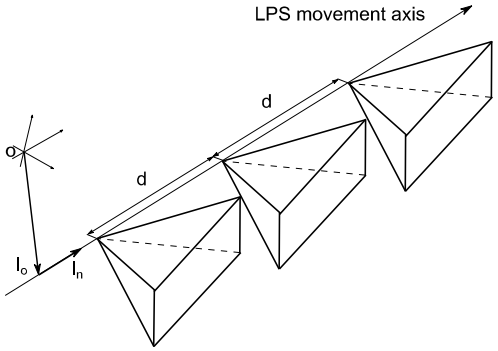


Fig. 2. Parameterization of the motorized LPS.

2 Estimation

2.1 Estimation of parameters for a single camera

Extraction of camera parameters can be broken down to estimating the 3D position, rotation and distortion coefficients of the camera. For that purpose, for a given camera setup, projections of 3D points are measured and then the inverse problem is solved to estimate R, t, I parameters. In other words, for a given set of points in space $M_k = \begin{bmatrix} x_k \\ y_k \\ z_k \end{bmatrix}, k = 1, \dots, N$ and their corresponding projections $m_k = \begin{bmatrix} u_k \\ v_k \end{bmatrix}, k = 1, \dots, N$,

$$\underset{R,t,I}{\operatorname{argmin}} \sum_{k=1}^{N} \left\| F_{R,t,I}(M_k) - m_k \right\|^2$$

is found. When all parameters are unknown and the function F is nonlinear, the solution should be obtained through nonlinear optimization algorithms. For that purpose, we use camera estimation algorithms implemented in the OpenCV library [2] based on [3, 4].

2.2 Motorized LPS parameter estimation

Estimation of the motorized LPS parameters refers here to estimating the movement precision over the LPS movement axis for specific fixed camera with unknown intrinsic parameters. Methods for unconstrained estimation of the camera locations at each capture step do not provide valuable results. Particularly, results obtained through estimation of the unconstrained locations with common intrinsic parameters and lens distortion coefficients [2, 3], i.e.

$$\underset{\substack{I,R_i,t_i,\\ i=1,\dots,K}}{\operatorname{argmin}} \sum_{i=1}^{K}\sum_{k=1}^{N} \left\| F_{R_i,t_i,I}(M_k) - m_{k,i} \right\|^2 \tag{2}$$

or independent estimation of the locations

$$\underset{R_i,t_i}{\operatorname{argmin}} \sum_{k=1}^{N} \left\| F_{R_i,t_i,I^*}(M_k) - m_{k,i} \right\|^2, i = 1,\dots,K \tag{3}$$

using beforehand independently estimated lens distortions and intrinsic ters $I^* = (f^*, c^*, k_1^*, k_2^*, k_3^*, p_1^*, p_2^*)$, are not precise enough. They give only a rough estimation of the camera movement through space.

In our proposed approach we suggest considering a linear dependence between the camera positions. In this case the problem can be formulated as

$$\underset{R,l_0,l_n}{\operatorname{argmin}} \sum_{i=1}^{K}\sum_{k=1}^{N} \left\| F_{R,l_0+d_i l_n,I^*}(M_k) - m_{k,i} \right\|^2 \tag{4}$$

where the rotation denoted by R is common for all positions, l_0, l_n define a line in the space, $\{d_i\}_{i=1,\dots,K}$ is the distribution of positions over that line, and I^* denotes the intrinsic parameters and lens distortions coefficients which are estimated beforehand using (2). The minimization problem is solved by using the Levenberg–Marquardt algorithm [5]. It is a non-linear minimization algorithm and therefore providing good initial estimates for R, l_0, l_n improves the estimation performance. In particular, initial estimates for the minimization algorithm (4) can be obtained by using independently estimated 3D position $t_i, i = 1,\dots,K$ and rotation $R_i, i = 1,\dots,K$ of each camera position by minimizing (3). The initial common rotation estimation R is the mean of all rotations $R_i, i = 1,\dots,K$ and l_0, l_n are the least square solution of

$$\begin{bmatrix} t_1 \\ \vdots \\ t_K \end{bmatrix} = [l_0\ l_n] \begin{bmatrix} 1 & \cdots & 1 \\ d & \cdots & Kd \end{bmatrix} \tag{5}$$

where d is a predefined uniform step size. Alternatively, l_0, l_n can be found based on principle component analysis for fitting a line to points $t_k, k = 1,\dots,N$ [6].

After finding solution for (4), $d_i, i = 1,\dots,K$ are estimated such that

$$\underset{d_i}{\operatorname{argmin}} \sum_{k=1}^{N} \left\| F_{R,l_0+d_i l_n,I^*}(M_k) - m_{k,i} \right\|^2, i = 1,\dots,K. \tag{6}$$

In fact, the interesting values are $d_i, i = 1,\dots,K$, which allow estimating or verifying the precision of the movement over the LPS compared to the intended uniform step size d. The proposed hybrid algorithm can be summarized as follows:

Input: K number of views, N number of observed fixed points of the scene and $M_k, k = 1,\dots,N$ their 3D coordinates in the space, $m_{k,i}, k = 1,\dots,N, i = 1,\dots,N$

corresponding projection image coordinates in pixels of i-th point in k-th view, d motorized LPS movement step size, and $I^* = f^*, c^*, k_1^*, k_2^*, k_3^*, p_1^*, p_2^*$ previously estimated intrinsic parameters and lens distortion coefficients.

1. Find R_i rotation matrix and t_i position ($i = 1, \dots, K$) of the camera for each view by solving (3).
2. Calculate mean rotation matrix R based on rotation matrices R_i. Calculate least square solution of (5) for l_0, l_n (normalize l_n, if it is necessary).
3. Using already found R, l_0, l_n as initial values, update them by solving (4)
4. Solve (6) to find d_i, using grid search method for each $i = 1, \dots, N$ independently.
5. Repeat step 3, 4 until $err = \sum_{i=1}^{K} \sum_{k=1}^{N} \left\| F_{P_i}(M_k) - m_{k,i} \right\|^2 < T$, where T is the predefined convergence tolerance.

3 Experimental results

3.1 Experiment with synthetic data

To evaluate the proposed algorithm, we generate a synthetic dataset based on the mathematical model presented in sections 1.2 and 1.3. The dataset is generated by projecting synthetic chessboard corners (Fig. 4(a)) to virtual cameras located along the line and adding noise to projected coordinates to model inaccuracy of the feature detection algorithms. The inaccuracy is assumed to be less than 1px in absolute value along each axis. A realistic set of LPS parameters is used in projection calculation, but lens distortions are not considered. For camera motion we consider a scenario where consecutive step size varies as a sine wave Fig. 3 (ground truth). As seen in the figure, estimated consecutive distances are close to the ground truth even in the presence of noise, which shows that the proposed algorithm is robust to inaccuracy in feature point detection. This makes our approach suitable for LPS accuracy evaluation in a real world setting.

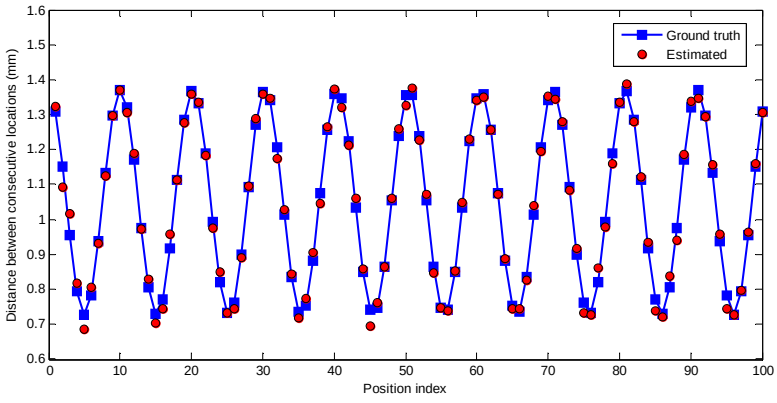
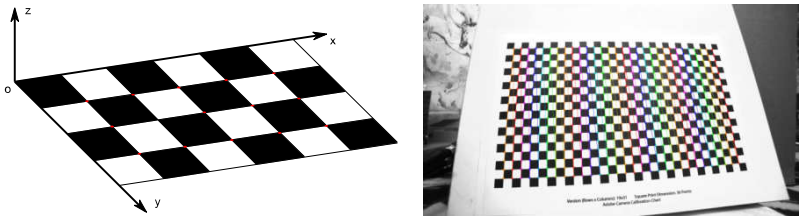


Fig. 3. Comparison of the estimated locations against the ground truth for synthetic dataset.

3.2 Experiment with real data

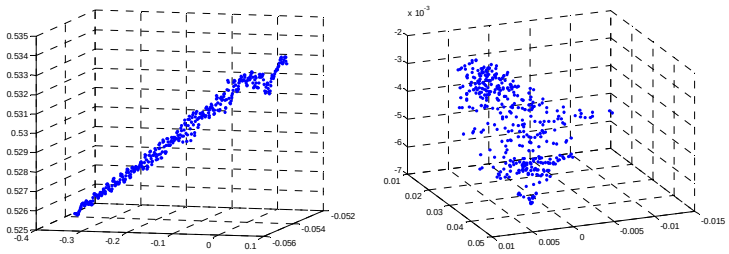
To form a set of 3D points, we use a chessboard texture on a plane such that each inner corner of the chessboard defines a point in 3D space as shown in Fig 4(a). Corner detection algorithm presented in OpenCV allows detecting corners with sub pixel precision in image coordinates, see Fig 4(b). For the given LPS, a nominal step size of 1 mm is used for camera motion and a chessboard is captured by the camera at each step for later processing and movement precision estimation, see Fig. 5.



(a) (b)

Fig. 4. (a) Origin of the 3D space positioned at the edge corner of a chessboard and all inner points shown as red dots related to selected axis. (b) Detected image coordinates of inner points of a chessboard (18x30 inner points).

Fig. 5. Example of views taken with 11mm distance using LPS. Original data contain 400 views of size 1980x1080px taken with 1mm distance.



(a) (b)

Fig. 6. 3D coordinates of estimated positions (a) in meters and rotations (b) in radians obtained by solving Eq. 2 of the camera at 400 locations over the LPS movement axis with 1mm step size.

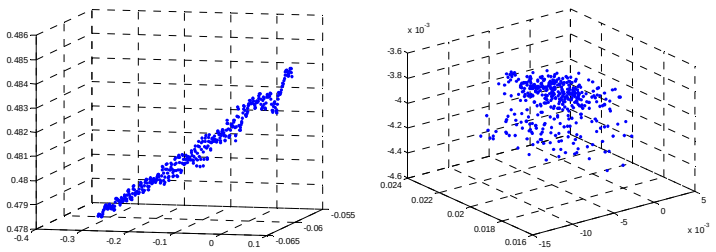
Solving Eq. 2 is not trivial because of the large number of unknowns. For example, in the case of $N = 32$ (chessboard with 4x8 inner corners) and $K = 400$ (K is the number of images), it is necessary to estimate 9 intrinsic parameters and $9K$ extrinsic parameters based on NK measurements, in other words, to solve a nonlinear optimization problem with 3609 unknowns based on 12800 measurements. Due to the large number of unknowns, the method is too slow in practice – for 30 iterations of minimization it takes about three hours of computation. Estimated location results are presented in Fig. 6, and the estimated intrinsic parameters are:

$f = 9m, c_x = 962.68px, c_y = 539.13px, k_1 = -0.233, k_2 = 0.247, k_3 = -0.173, p_1 = 0.0028, p_2 = -0.0035.$

Beforehand estimation of the lens distortion coefficients independent from LPS can be done by using only e.g. 6 arbitrary positioned images of the chessboard, similar to the one shown in Fig. 4(b). For a better estimation, it is necessary to have images where a large part of the field of view is occupied by a chessboard with a large number of inner corners. In that case, in Eq. 2 one will have a large number of measurements (3240) with relatively small number of unknowns (69). Then one can effectively solve Eq. 2 using the Levenberg–Marquardt algorithm. The obtained results are:

$f = 8.18m, c_x = 951.59px, c_y = 551.08px, k_1 = -0.169, k_2 = 0.111, k_3 = -0.02, p_1 = -0.0007, p_2 = -0.00002.$

After determining the lens distortion coefficients and camera intrinsic parameters, the camera positions and rotations are calculated by solving Eq. 3 and the results are shown in Fig. 7. They are quite similar to the results shown in Fig. 6, however, their calculation takes only a few seconds. Unfortunately, the estimation quality is not high enough to allow making conclusions about the LPS working accuracy.



(a) (b)

Fig. 7. 3D coordinates of estimated positions (a) and rotations (b) of the camera obtained by solving Eq. 3, at 400 locations over the LPS movement axis with 1mm step size.

The proposed constraint in Eq. 4 together with Eq. 6 facilitates getting meaningful results about the accuracy of the LPS movement as shown in Fig. 8. For the given LPS, maximum estimated positioning misalignment of the LPS has an absolute value less than 0.04mm. Together with joint camera rotation estimation for all positions, results provide LPS model parameters for further processing and properly interpreting sampled light field data. We also noticed that the precision of the lens distortion coefficient estimation highly affects the camera location estimation. However, the results

clearly show that the proposed method is able to detect incorrect and unexpected camera movement by the LPS. This is illustrated in Fig. 9 by adding outliers in the original dataset.

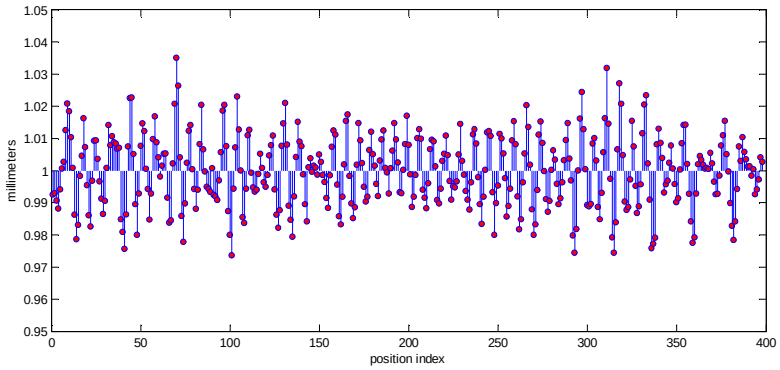


Fig. 8. Distances between estimated consecutive locations ($d_i - d_{i+1}$) of the camera over the line in space. They are calculated based on the proposed method from a data set containing images with 1 mm step size along the LPS movement axis.

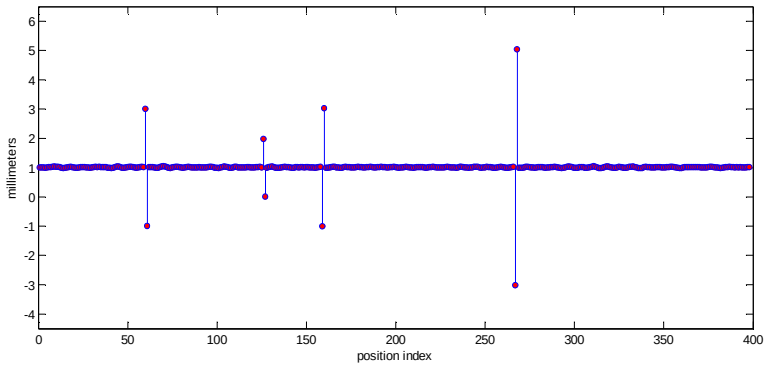


Fig. 9. Similar results as in Fig. 7 for same dataset, but with several outliers added artificially by replacing images in dataset, which are clearly visible from the results.

4 Conclusion

A motorized LPS allowing very fine light field sampling has been constructed and a method for its precision verification has been developed. Proposed calibration algorithm also provides estimates of the camera lens distortions and camera rotation during capture. Having this data, one can model the light field sampling process with a higher accuracy. As a future work we plan to extend capturing process from linear capturing system to a planar capturing system and attempt to provide a more advanced estimation procedure for evaluating the accuracy of a 2D positioning system.

S. Vagharshakyan, A. Durmush, O. Suominen, R. Bregović, and A. Gotchev, "Accuracy Evaluation of a Linear Positioning System for Light Field Capture," *7th Asian Conf. on Intelligent Information and Database Systems (ACIIDS)*, Bali, Indonesia, March 2015 – published in LNCS vol. 9012, pp. 388-397.

References

1. Brown, C. Duane, Decentering distortion of lenses, Photogrammetric Engineering, 32 (3): 444–462, 1966
2. OpenCV (Open Source Computer Vision Library) www.opencv.org
3. Jean-Yves Bouguet, Camera Calibration Toolbox for Matlab, www.vision.caltech.edu/bouguetj/calib_doc/
4. Z. Zhang, A Flexible New Technique for Camera Calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22 (11):1330-1334, 2000
5. D. Marquardt, An Algorithm for Least-Squares Estimation of Nonlinear Parameters. SIAM Journal on Applied Mathematics 11 (2): 431–441, 1963
6. Hawkins, M. Douglas, On the Investigation of Alternative Regressions by Principal Component Analysis. Journal of the Royal Statistical Society, Series C, 22 (3): 275–286, 1973