Lens Distortion Correction Using ELM

Chang Liu and Fuchun Sun

State Key Laboratory of Intelligent Technology and Systems,
Tsinghua National Laboratory for Information Science and Technology,
Dept. of Computer Science and Technology, Tsinghua University, P.R. China, 100084
cliu13@mails.tsinghua.edu.cn,
fcsun@mail.tsinghua.edu.cn

Abstract. Lens distortion is one of the major issues in camera calibration since it causes the perspective projection of the camera model to no longer hold. Thus to eliminate lens distortion becomes an essential part of camera calibration. This paper proposes a novel method of correcting lens distortion by implementing extreme learning machine, a new learning algorithm for single-hidden layer feedforward networks. A camera calibration model which contains linear phase for calibration, and non-linear phase for lens distortion correction is introduced. The performance is evaluated in comparison with traditional learning methods, and the results show that the proposed model produces much better performance than that of the others.

Keywords: lens distortion correction, extreme learning machine, direct linear transformation, non-linear model, camera calibration, feedforward neural network.

1 Introduction

Camera calibration is the essential premise of stereo computer vision and three-dimensional reconstruction. The purpose of camera calibration is to reconstruct any world point from image points based on calibrated camera model. Plenty of camera calibration methods have been proposed, and they can be divided mainly into two categories: traditional calibration methods and self-calibration methods.

For a traditional method, a calibration object, whose structure of geometry is known in advance, is always needed. It constructs the constraints of the camera model parameters through the correspondence between the spatial point and the image points (two images for binocular vision), and then solves the optimization problem to obtain the model parameters. For example, the Direct Linear Transformation method of Abdel-Aziz and Karara [1] tries to obtain the camera model parameters by solving a set of linear equations; the two-step calibration method based on radial constrains given by Tsai [2] calculates the camera model parameters via RAC constrains; Zhang's new flexible camera calibration method [3] uses a calibration plane with axis Z in world coordinate equals zero to find the optimization solution.

J. Cao et al. (eds.), *Proceedings of ELM-2014 Volume 2*,
Proceedings in Adaptation, Learning and Optimization 4, DOI: 10.1007/978-3-319-14066-7_3

For a self-calibration method the calibration object is not required, since it directly performs calibration with the relationship of the corresponding points of multi images. For example, Faugeras proves the existence of quadratic nonlinear constraints in each two images, and uses LM algorithm to obtain the parameters [4]. Triggs implanted the consistency of the European transform of the absolute conic to perform the calibration [5].

Other methods include: vanishing points for orthogonal directions [6], self-calibration for varifocal cameras [7] and so on.

The procedure of camera calibration usually contains two phases: the linear phase which tries to find the projection relationship between the world points and the image points by determining the parameters of the camera model, and the non-linear phase which tries to solve the lens distortion of camera.

Extreme learning machine (ELM) is a new learning method for single-hidden layer neural network where the hidden layer neurons can be randomly generated independent of training data and application environment [8]. ELM has proved fine ability in terms of non-linear regression and classification. For example, Rong proposed a recognition scheme for identifying the aircrafts of different types using ELM to train the multiple single-hidden layer feedforward networks [9]. Sun introduced a sale forecasting model which solves a regression problem with the aid of ELM [10]. Xu Y raised a real time transient stability assessment model using ELM to gain better computation speed and accuracy [11] and so on.

This paper presents a new model for camera calibration and lens distortion correction based on ELM. To the best of our knowledge, the application of ELM has not been used in the literature before. The simulation results show that the proposed model performs well in three-dimensional reconstruction and outperforms other traditional regression methods, Back-Propagation Network and Support Vector Machine.

2 Extreme Learning Machine

A typical SLFN usually contains three layers: an input layer with n nodes, a single hidden layer with $\tilde{N}$ nodes and an output layer with m nodes.

For N arbitrary distinct samples $(\boldsymbol{x}_i, \boldsymbol{t}_i)$, where $\boldsymbol{x}_i = (x_{i_1}, x_{i_2}, \cdots, x_{i_n})^T$ and $\boldsymbol{t}_i = (t_{i_1}, t_{i_2}, \cdots, t_{i_m})^T$, the output function of the network is expressed as:

$$f_{\tilde{N}}(\boldsymbol{x}) = \sum_{i=1}^{\tilde{N}} \beta_i G(\boldsymbol{w}_i, b_i, \boldsymbol{x}_j) = \sum_{i=1}^{\tilde{N}} \boldsymbol{\beta}_i g(\boldsymbol{w}_i \cdot \boldsymbol{x}_j + b_i) = \boldsymbol{o}_j, j = 1,2,\cdots,N \tag{1}$$

where $\boldsymbol{\beta}_i = (\beta_{i_1}, \beta_{i_2}, \cdots, \beta_{i_m})^T$ is the weight vector from the hidden layer nodes to the output layer nodes for the ith sample; $\boldsymbol{w}_i = (w_{i_1}, w_{i_2}, \cdots, w_{i_n})^T$ denotes the weight vector from the input layer nodes to the hidden layer nodes and b_i is the bias of the ith hidden layer node. The activation function can be linear function or sigmoid function or other types of functions and here it is defined as linear function.

The traditional method of solving the SLFN is to minimize the cost function of error with gradient decent method. With back-propagation method, it updates the weight vector from learning. Two major problems of this method are, firstly, its convergence speed can be slow and secondly, it may stop at the local minima rather the optimal solution.

For a standard SLFN with $\widetilde{N}$ nodes, it can approximate N samples with zero error, which implies $\sum_{j=1}^{\widetilde{N}}\left\|\boldsymbol{o}_j-\boldsymbol{t}_j\right\|$. Given (1), it can also be expressed as:

$$\sum_{i=1}^{\widetilde{N}}\boldsymbol{\beta}_i g\left(\boldsymbol{w}_i\cdot\boldsymbol{x}_j+b_i\right)=\boldsymbol{t}_j, \mathrm{j}=1,2,\cdots,N \tag{2}$$

Written in matrix form indicates:

$$\mathbf{H}\beta=\mathbf{T} \tag{3}$$

where

$$\mathbf{H}(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_{\widetilde{N}},b_1,\cdots,b_{\widetilde{N}},\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{\widetilde{N}})$$
$$=\begin{bmatrix} g(\boldsymbol{w}_1\cdot\boldsymbol{x}_1+b_1) & \cdots & g(\boldsymbol{w}_{\widetilde{N}}\cdot\boldsymbol{x}_1+b_1) \\ \vdots & \ddots & \vdots \\ g(\boldsymbol{w}_1\cdot\boldsymbol{x}_{\widetilde{N}}+b_1) & \cdots & g(\boldsymbol{w}_{\widetilde{N}}\cdot\boldsymbol{x}_N+b_1)\end{bmatrix}_{N\times\widetilde{N}}$$

$$\beta=[\beta_1^T,\beta_2^T,\cdots,\beta_{\widetilde{N}}^T]_{\widetilde{N}\times m}^T \text{ and } \mathbf{T}=[\mathbf{t}_1^T,\mathbf{t}_2^T,\cdots,\mathbf{t}_{\widetilde{N}}^T]_{\mathrm{N}\times m}^T. \tag{4}$$

H denotes the hidden layer output matrix; the ith column of **H** is the ith hidden node output with respect to inputs $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_{\widetilde{N}}$.

It has been proved by Huang G.B[8] in that if the activation function g is infinitely differentiable, then the input weight vector $\boldsymbol{w}$ and the hidden layer bias b can be randomly assigned. Thus solving the SLFN problem reduces to solve the output weight vector β, which can be accomplished by solving the Moore-Penrose of **H** from (3), noted as:

$$\beta=\mathbf{H}^{+}\mathbf{T} \tag{5}$$

where $\mathbf{H}^{+}$ is the Moore-Penrose of $\mathbf{H}$.

Thus, for a standard SLFN with infinitely differentiable activation function, the Extreme Learning Machine (ELM) algorithms is described as follows:

Algorithm ELM.

Input: training samples $\mathrm{X}=\{(\boldsymbol{x}_i,\boldsymbol{t}_i)|\boldsymbol{x}_i\in\boldsymbol{R}^n,\boldsymbol{t}_i\in\boldsymbol{R}^m,i=1,2,\cdots,\mathrm{N}\}$
Output: output weight β
Algorithm steps:
Step1: Random assign the input weight $\boldsymbol{w}$ and the hidden layer bias b
Step2: Calculate the hidden layer output matrix **H**
Step3: Calculate the output weight β

The Moore-Penrose of the output matrix **H** can always be obtained by singular value decomposition (SVD).

3 Lens Distortion in Camera Calibration

3.1 Linear Camera Model

Camera calibration means to calculate the intrinsic parameters and extrinsic parameters of the camera, in order to reconstruct any point in the camera image coordinate to the world 3D coordinate.

The theoretic camera model is the linear model, or pin-hole model, in which the image point and the world point comply with a linear transform. The pinhole camera model is based on the principle of collinearity, where each point in the object space is projected by a straight line through the projection center in to the image plane [12], as is shown in Figure 1.

For a point P in the world 3D coordinate $\mathrm{P}(X_W, Y_W, Z_W)$, it can be represented with a linear transform:

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \tag{6}$$

where s is the scale factor, [u v 1]' is the homogeneous coordinates of point P', the corresponding point of P in the image plane; f_x and f_y denotes the scale factors along the coordinates u and v in the image plane; (u_0, v_0) is the principle point P_o of the image plane; $\boldsymbol{R}$ and $\boldsymbol{T}$ are the rotation matrix and the translation matrix respectively, which represents the relationship between the image coordinate and the world coordinate.

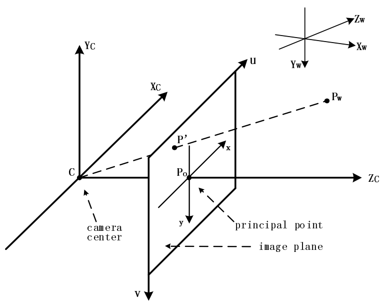


Fig. 1. Coordinates transform from image plane to 3D

(6) can also be expressed as:

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{M_1 M_2 X_W} \tag{7}$$

where $\mathbf{M}_1 = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$ is called the intrinsic matrix and $\mathbf{M}_2 = \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ \mathbf{0}^T & 1 \end{bmatrix}$ is called the extrinsic matrix. $\mathbf{X_W}$ denotes the homogeneous coordinates vector of 3D point $P(X_W, Y_W, Z_W)$.

3.2 Direct Linear Transformation and 3D Reconstruction with Binocular Vision

Direct Linear Transformation

Direct Linear Transformation (DLT) was proposed by Abdel-Aziz and Karara to calibrate the parameters of camera[1]. DLT establishes the geometric linear model of camera, which can be solved directly with linear equation system.

DLT is defined as (8):

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{K}(\mathbf{R}\ \mathbf{t}) \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} = \mathbf{P}_{3x4} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \tag{8}$$

Expand (8) and eliminate s presents:

$$\begin{aligned} p_{11}X_w + p_{12}Y_w + p_{13}Z_w + p_{14} - p_{31}uX_w - p_{32}uY_w - p_{33}uZ_w - p_{34}u = 0 \\ p_{21}X_w + p_{22}Y_w + p_{23}Z_w + p_{24} - p_{31}vX_w - p_{32}vY_w - p_{33}vZ_w - p_{34}v = 0 \end{aligned} \tag{9}$$

Suppose N 3D points and their corresponding image points are already known, then the linear equation system with 2*N equations:

$$\mathbf{AL} = \mathbf{0} \tag{10}$$

where $\mathbf{A}$ is a 2N*12 matrix, $\mathbf{L}$ is a vector consisting of elements from the perspective projection matrix $\mathbf{P}_{3x4}$. Thus, the camera calibration problem is reduced to solve a least square problem:

$$\begin{aligned} \boldsymbol{L_1} = (\boldsymbol{C}^T\boldsymbol{C})^{-1}\boldsymbol{C}^T\boldsymbol{B} \\ s.t. \quad p_{34} = 1 \end{aligned} \tag{11}$$

where $\boldsymbol{L_1}$ is a vector constructed by the first 11 elements of $\mathbf{L}$, and $\boldsymbol{C}$ is a matrix constructed by the first 11 columns of $\mathbf{A}$ and $\boldsymbol{B}$ is the 12th column of $\mathbf{A}$.

3D Reconstruction with Binocular Vision

Once the projection relationship is determined, with two cameras the world coordinate of a point can be reconstructed from two set of image coordinate gained from the images token by the two cameras.

3.3 Lens Distortion in Camera Calibration

The above section explains in detail the pin-hole model of camera. However, the pin-hole model is the ideal model for camera calibration since it assumes the projection from the world point to the image plane strictly follows the linear projection principle. In reality, yet, the linear projection principle does not usually hold because of the distortion of the camera.

The distortion of the camera is a phenomenon where straight lines can no longer remain straight after projection, and the further the line from the camera center, the more obvious the phenomenon is.

The most common camera distortions are the Barrel distortion, the Pincushion distortion and the Mustache distortion, and the mathematical model which describes the non-linear camera distortion is given by Atkison [14] and Weng [15]:

$$\bar{x} = x + \delta_x(x, y)$$
$$\bar{y} = y + \delta_y(x, y) \tag{12}$$

where $(\bar{x}, \bar{y})$ represents the ideal image coordinate of the world point P of the pin-hole model; (x, y) represents the actual image coordinate; δ_x and δ_y represents the non-linear distortion value, and can be given by:

$$\delta_x(x, y) = x(1 + k_1 r^2 + k_2 r^4 \cdots) + (p_1(2x^2 + r^2) + 2p_2 xy) + s_1 r^2$$
$$\delta_y(x, y) = y(1 + k_1 r^2 + k_2 r^4 \cdots) + (p_2(2x^2 + r^2) + 2p_1 xy) + s_2 r^2 \tag{13}$$

where $r^2 = x^2 + y^2$, and the first item is called the radial distortion, the second item decentering distortion and the third item thin prism distortion.

Although with (13) the non-linear projection problem can be alleviated to some extent, yet some issues still remain. For example, it is usually not clear to give the definite expression of the non-linear distortion given (13), since the types of distortion in which a camera can possess is beyond knowledge in advance. Moreover, even the types of distortion are confirmed, the degree of the radial distortion still needs further experiments. Tsai[3] points out that, to introduce too many non-linear arguments may cause the solution to be unstable rather than improve the accuracy. Thus the number of non-linear arguments can be a trade-off problem.

In the following section, a fast and explicit method of eliminating the non-linear distortion of camera using extreme learning machine, regardless of the types and arguments of distortion, is proposed, which proves to be simpler and with better performance through experiments.

4 Lens Distortion Correction Model using ELM

4.1 Model Structure

Recall in section 3.1, the purpose of camera calibration is to determine the intrinsic and the extrinsic parameters of the camera, and to amend non-linear lens distortion as well, in order to obtain correct projection relationship between the image points and the world 3D points; since only after the camera is correctly calibrated can it be used for 3D reconstruction. However, as discussed in section 3.3, the traditional ways to eliminate lens distortion bring about some other issues like the choice of the degree of the radial distortions and so on.

The above process may also be illustrated as the process to project the image points directly to the world 3D points, regardless of the intermediate process. The model contains a linear part and a non-linear part, which considers the calibration process as a black box. With this idea, the lens distortion correction model using ELM is proposed, and the algorithm of the model is detailed in section 4.2.

4.2 Algorithm Procedure

In order to obtain the world 3D coordinate, the algorithm is divided into two phases: the training phase and the testing phase. The training phase contains two stages, corresponding to the linear part and the non-linear part respectively. As discussed in chapter 3 and 4, the linear part is modeled based on the direct linear transformation, and the non-linear part is modeled based on the extreme learning machine. The algorithm procedure is detailed in Figure 2.

The training phase is composed with two stages: the linear stage and the non-linear stage, as is shown in Figure 2.

The linear stage takes the combination of image points coordinates from two cameras and the corresponding world points coordinates as the input training sample X_1. Note that the world coordinate (X_w, Y_w, Z_w) is pre-obtained and is regarded as the criteria for the expected output world coordinate. With DLT method, the perspective projection matrix P_1 and P_1 for the two cameras is calculated and then the training sample for 3D reconstruction can be constructed, whose world coordinate is marked as $(X^{(2)}, Y^{(2)}, Z^{(2)})$.

The non-linear stage utilizes the sample X_2 from linear stage and combines the corresponding world coordinate $(X_w^{(2)}, Y_w^{(2)}, Z_w^{(2)})$ as the training sample for the extreme learning machine, and at last the expected ELM output weight β.

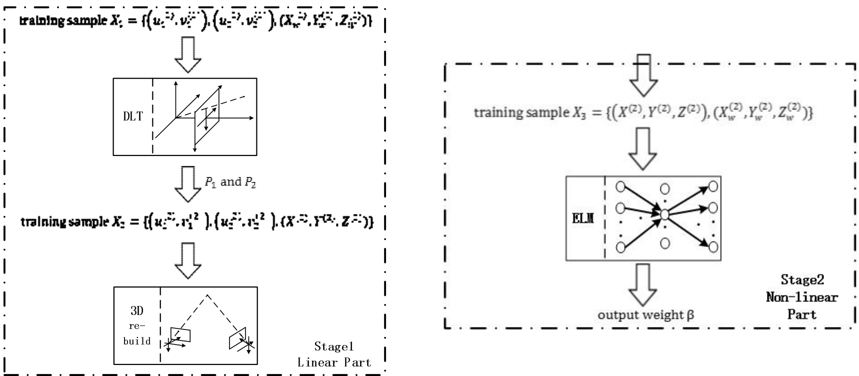


Fig. 2. Model Algorithm Procedure

5 Performance Evaluation

5.1 Dataset Description

In our experiment, the dataset is constructed manually with two cameras during the calibration and reconstruction procedure. Sixteen photos were taken by each camera, with ten gauge points on each picture. The 3D world coordinates were obtained by external calibration tool, and the world coordinates were determined with regard to a certain world point. Although the exact reference point cannot be obtained, the calibration and the reconstruction work can still be done without the knowledge of the point, as long as the relative relationship between the two cameras is known.

After eliminating the outlier points, a total dataset of 120 points is effective. Half of the dataset is used as training sample, which are selected randomly, leaving the rest as the testing sample.

5.2 Comparison Methods

The performance of the proposed ELM Lens Distortion Correction model is compared mainly with the popular algorithms for SLFN, namely the Back Propagation Network (BP network) and the Support Vector Machine (SVM). In the BP network, LM method is chosen and the activation function for hidden layer uses sigmoid function. In SVM, the simulation is carried out with LIBSVM [16], with RBF kernel function. In ELM, the All the three methods are carried out on a personal computer, which possesses a 16GB RAM, a 3.6GHz CPU. The simulation environment is MATLAB 2013b.

5.3 Performance Evaluation

MSE & Dev for Training & Testing

The number of hidden layer nodes is fixed at 40, since the size of the dataset is not large. Note the support vectors number for SVR is determined by LIBSVM tool. The number of the support vectors is 44, which also implies the rationality of the 40 nodes for ELM and BP. Table 3 shows the root mean square error and the standard deviations for the training phase and the testing phase of each algorithm. The performance is evaluated with the mean RMSE and DEV for X, Y and Z axis, and each algorithms is evaluated for 100 times and is averaged over the evaluated times.

From Table 1 it can be drawn that ELM generally performs better in terms of RMSE and DEV.

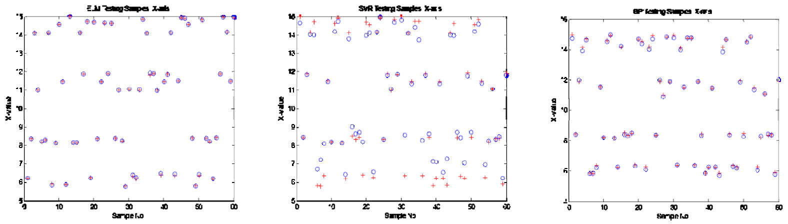
Table 1. Performance Comparison for ELM, BP and SVR

a. Performance for X-axis

Algorithms	Number of Nodes/SVs	Training		Testing	
		RMSE	Dev	RMSE	Dev
ELM	40	0.0128	0.0041	0.1252	0.1034
BP(LM)	40	0.0956	0.2359	0.2741	0.4515
SVR	44	0.1066	0.2810	0.2046	0.4253

Training & Testing Reconstruction Comparison

Figure 3 gives the reconstruction result figure for the above three algorithms. Note the red cross '+' stands for real outputs and the blue circle 'o' stands for the algorithms outputs, and only the testing samples of X-axis is presented in the figure. The results for Y and Z axis are similar with X axis, and are not shown for simplicity. From Figure 3, it can be drawn ELM performs better in reconstruction work than the other two algorithms.



ELM SVR BP

Fig. 3. X-axis testing reconstruction

Training & Testing Accuracy of ELM

Figure 4 presents the testing accuracy of ELM model with regard to the change of the hidden layer nodes. It can be concluded that the performance gets better when the number of hidden layer nodes increases.

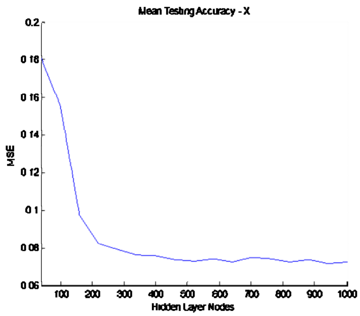


Fig. 4. ELM Testing Accuracy for X-axis

6 Conclusion

This paper proposes a novel method for camera calibration and lens distortion correction which employs the extreme learning machine. In the linear phase, the model uses DLT method to find the perspective projection relationship between the image coordinate and the world coordinate; in the non-linear phase, the model corrects the distortion by implementing ELM learning mechanism. The proposed model has been tested in real calibration work and the results show the validity of the model. Comparisons are made between the model and traditional regression models such as BP network and SVR. Yet future work on testing the robustness and fitness for big dataset is under investigation.

References

1. Abdel-Aziz, Y.I., Karara, H.M.: Direct Linear Transformation from Comparator Coordinates into Objectspace Coordinates in Close-range Photogrammetry. In: ASP Symp. on Close Range Photogrammetry. American Society of Photogrammetry, Falls Church (1971)
2. Tsai, R.Y.: An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition (1986)
3. Zhang, Z.: A Flexible New Technique for Camera Calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(11), 1330–1334 (2000)
4. Faugeras, O.D., Luong, Q.T., Maybank, S.J.: Camera Self-calibration: Theory and Experiments. In: Sandini, G. (ed.) ECCV 1992. LNCS, vol. 588, pp. 321–334. Springer, Heidelberg (1992)
5. Triggs, B.: Autocalibration and the Absolute Quadric. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 321–334 (1997)
6. Caprile, B., Torre, V.: Using Vanishing Points for Camera Calibration. International Journal of Computer Vision 4(2), 127–139 (1990)
7. Pollefeys, M., Koch, R., Van, G.L.: Self-calibration and Metric Reconstruction in spite of Varying and Unknown Intrinsic Camera Parameters. International Journal of Computer Vision 32(1), 7–25 (1999)
8. Huang, G.B., Zhu, Q.Y., Siew, C.K.: Extreme Learning Machine: Theory and Applications. Neurocomputing 70(1), 489–501 (2006)
9. Rong, H.J., Jia, Y.X., Zhao, G.S.: Aircraft recognition using modular extreme learning machine. Neurocomputing (2013)
10. Sun, Z.L., Choi, T.M., Au, K.F.: Sales Forecasting Using Extreme Learning Machine with Applications in Fashion Retailing. Decision Support Systems 46(1), 411–419 (2008)
11. Xu, Y., Dong, Z.Y., Meng, K.: Real-time Transient Stability Assessment Model Using Extreme Learning Machine. IET Generation, Transmission & Distribution 5(3), 314–322 (2011)
12. Heikkila, J., Silvén, O.: A Four-step Camera Calibration Procedure with Implicit Image Correction. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1106–1112. IEEE (1997)
13. Source from, `http://en.wikipedia.org/wiki/Distortion_(optics)`
14. Atkinson, K.B.: Developments in Close Range Photogrammetry. Elsevier Science & Technology (1980)
15. Weng, J., Cohen, P., Herniou, M.: Camera Calibration with Distortion Models and Accuracy Evaluation. IEEE Transactions on Pattern Analysis and Machine Intelligence 14(10), 965–980 (1992)
16. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST) 2(3), 27 (2011)