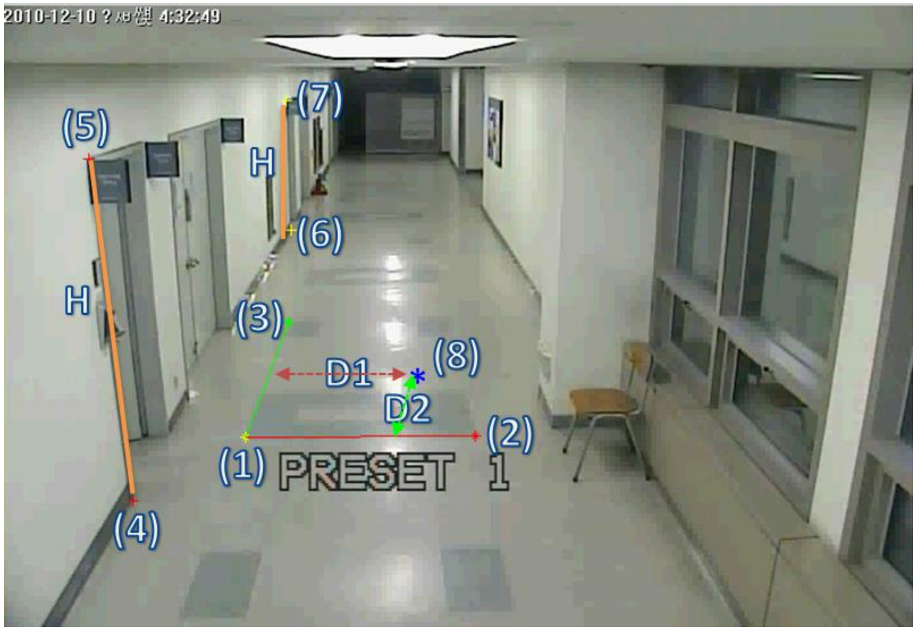
Part 1: Simplified camera calibration

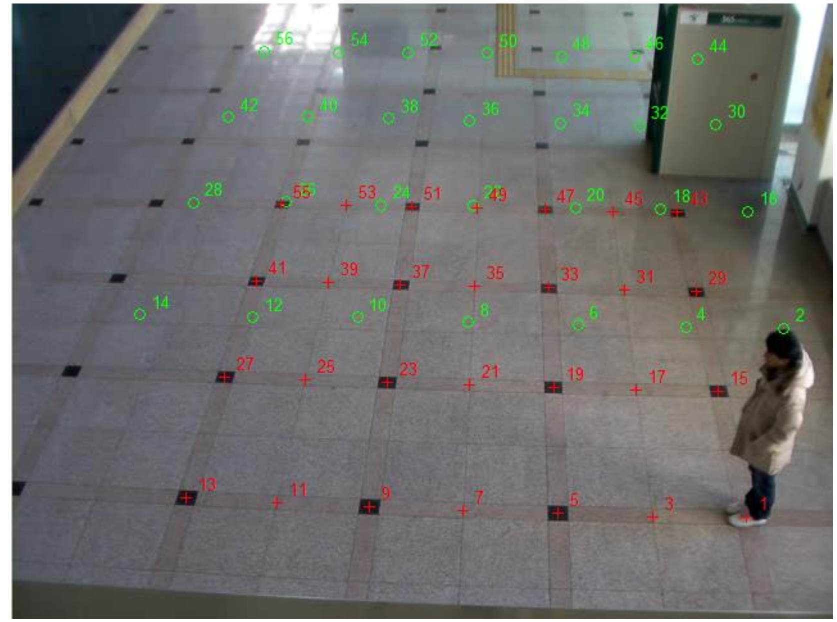
Shengzhe Li

CVLab, Inha Univ.

2014. 6.

Problem





Camera calibration experiments performed by Bin et al. at CVLab in 2011.
(Left) vanishing point based method. 8 points, height H, distance D1, D2.
(Right) DLT method. As many as possible points.

These method are accurate, but complicated

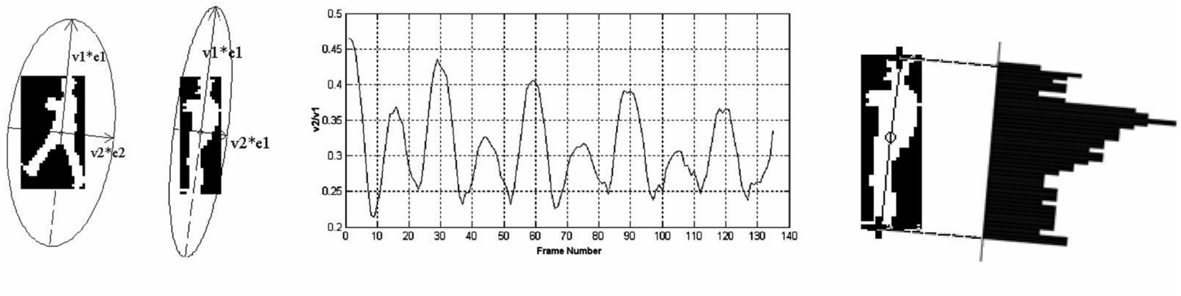
It is natural to associate **a walking or standing human** with the camera calibration problem in the context of video surveillance

Calibration by walking human

- Lv, F., Zhao, T., Nevatia, R.: Self-calibration of a camera from video of a walking human. In: International Conference on Pattern Recognition (ICPR). Volume 1. (2002) 562–567
- Lv, F., Zhao, T., Nevatia, R.: Camera calibration from video of a walking human. IEEE Trans Pattern Anal Mach Intell **28** (2006) 1513–1518
- Krahnstoever, N., Mendonca, P.R.: Bayesian autocalibration for surveillance. In: International Conference on Computer Vision (ICCV). (2005)
- Junejo, I., Foroosh, H.: Robust auto-calibration from pedestrians. In: IEEE International Conference on Video and Signal Based Surveillance (AVSS). (2006)
- Liu, J., Collins, R.T., Liu, Y.: Surveillance camera autocalibration based on pedestrian height distributions. In: British Machine Vision Conference, Dundee. (2011)
- Liu, J., Collins, R.T., Liu, Y.: Robust autocalibration for a surveillance camera network. In: IEEE Workshop on Applications of Computer Vision (WACV). (2013) 433–440

Lv et al. 2002, 2006

- Vanising point based; automatically extract head-foot points on leg crossing frames; nonlinear minimization algorithm to optimize parameters.



(a) (b) (c) (d)

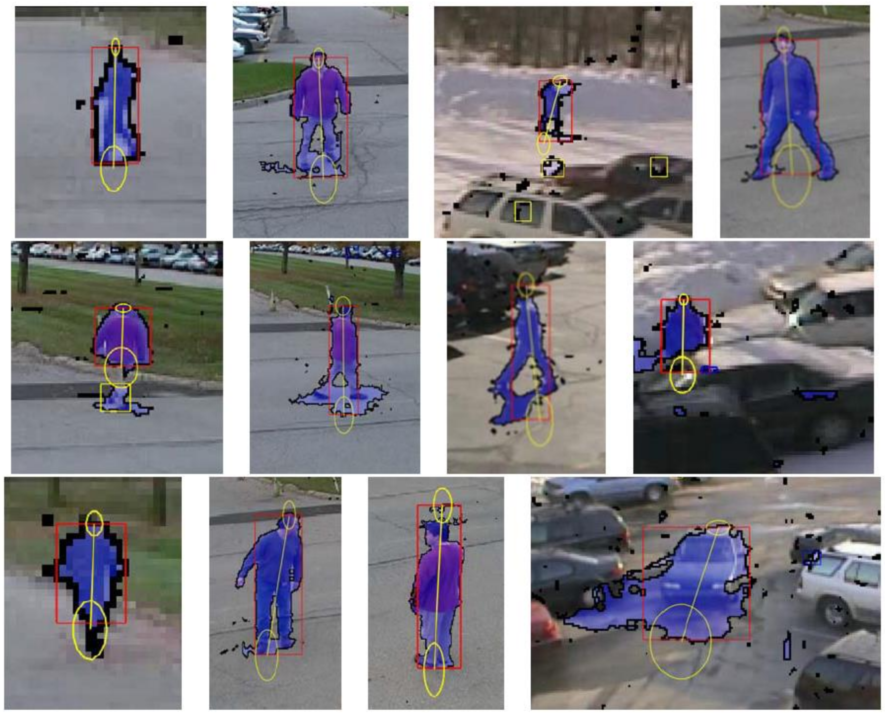
Extracting human leg-crossing phases from a video sequence. (a) and (b) Eigen analysis on the human shape at two different phases. (c) Plot of q_t over time. (d) Head and feet positions are located by finding two end points along the principal axis.

(a) (b) (c) (d) (e) (f)

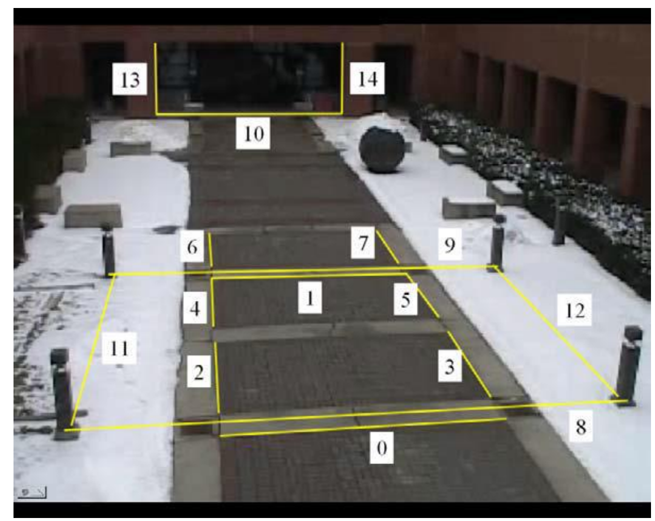
One detected leg-crossing frames of each scene.

Krahnstoever et al. 2005

- Vanishing point based; Homology estimation; Bayesian framework to optimize parameters;



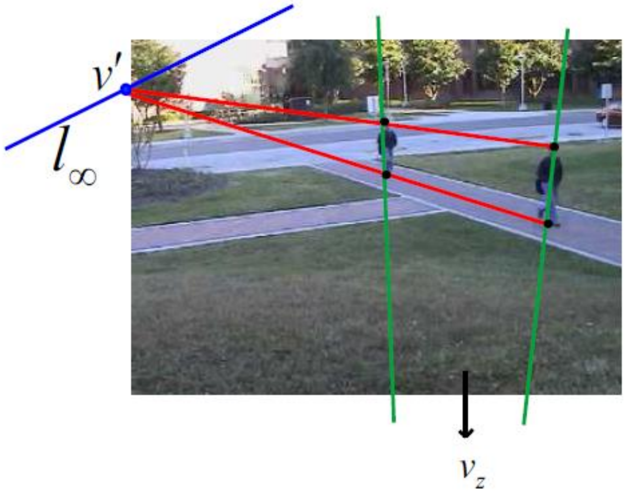
Foot and Head Location Estimation.



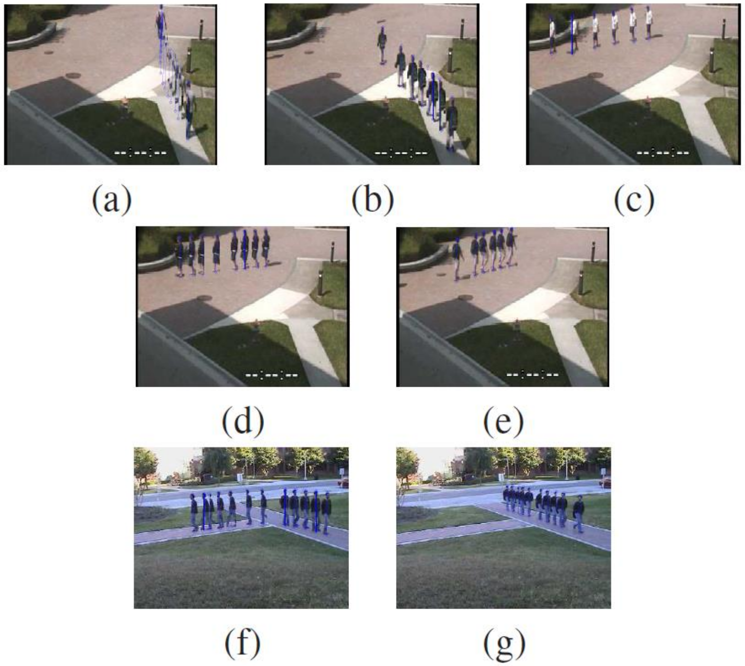
Ground Truth Distances.

Junejo et al. 2006

- Similar to Krahnstoever; the modification in the outlier removal stage



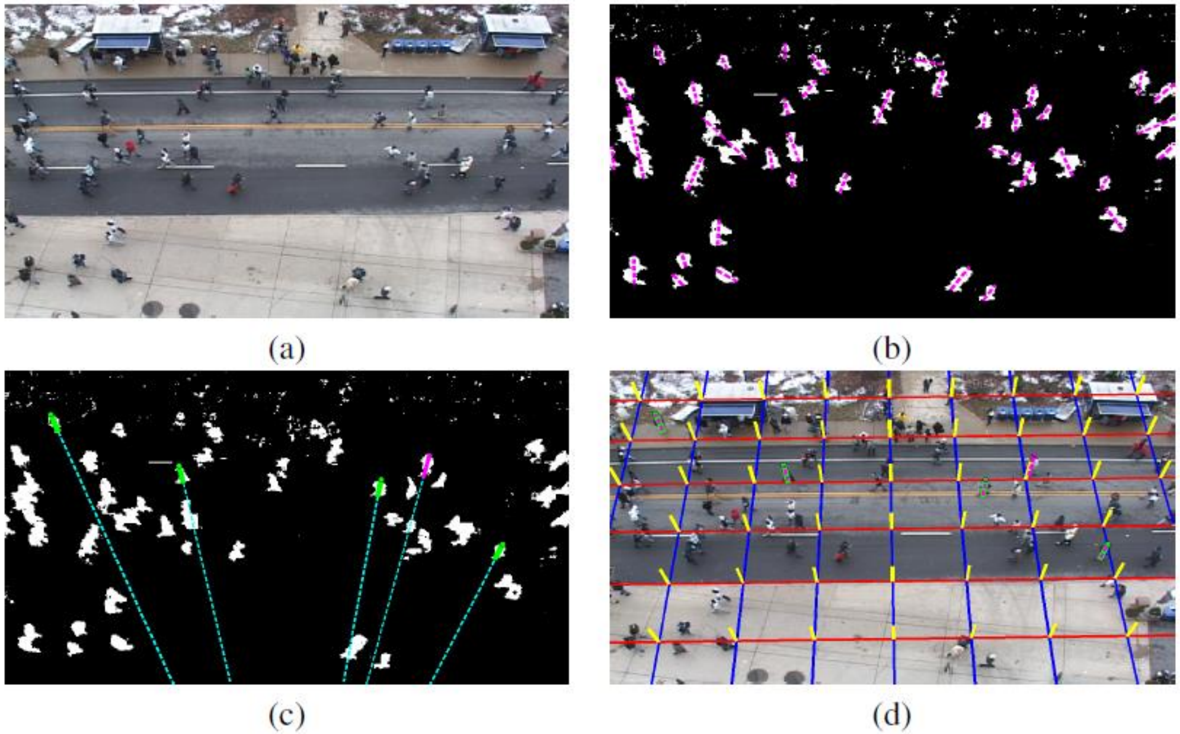
Auto-Calibration Geometry:A pedestrian, in two views, provides vertical vanishing points and an another vanishing points lying on the horizon line of the ground plane.



The figure depicts instances of the data set used for testing the proposed method.

Liu et al. 2011, 2013

- Automatically estimates using prior knowledge about the distribution of relative human heights



An example of surveillance camera calibration on Seq.1.
(a) one video frame; (b) foreground masks (c) after RANSAC vanishing point estimation and height distribution analysis (d) final calibration results

Papers on height estimation

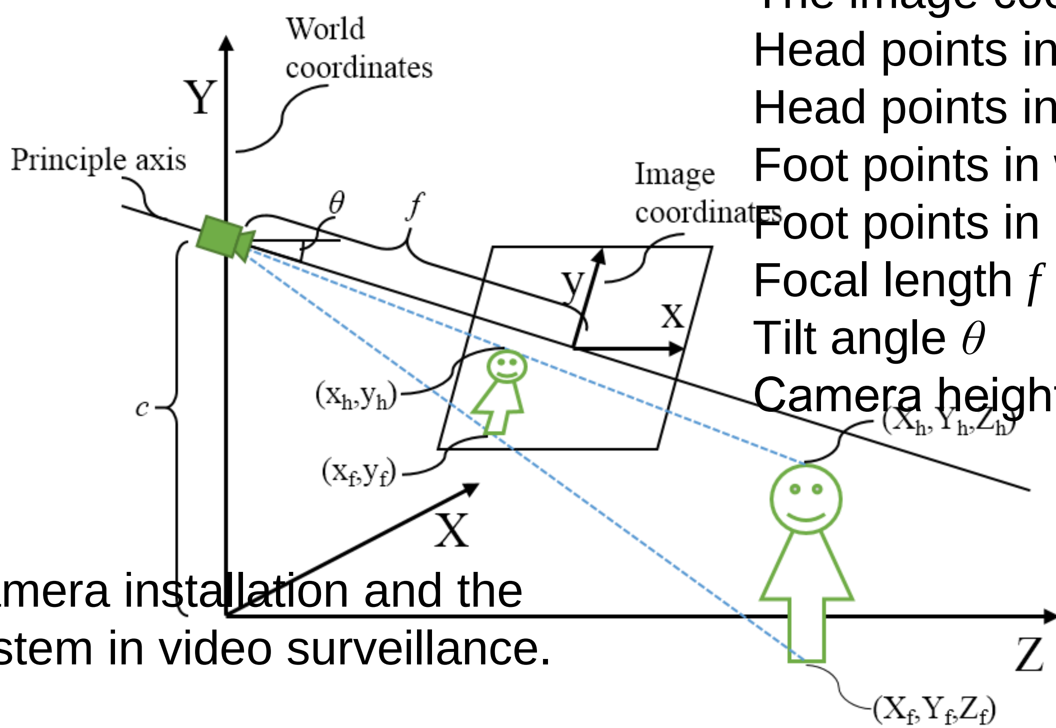
- Lee, K.Z.: A simple calibration approach to single view height estimation. In: IEEE Conference on Computer and Robot Vision (CRV). (2012) 161–166
- Gallagher, A.C., Blose, A.C., Chen, T.: Jointly estimating demographics and height with a calibrated camera. In: International Conference on Computer Vision (ICCV). (2009) 1187–1194
- Kispl, I., Jeges, E.: Human height estimation using a calibrated camera. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2008)

- DLT or Vanishing points method.

Main idea

- The reason why camera calibration is complicated is that there are too much calibration parameters (five intrinsic and six extrinsic parameters).
- Reducing the number of calibration parameters can simplify the problem.
- Considering that most cameras for video surveillance are installed in high positions with a slightly tilted angle
- It is possible to retain only **three calibration parameters** in the original camera model, namely the *focal length*, *tilting angle* and *camera height*.

Coordinate system, notations



The world coordinates $[X,Y,Z]^T$
The image coordinates $[x,y]^T$
Head points in world $[X_h,Y_h,Z_h]^T$
Head points in image $[x_h,y_h]^T$
Foot points in world $[X_f,Y_f,Z_f]^T$
Foot points in image$[x_f,y_f]^T$
Focal length f
Tilt angle θ
Camera height c

The typical camera installation and the coordinate system in video surveillance.

Simplified Calibration

Most cameras for video surveillance are installed in high positions with a slightly tilted angle. In such installation, the rotation angles along axis Y and Z can be assumed as 0 (which are also known as pan and roll), as well as the translations along axis X and Z. Therefore,

- $$P = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & c \\ 0 & 0 & 1 & 0 \end{bmatrix}$$
$$= \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f\cos\theta & -f\sin\theta & cf\cos\theta \\ 0 & \sin\theta & \cos\theta & c\sin\theta \end{bmatrix}$$

Simplified Calibration

These three parameters can determine the mapping from the world coordinates $[\mathrm{X},\mathrm{Y},\mathrm{Z}]^T$ to the image coordinates $[\mathrm{x},\mathrm{y},\mathrm{w}]^T$ as

$$\begin{bmatrix}\mathrm{x}\\ \mathrm{y}\\ \omega\end{bmatrix} = P\begin{bmatrix}\mathrm{X}\\ \mathrm{Y}\\ \mathrm{Z}\\ 1\end{bmatrix}$$

$$= \begin{bmatrix} f & 0 & 0 & 0\\ 0 & f\cos\theta & -f\sin\theta & cf\cos\theta\\ 0 & \sin\theta & \cos\theta & c\sin\theta \end{bmatrix}\begin{bmatrix}\mathrm{X}\\ \mathrm{Y}\\ \mathrm{Z}\\ 1\end{bmatrix}$$

$$= \begin{bmatrix} f\mathrm{X}\\ f\mathrm{Y}\cos\theta - f\mathrm{Z}\sin\theta + cf\cos\theta\\ \mathrm{Y}\sin\theta + \mathrm{Z}\cos\theta + c\sin\theta \end{bmatrix}$$

which can be represented in Cartesian coordinates as

$$\begin{bmatrix}\mathrm{x}\\ \mathrm{y}\end{bmatrix} = \begin{bmatrix} \dfrac{f\mathrm{X}}{\mathrm{Y}\sin\theta + \mathrm{Z}\cos\theta + c\sin\theta}\\ \dfrac{f\mathrm{Y}\cos\theta - f\mathrm{Z}\sin\theta + cf\cos\theta}{\mathrm{Y}\sin\theta + \mathrm{Z}\cos\theta + c\sin\theta} \end{bmatrix}$$

Simplified Calibration

A basic relationship between the world coordinates Y, Z and the image coordinates y, which is given as

$$\mathrm{y} = \frac{f\mathrm{Y}\cos\theta - f\mathrm{Z}\sin\theta + cf\cos\theta}{\mathrm{Y}\sin\theta + \mathrm{Z}\cos\theta + c\sin\theta}$$

$$= \frac{f\mathrm{Y} - f\mathrm{Z}\tan\theta + cf}{\mathrm{Y}\tan\theta + \mathrm{Z} + c\tan\theta}.$$

Since each pair of the head and foot of the y coordinates, denoted as $\mathrm{y_h}$ and $\mathrm{y_f}$, can be measured from the image. By above Eq., a set of equations with three unknowns can be built as

$$\begin{cases} \mathrm{y_f} = \frac{-f\mathrm{Z}\tan\theta + cf}{\mathrm{Z} + c\tan\theta} \\ \mathrm{y_h} = \frac{f\mathrm{Y_h} - f\mathrm{Z}\tan\theta + cf}{\mathrm{Y_h}\tan\theta + \mathrm{Z} + c\tan\theta} \end{cases}.$$

Eliminating Z,

$$\hat{\mathrm{y}}_\mathrm{h} = \frac{f\left(c\tan^2\theta + \mathrm{Y_h} + c\right)\mathrm{y_f} + f^2\tan\theta\mathrm{Y_h}}{\tan\theta\mathrm{Y_h}\mathrm{y_f} + f\left(\tan^2\theta\mathrm{Y_h} + c\tan^2\theta + c\right)}.$$

Simplified Calibration

The parameters can be found by the nonlinear regression as

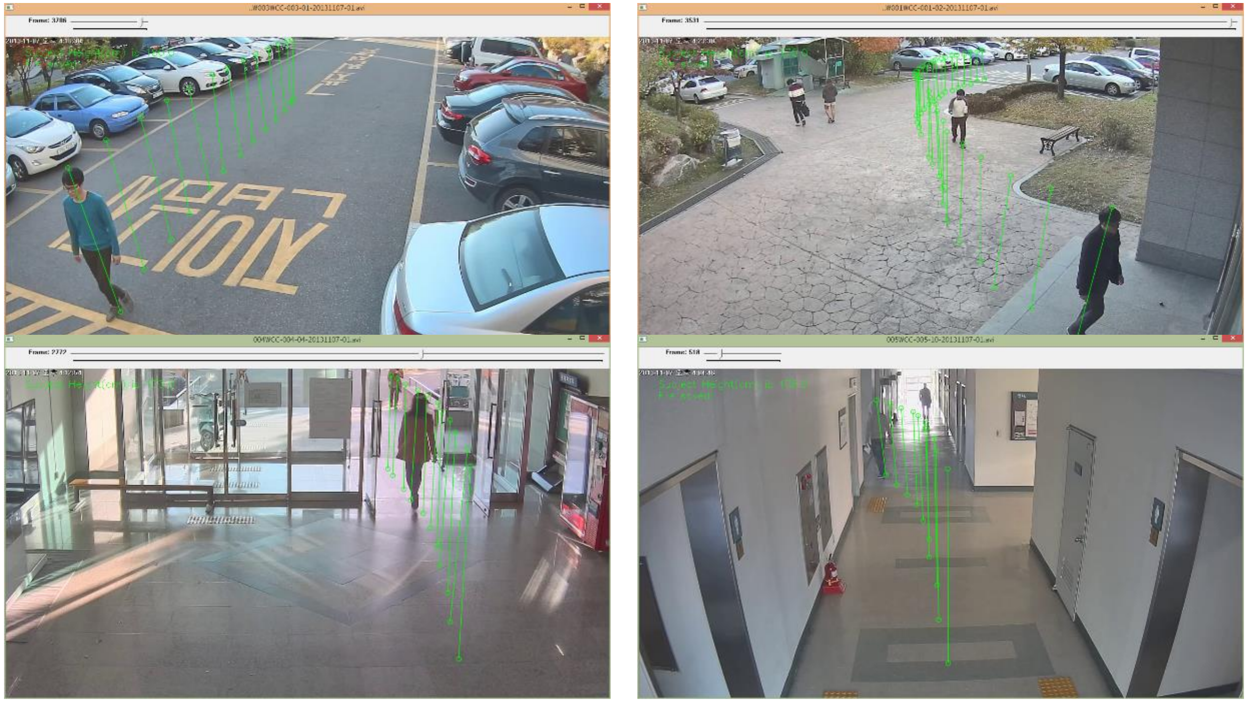
$$\begin{bmatrix} f \\ \theta \\ c \end{bmatrix} = \underset{f,\theta,c_y}{\operatorname{argmin}} \sum_i (\hat{y}_h - y_h)^2.$$

Once the calibration parameters of a camera are obtained, the physical height of a person can be estimated from a pair of head and foot points observed from the image.

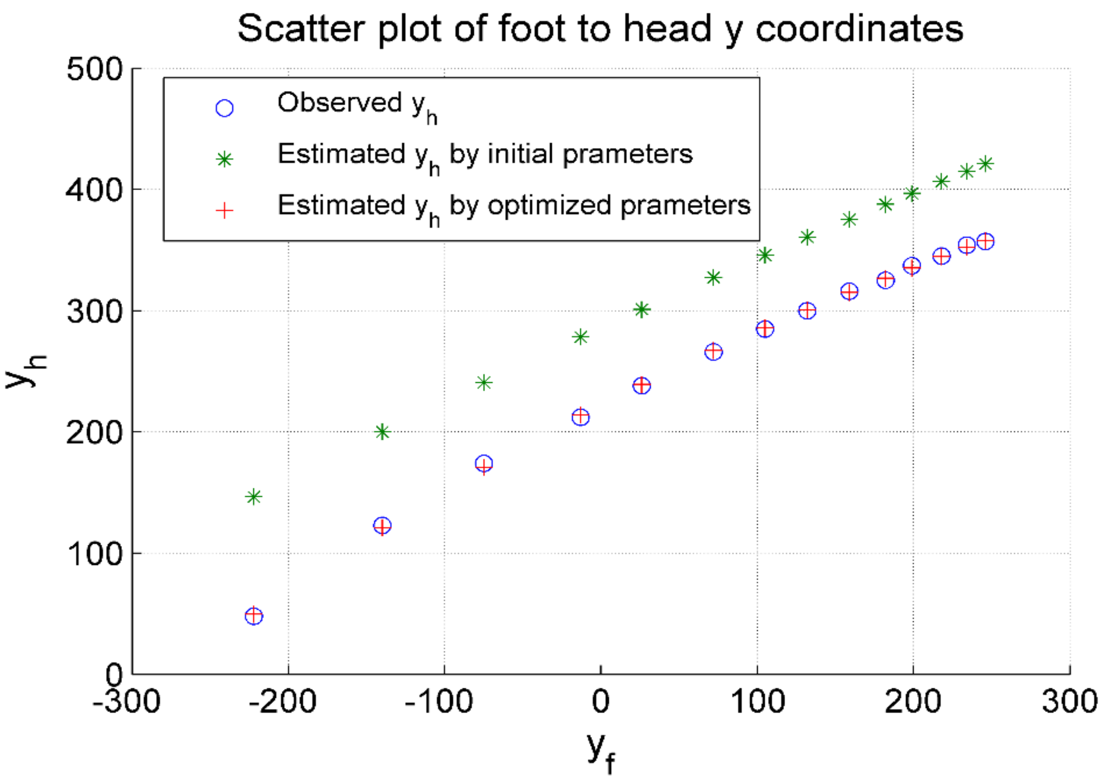
$$\hat{\mathrm{Y}}_{\mathrm{h}} = \frac{fc\left(\tan^2\theta + 1\right)(\mathrm{y_f} - \mathrm{y}_h)}{\tan\theta\,\mathrm{y_h}\mathrm{y_f} - f\mathrm{y_f} + f\tan^2\theta\,\mathrm{y_h} - f^2\tan\theta}.$$

Dataset for evaluation

- Number of subjects: 6
- Number of cameras: 5
- Video resolution: 1280 × 720
- Location: Inha Univ. Hitech Bldg.
- Mark: Manual



Calibration Result



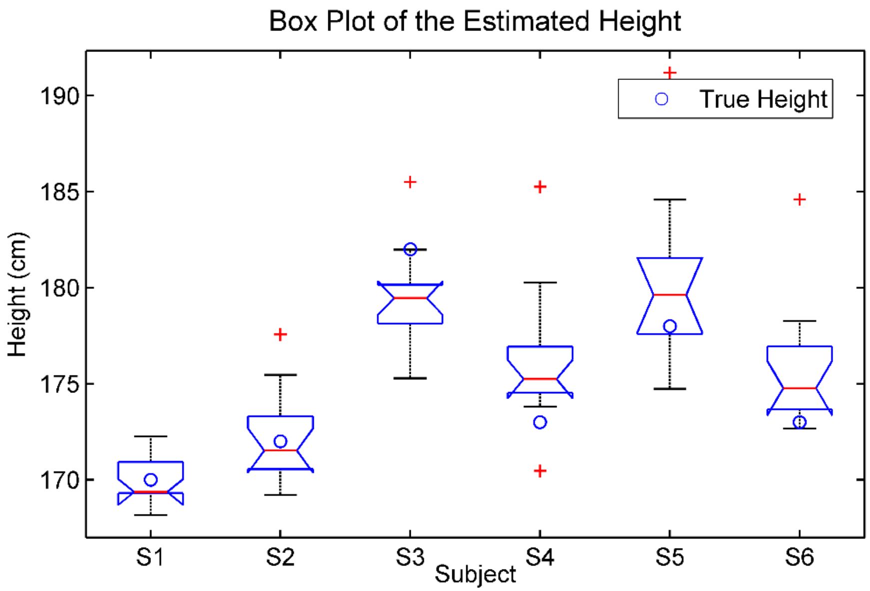
- A scatter plot of the y coordinates of the observed and estimated head points with respect to the observed foot points. The initial parameters f = 720, θ = -30 and c = -300 are approximated via visual estimation and the optimal parameters are found as f = 547.7, θ = -38.6 and c = -270.2 by the nonlinear regression method.

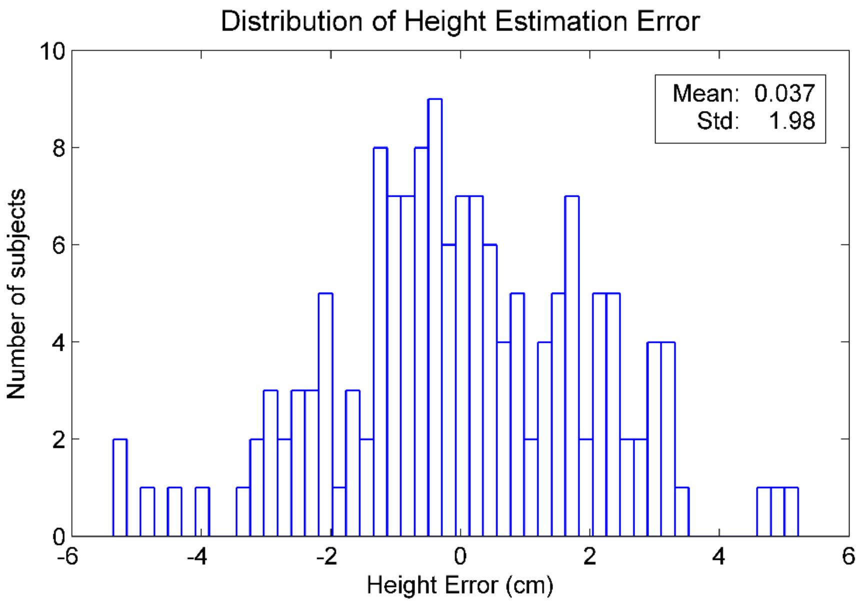
Height Estimation Results

		Test					
Train		S1	S2	S3	S4	S5	S6
	S1	170(0)	171.7(0.3)	181.7(1.6)	173.7(1.9)	179.8(1.1)	174.8(0.9)
	S2	170.2(0.4)	172(0.1)	182(1.7)	173.9(1.7)	180.2(0.8)	175(1)
	S3	170.6(1.4)	172.4(1.6)	182(0)	174.1(2.6)	180(2.3)	175.2(1.8)
	S4	169.2(1.5)	170.9(1.5)	181.3(3)	173(0)	179.2(1.8)	174.2(1.6)
	S5	168.1(1.5)	169.9(1.1)	180(2.3)	171.8(1.3)	178(0)	172.9(0.9)
	S6	168.3(1.2)	170(1)	180.1(1.8)	171.9(1.5)	178.1(1.2)	173(0)
True		170	172	182	173	178	173

Cross Validation of the height estimation. The number in the parenthesis indicates the standard deviation of estimated heights from the six cameras. The true heights are given in the last row.

Height Estimation Results





Comparison

	Calibration object	Method	Mean Abs. Error	Std. of Error	Maximum Error
N. Krahnstoever	Walking human	Automatic; Vanishing point	5.80%		
K. Z. Lee	Cubix box or line	Manual; Vanishing point			5.50%
A. C. Gallager	Grid pattern	Manual; Zhang's method		2.67cm	3.28cm
Kispal				3.1cm	5.5cm
Proposed	Walking human	Manual; Non-linear regression	1.55cm (0.8%)	1.98cm (1.1%)	5.36cm (3.0%)

Comparison of the proposed method with the existing height estimation methods

Conclusion

- The proposed method requires neither any special calibration object nor a special pattern on the ground, such as parallel or perpendicular lines; the proposed method does not rely on computing the vanishing points, which is difficult to estimate in practice
- The cross validation results show that an mean absolute error is only about 1.55cm from the ground-truth data.
- The proposed method can be integrated with automated human detection methods to fully perform autocalibration. This remains as a future study. Another future work can be introducing the parameter of lens distortion into the simplified camera model, which is not considered in this paper.

Part 2: Simplified Camera calibration with distortion correction

Shengzhe Li
CVLab, Inha Univ.
2014. 11.

New problem 1

GT : 170cm

New problem 2

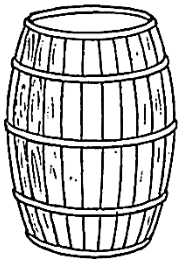
GT : 396cm
AR : 322cm

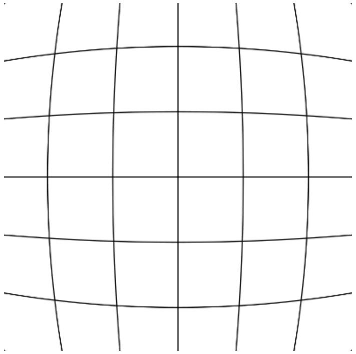
GT : 271cm
AR : 278cm

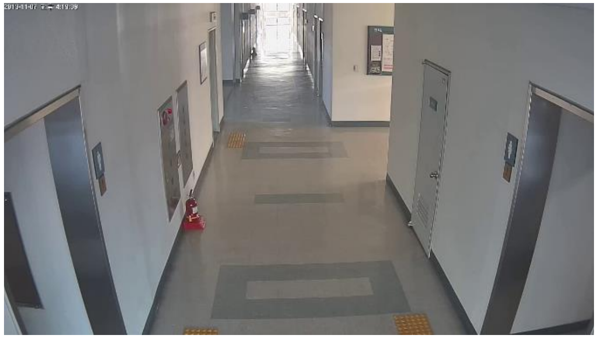
The 4th order distortion model

- The types of distortion (from Wikipedia)

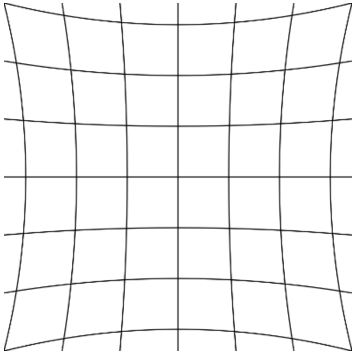
Barrel distortion







Pincushion distortion



The 4th order distortion model

Camera distortion can usually be expressed as

$$\begin{cases} x_d = x_u\left(1 + k_{d1} \cdot r_u^2 + k_{d2} \cdot r_u^4\right) \\ y_d = y_u\left(1 + k_{d1} \cdot r_u^2 + k_{d2} \cdot r_u^4\right) \end{cases}$$

where x_u, y_u are undistorted(ideal) coordinates, x_d, y_d are distorted coordinates(real) and r_u is the radius. k_{d1} and k_{d2} are the distortion parameters.

The inverse of camera distortion model has same form but different coefficients

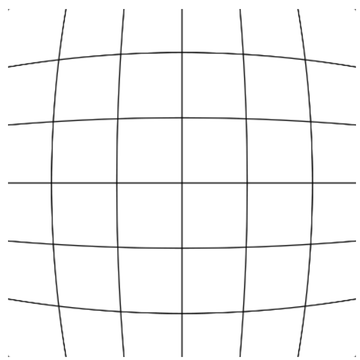
$$\begin{cases} x_u = x_d\left(1 + k_{u1} \cdot r_d^2 + k_{u2} \cdot r_d^4\right) \\ y_u = y_d\left(1 + k_{u1} \cdot r_d^2 + k_{u2} \cdot r_d^4\right) \end{cases}.$$

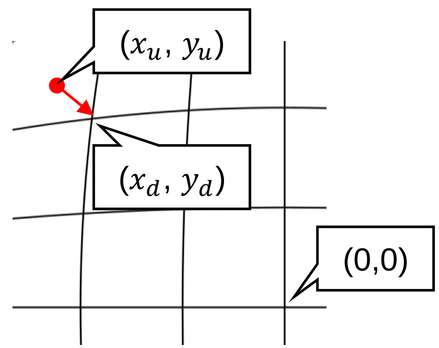
The 4th order distortion model

The type of distortion depends on the sign of k_{d1}.

$$\begin{cases} x_d = x_u\left(1 + k_{d1}\cdot r_u^2 + k_{d2}\cdot r_u^4\right) \\ y_d = y_u\left(1 + k_{d1}\cdot r_u^2 + k_{d2}\cdot r_u^4\right) \end{cases}$$

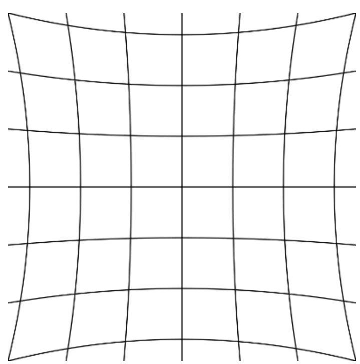
Barrel distortion





$k_{d1} < 0$

Pincushion distortion



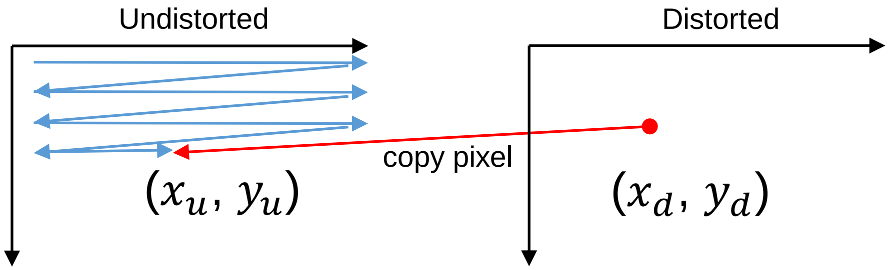
…

$k_{d1} > 0$

The use of distortion model

- Correct image distortion
  - Given a distorted image how to get the undistorted image

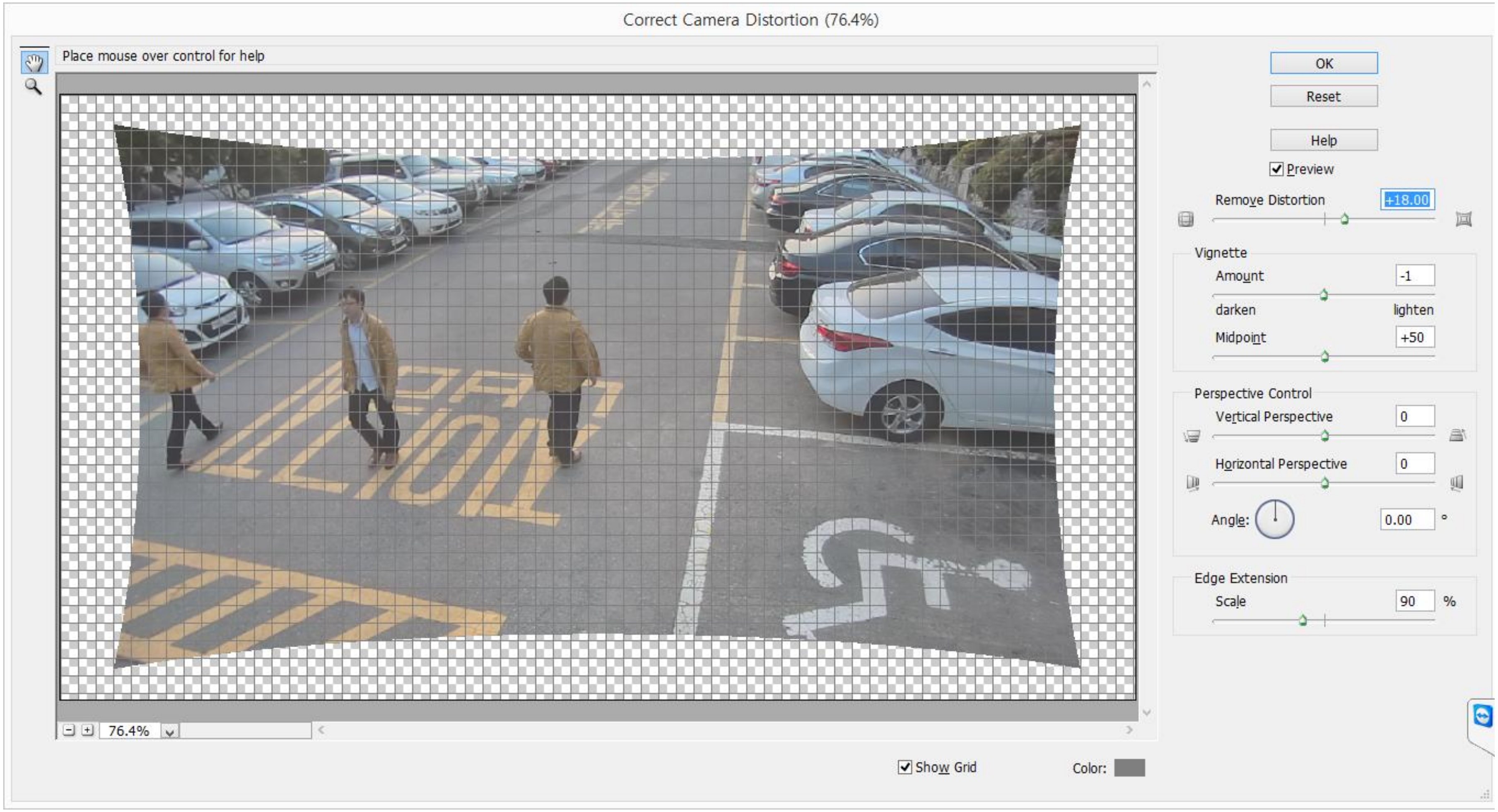
$$\begin{cases} x_d = x_u\left(1 + k_{d1} \cdot r_u^2 + k_{d2} \cdot r_u^4\right) \\ y_d = y_u\left(1 + k_{d1} \cdot r_u^2 + k_{d2} \cdot r_u^4\right) \end{cases}$$



- Correct coordinates distortion
  - Given a pixel's coordinates in distorted image how to get the undistorted coordinates

$$\begin{cases} x_u = x_d\left(1 + k_{u1} \cdot r_d^2 + k_{u2} \cdot r_d^4\right) \\ y_u = y_d\left(1 + k_{u1} \cdot r_d^2 + k_{u2} \cdot r_d^4\right) \end{cases}$$

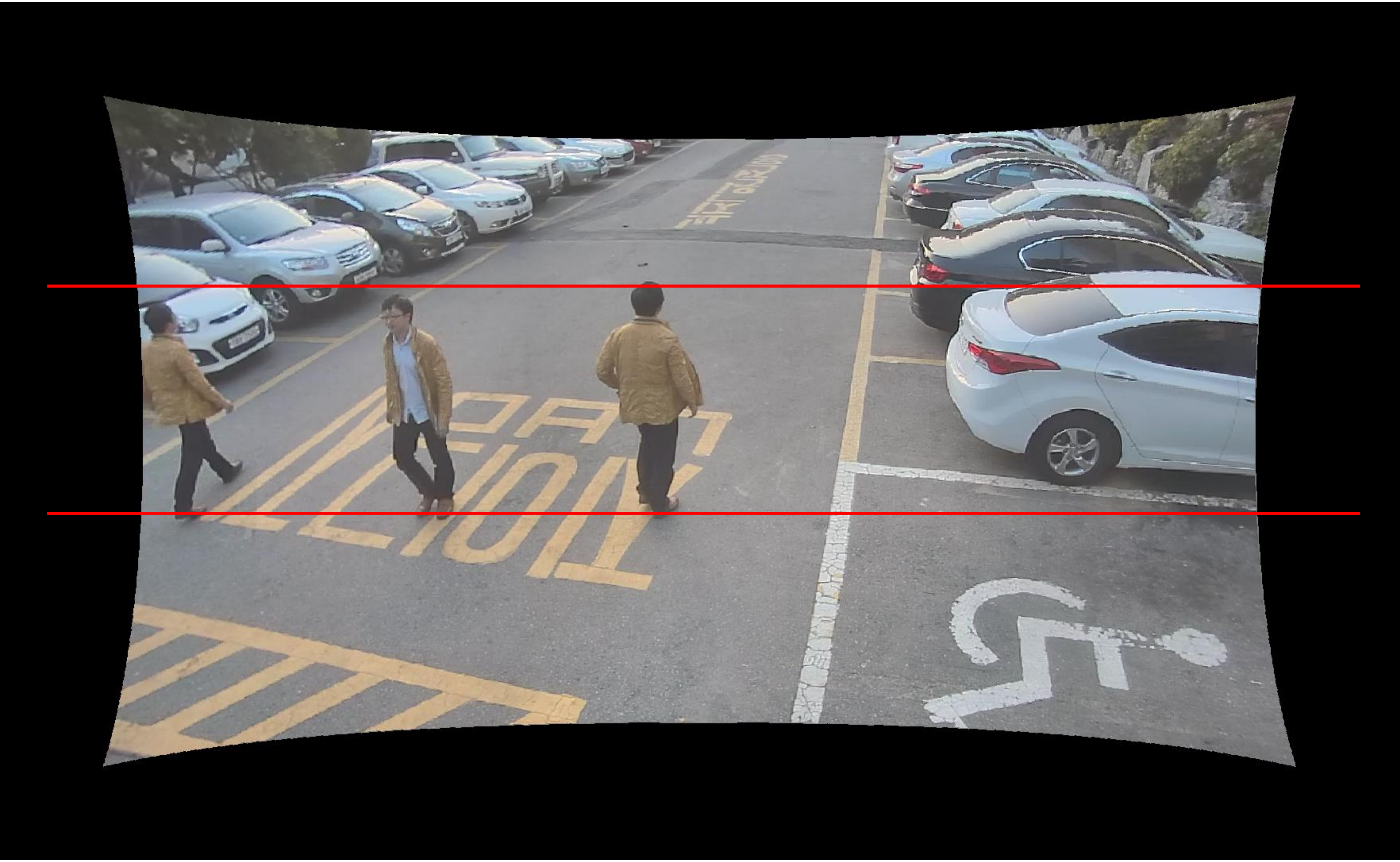
Correct image distortion in PS



Distort

Undistort k_{d1}=-0.2

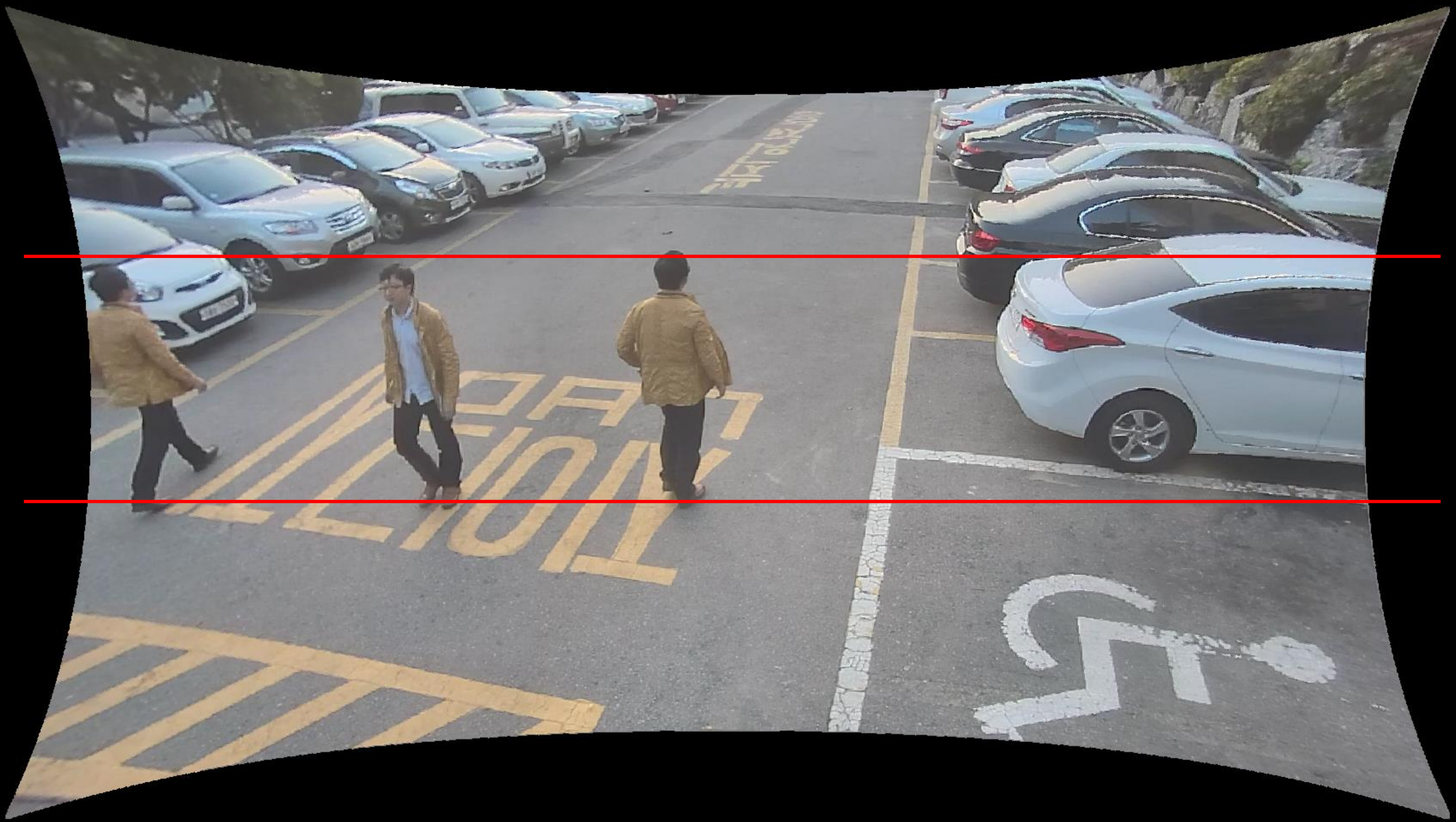
Undistort k_{d1}=-0.4



Undistort k_{d1}=-0.4 k_{d2}=0.2

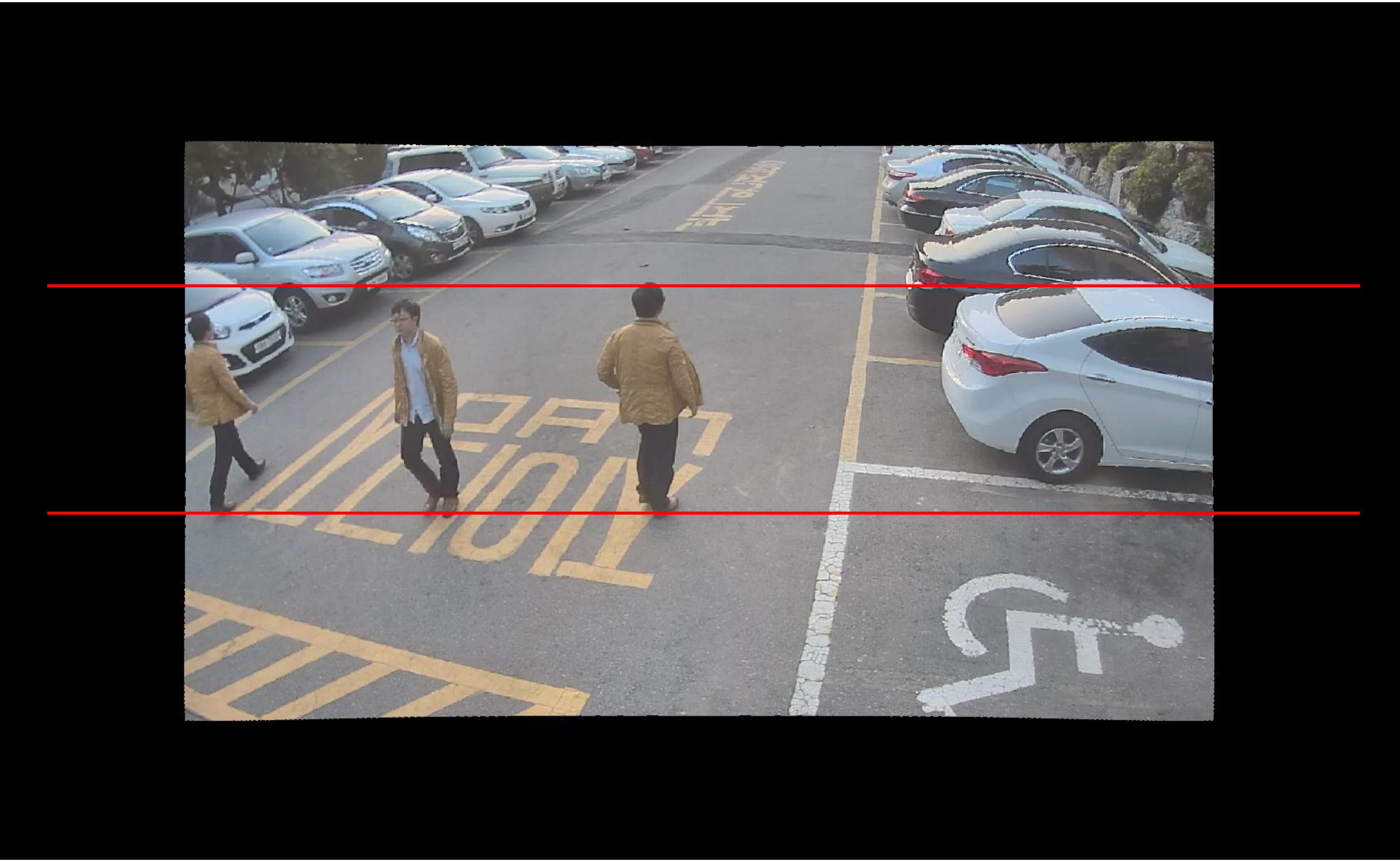
Undistort k_{d1}=-0.6

Undistort k_{d1}=-0.6 k_{d2}=0.2



Undistort k_{d1}=-0.6 k_{d2}=0.4

Undistort k_{d1}=-0.6 k_{d2}=0.4, Distort k_{u1}=0.6



Correct coordinates distortion

- To correct coordinates in the distorted image, we need k_u.

$$\begin{cases} x_u = x_d\left(1 + k_{u1} \cdot r_d^2 + k_{u2} \cdot r_d^4\right) \\ y_u = y_d\left(1 + k_{u1} \cdot r_d^2 + k_{u2} \cdot r_d^4\right) \end{cases}$$

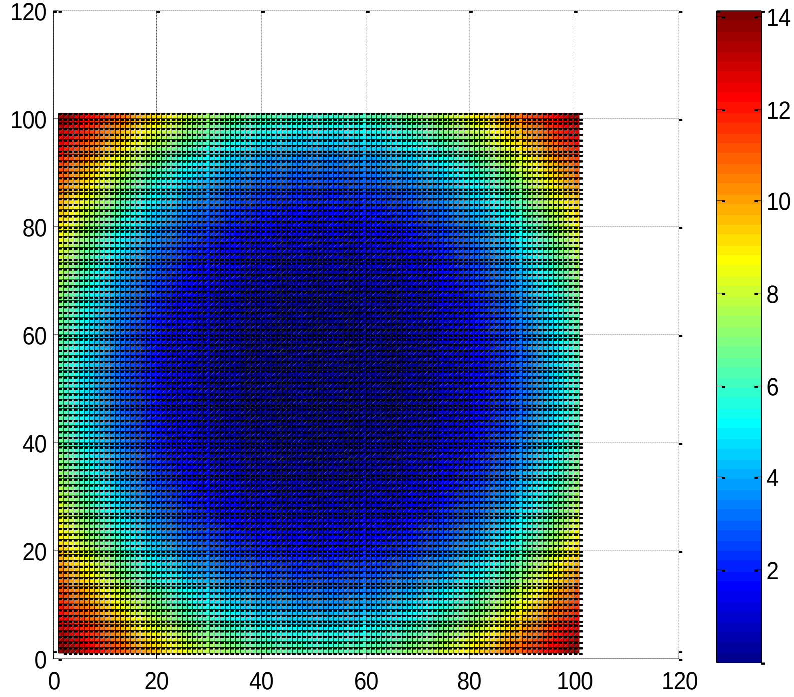
- Usually k_d can be obtained easily.
  - manual adjustment (in PS)
  - automatic method might be possible

- For each k_d there should be a corresponding k_u.
  - correct image distortion for each frame (very time consuming)
  - can be solved using the Cardan method [Devernay2001]
  - **numeric estimation**

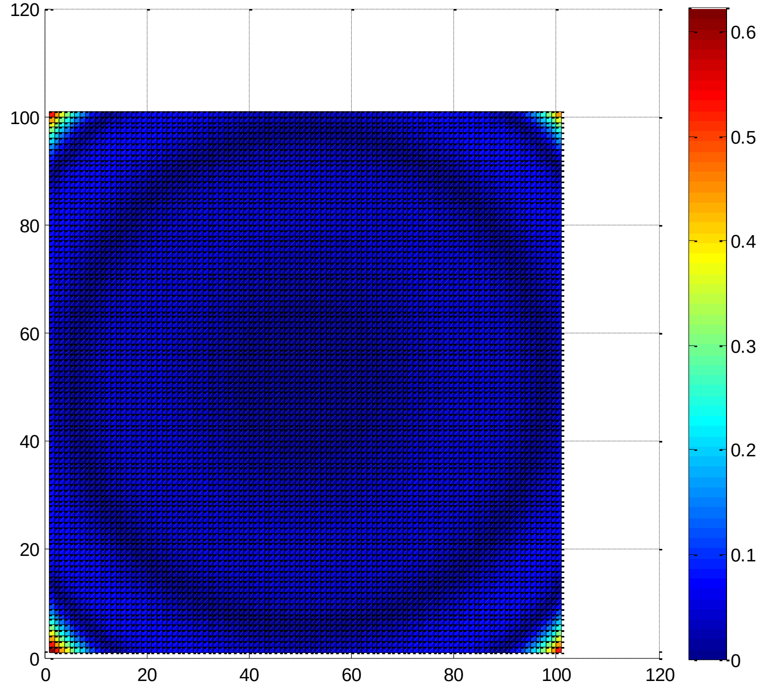
Numeric estimation of k_u

- Steps for numeric estimation of k_u
  1. Generate a undistorted coordinates $X_u = [x_u, y_u]$.
  2. Distort X_u to X_d by given k_u.
  3. Find k_d that minimize $\|\hat{X}_u - X_u\|$.

Coordinate error between X_u and X_d



Coordinate error between X_u and $\hat{X}_u$



Experiments

- After obtaining k_u, the points for calibration are converted to undistorted coordinates.
- Estimate the calibration parameters using corrected points by the nonlinear regression as

$$\begin{bmatrix} f \\ \theta \\ c \end{bmatrix} = \underset{f,\theta,c_y}{\operatorname{argmin}} \sum_i (\hat{y}_h - y_h)^2.$$

- Evaluate height estimation error and compare the results with previous experiment.

With distortion correction

GT : 170cm

With distortion correction

GT : 396cm
AR : 357cm

GT : 271cm
AR : 278cm

Without distortion correction (1st version Y only)

meanerror =

3.7522e-04

stderror =

0.0198

maxerror =

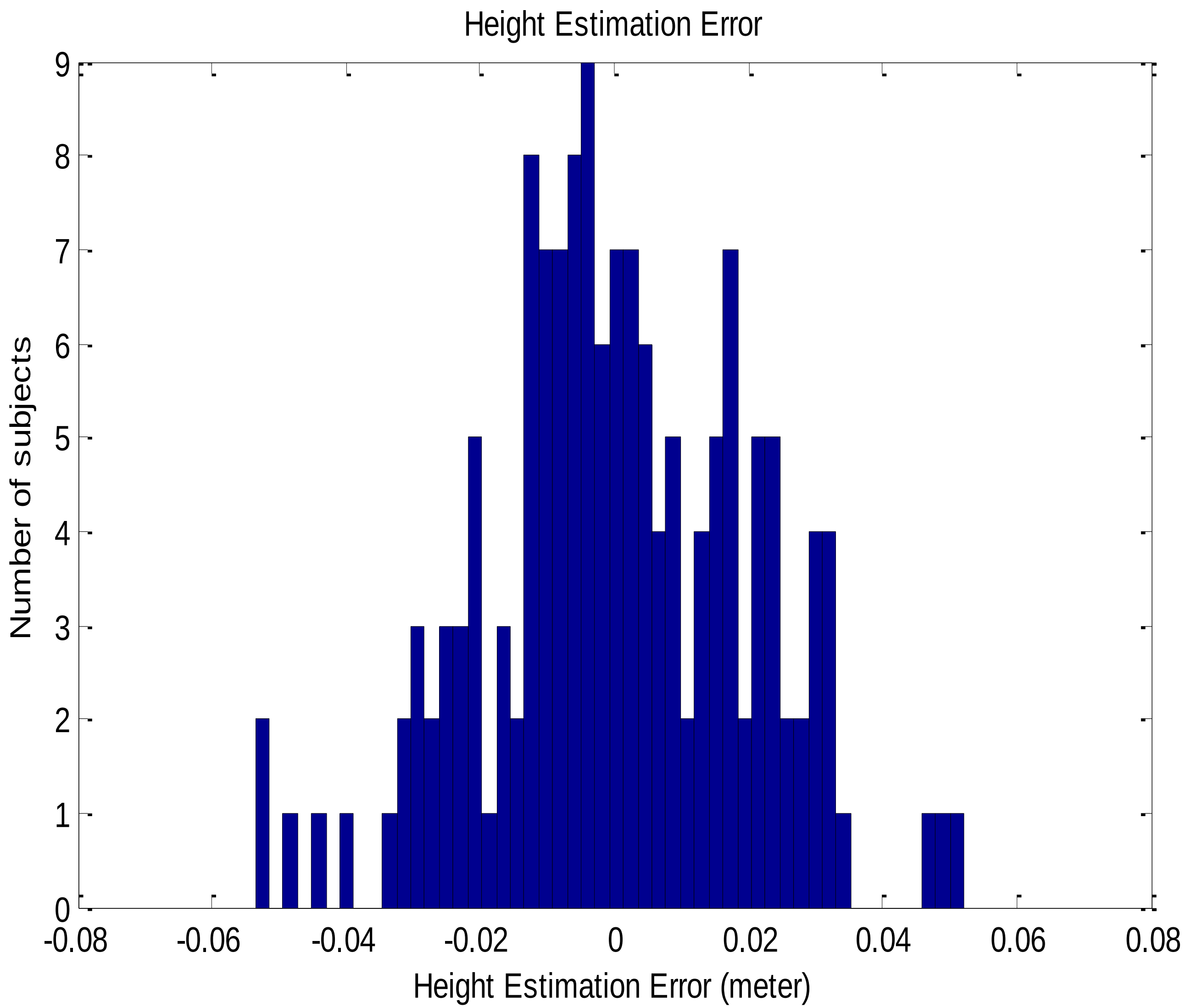
0.0522

minerror =

-0.0536

meanabserror =

0.0156



With distortion correction (1st version Y only)

meanerror =

1.9802e-04

stderror =

0.0187

maxerror =

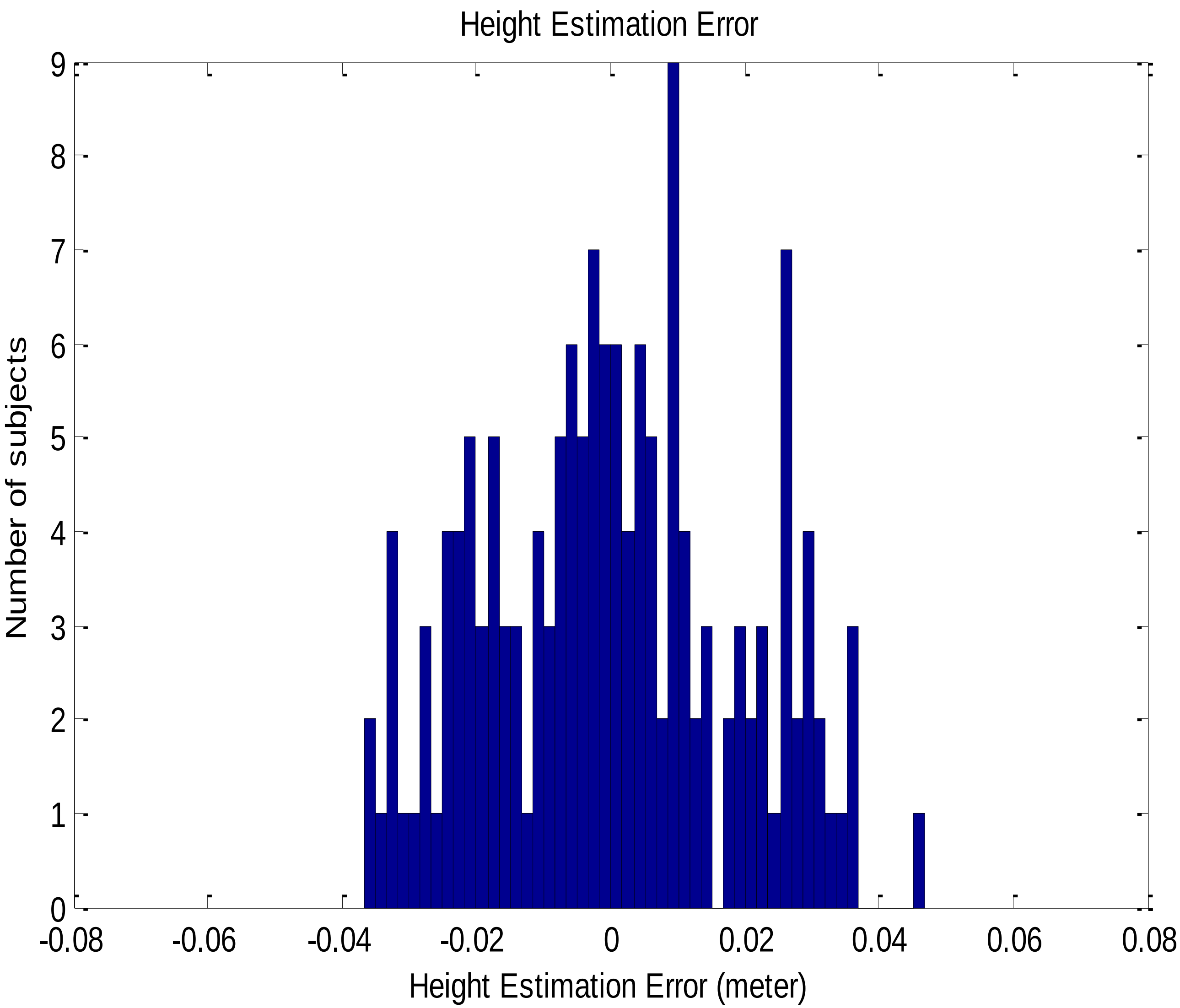
0.0470

minerror =

-0.0367

meanabserror =

0.0152



Without distortion correction (2^{nd} version XY)

meanerror =

-0.0081

stderror =

0.0263

maxerror =

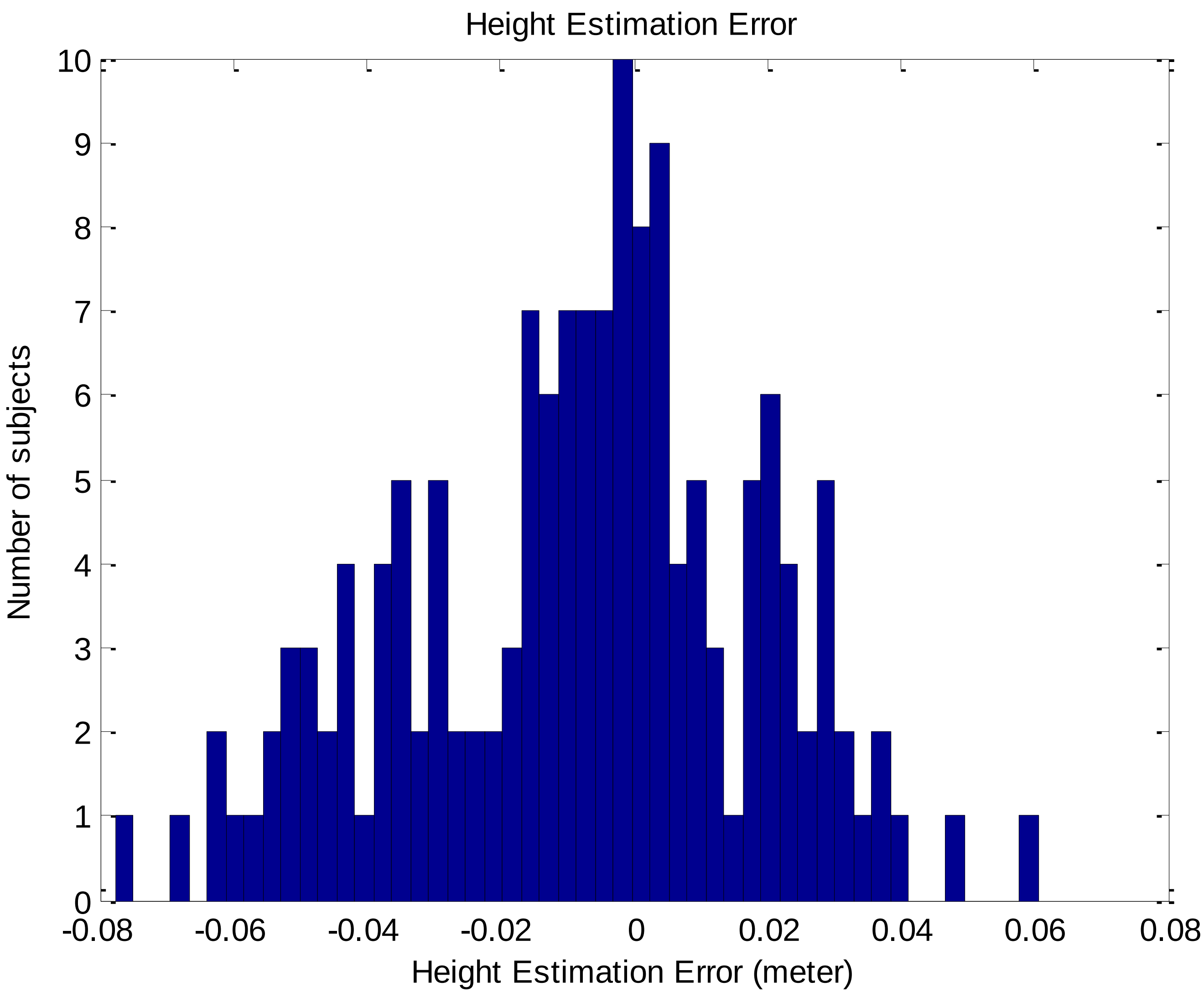
0.0605

minerror =

-0.0777

meanabserror =

0.0212



With distortion correction (2nd version XY)

meanerror =

-0.0059

stderror =

0.0230

maxerror =

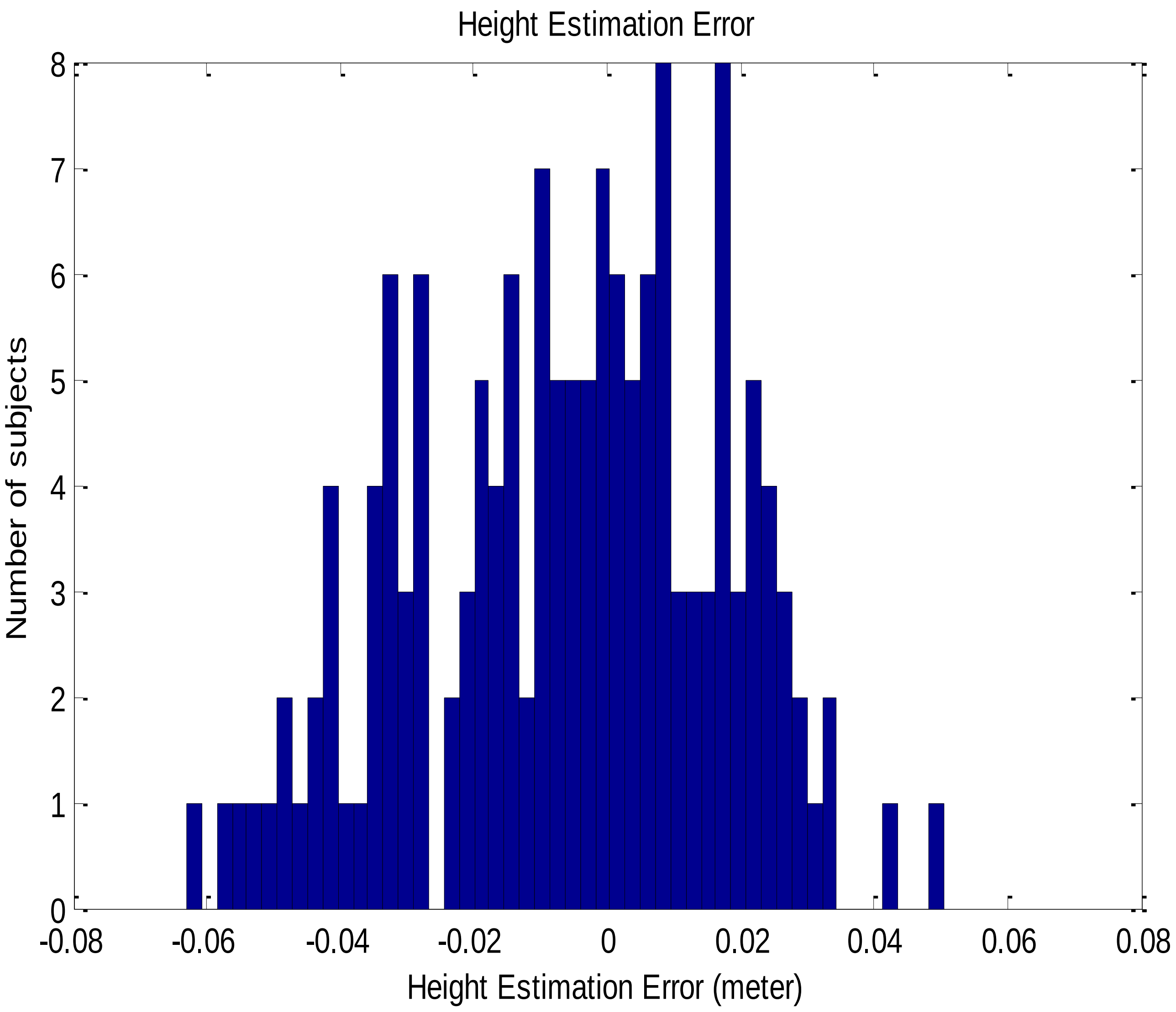
0.0503

minerror =

-0.0630

meanabserror =

0.0189



Conclusion

- Camera distortion correction offers more accurate height estimation especially for boarder area as well as the length in the floor.