Camera Self-Calibration from Tracking of Moving Persons

ICPR 2016 Oral Session MoAT3

Zheng Tang [1], Yen-Shuo Lin [2], Kuan-Hui Lee [1], Jenq-Neng Hwang [1], Jen-Hui Chuang [2], Zhijun Fang [3]

[1] University of Washington, United States;

[2] National Chiao Tung University, Taiwan;

[3] Shanghai University of Engineering Science, China

Agenda

- Introduction
- System Overview
- Computation of Vanishing Points
  - Object Tracking and Head/Foot Localization
  - Vanishing Points Estimation
- Self-Calibration by Optimization
- Experimental Results
- Conclusion

Agenda

- **Introduction**
- System Overview
- Computation of Vanishing Points
  - Object Tracking and Head/Foot Localization
  - Vanishing Points Estimation
- Self-Calibration by Optimization
- Experimental Results
- Conclusion

Introduction

- Geometry of camera model

$$[u, v, 1]^T \sim \boldsymbol{P} \cdot [X, Y, Z, 1]^T$$

$$\boldsymbol{P} = \boldsymbol{K} \cdot [\boldsymbol{R}|\boldsymbol{t}]$$

$$\boldsymbol{K} = \begin{bmatrix} \boldsymbol{f_x} & s & \boldsymbol{c_x} \\ 0 & \boldsymbol{f_y} & \boldsymbol{c_y} \\ 0 & 0 & 1 \end{bmatrix} \quad \boldsymbol{t} = \begin{bmatrix} t_X \\ \boldsymbol{t_Y} \\ t_Z \end{bmatrix}$$

$$\boldsymbol{R} = \boldsymbol{R}_Z \boldsymbol{R}_X \boldsymbol{R}_Y$$

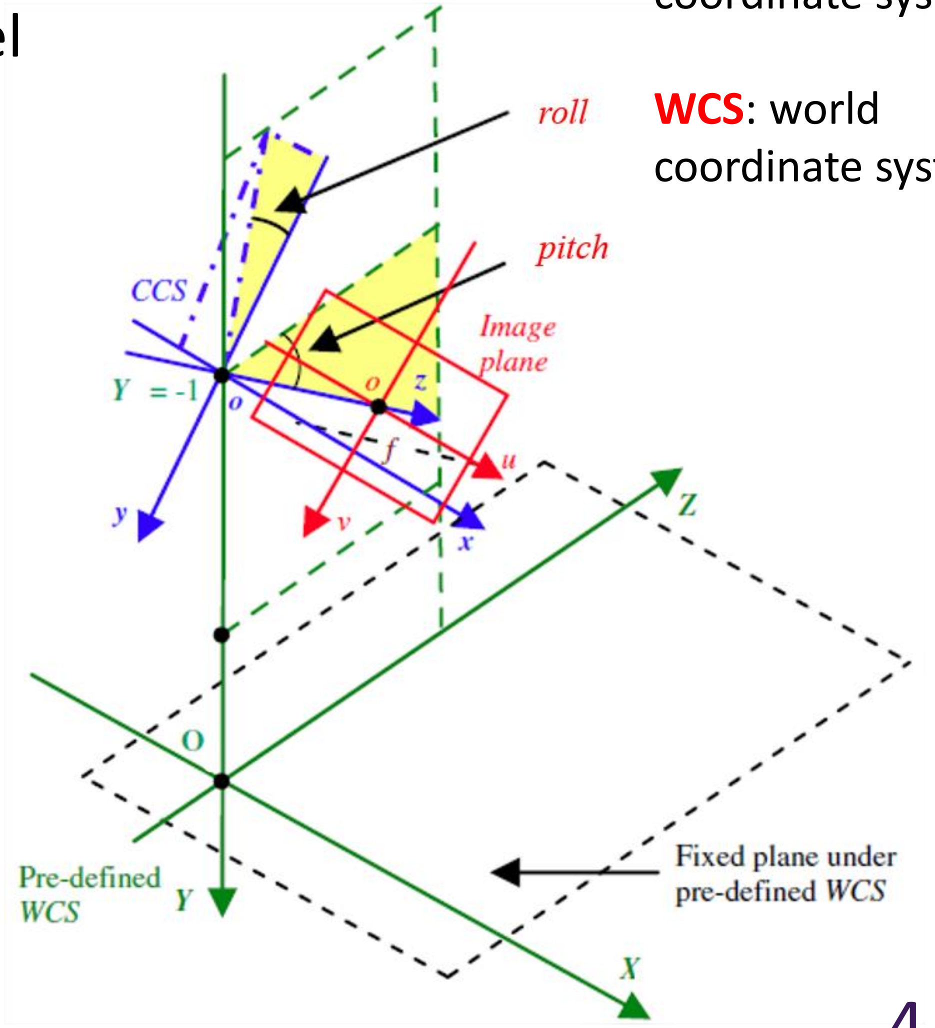
$$\boldsymbol{R}_Z = \begin{bmatrix} \cos(\boldsymbol{roll}) & -\sin(\boldsymbol{roll}) & 0 \\ \sin(\boldsymbol{roll}) & \cos(\boldsymbol{roll}) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\boldsymbol{R}_X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\boldsymbol{pitch}) & -\sin(\boldsymbol{pitch}) \\ 0 & \sin(\boldsymbol{pitch}) & \cos(\boldsymbol{pitch}) \end{bmatrix}$$

$$\boldsymbol{R}_Y = \begin{bmatrix} \cos(\boldsymbol{yaw}) & 0 & -\sin(\boldsymbol{yaw}) \\ 0 & 1 & 0 \\ \sin(\boldsymbol{yaw}) & 0 & \cos(\boldsymbol{yaw}) \end{bmatrix}$$

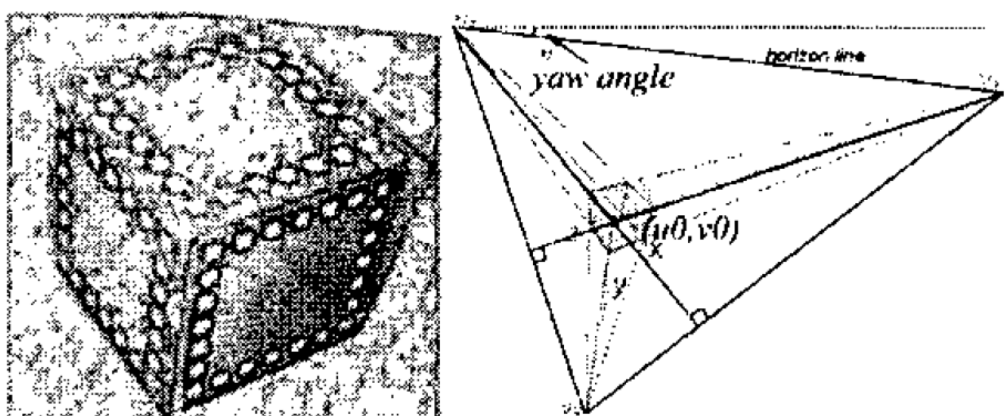
CCS: camera coordinate system

WCS: world coordinate system

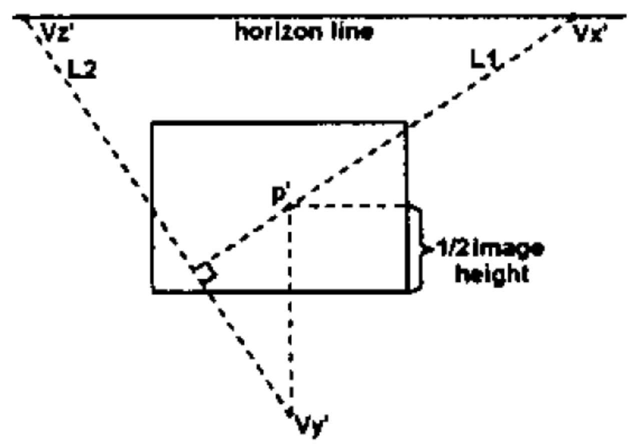


Introduction

- Categories of camera calibration
  - Calibration using calibrated templates
  - Self-calibration/auto-calibration
    - Static scene structures
    - Object motion -> **Moving-person tracking**

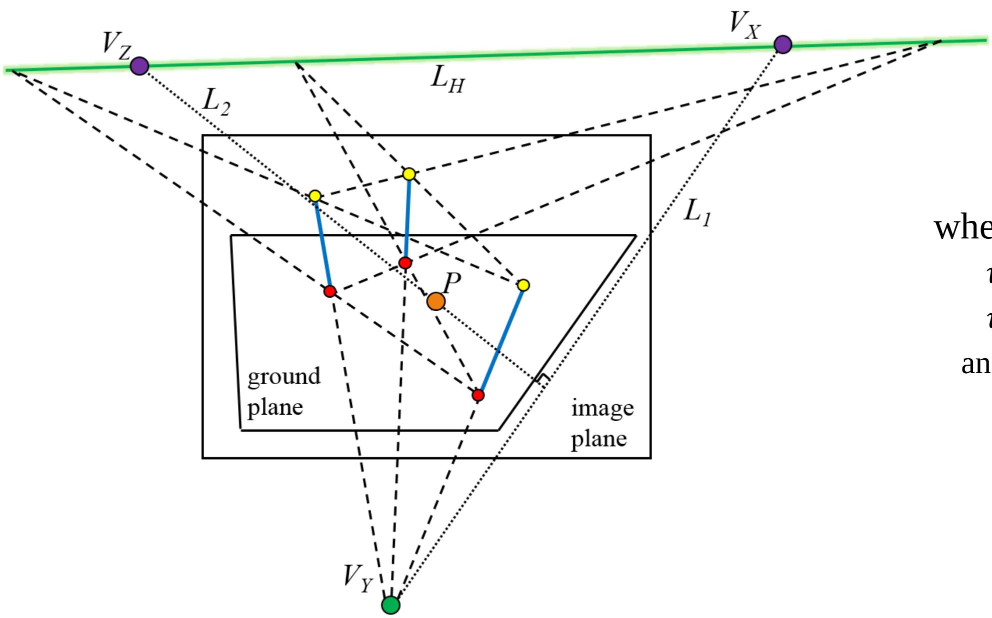


(a) (b)



Introduction

- Self-calibration from tracking of moving persons [Lv *et al.*, 2002]



$$roll = \tan^{-1}\left(\frac{v_{V_Z} - v_{V_X}}{u_{V_X} - u_{V_Z}}\right)$$

$$f_x = f_y = \sqrt{-\left(v_{V_X}^{rot} \cdot v_{V_Z}^{rot} + u_{V_X}^{rot} \cdot u_{V_Z}^{rot}\right)}$$

where $v_{V_X}^{rot} = \cos(roll)\,(v_P - v_{V_X}) - \sin(roll)\,(u_{V_X} - u_P)$,
$v_{V_Z}^{rot} = \cos(roll)\,(v_P - v_{V_Z}) - \sin(roll)\,(u_{V_Z} - u_P)$,
$u_{V_X}^{rot} = \cos(roll)\,(u_{V_X} - u_P) + \sin(roll)(v_P - v_{V_X})$,
and $u_{V_Z}^{rot} = \cos(roll)\,(u_{V_Z} - u_P) + \sin(roll)(v_P - v_{V_Z})$,

$$pitch = \tan^{-1}\left(\frac{v_{V_X}^{rot}}{f_x}\right)$$

$$yaw = -\tan^{-1}\left(\frac{f_x}{\cos(pitch) \cdot u_{V_X}^{rot}}\right)$$

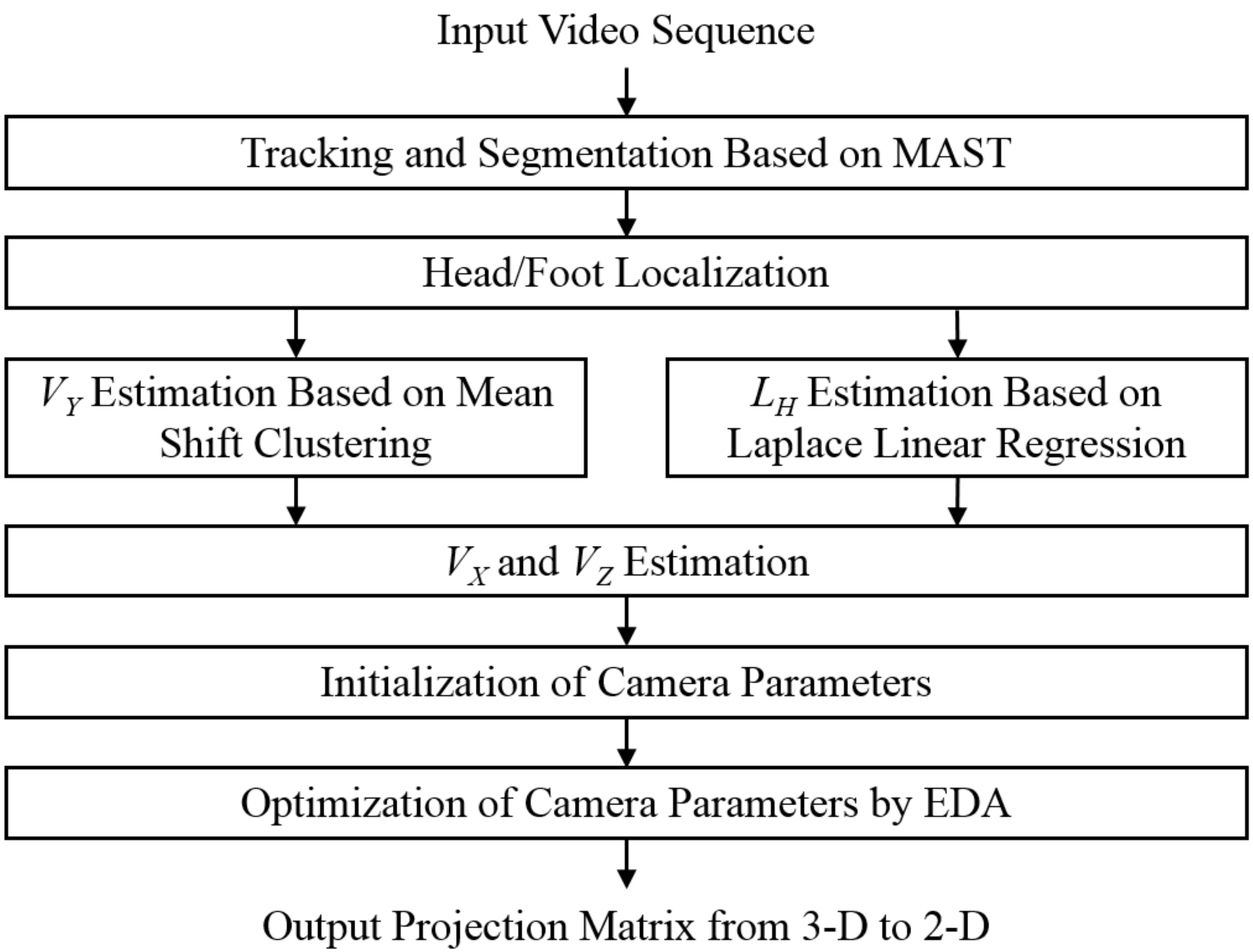
Original assumptions:
(1) Central principal point
(2) Unit aspect ratio
(3) Zero skew

Challenges:
(1) How to find the accurate V_Y and L_H?
(2) How can we optimize all camera parameters (relax original assumptions)?

Agenda

- Introduction
- **System Overview**
- Computation of Vanishing Points
  - Object Tracking and Head/Foot Localization
  - Vanishing Points Estimation
- Self-Calibration by Optimization
- Experimental Results
- Conclusion

System Overview



Agenda

- Introduction
- System Overview
- **Computation of Vanishing Points**
  - **Object Tracking and Head/Foot Localization**
  - Vanishing Points Estimation
- Self-Calibration by Optimization
- Experimental Results
- Conclusion

System Overview

Input Video Sequence

Tracking and Segmentation Based on MAST

Head/Foot Localization

V_Y Estimation Based on Mean Shift Clustering

L_H Estimation Based on Laplace Linear Regression

V_X and V_Z Estimation

Initialization of Camera Parameters

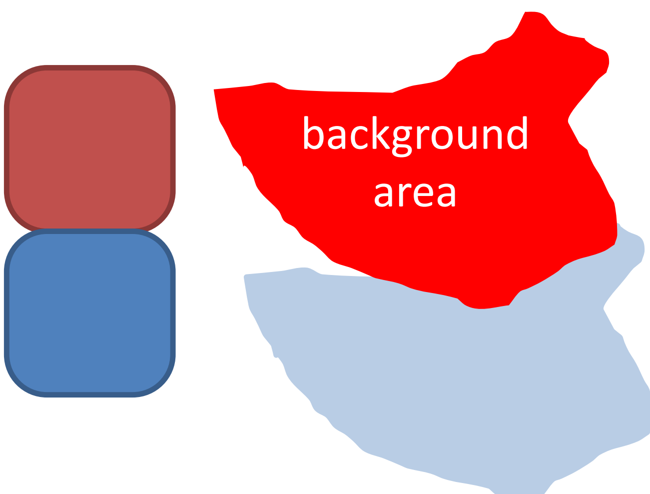
Optimization of Camera Parameters by EDA

Output Projection Matrix from 3-D to 2-D

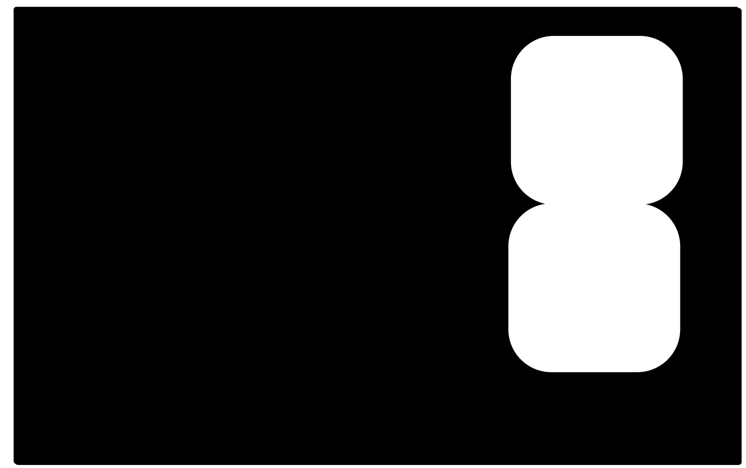
Object Tracking and Segmentation Based on MAST

Multiple-kernel **A**daptive **S**egmentation and **T**racking

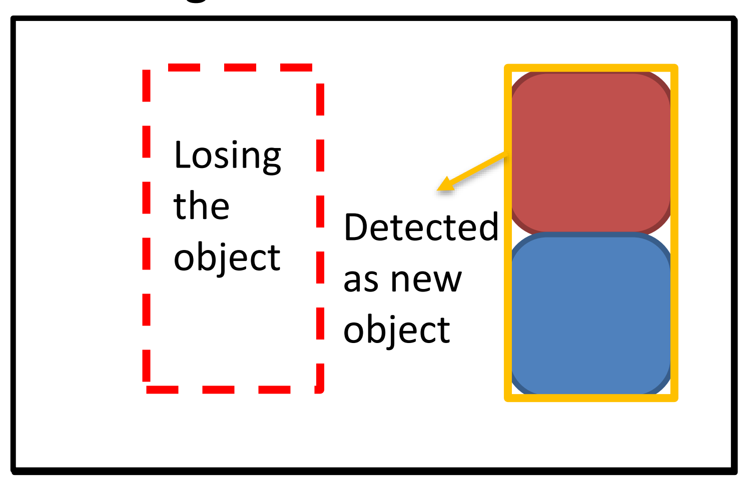
- Object merging



Segmentation result



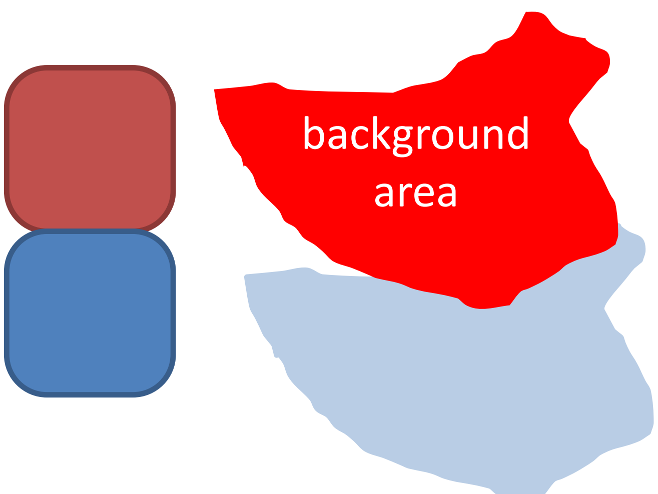
Tracking result



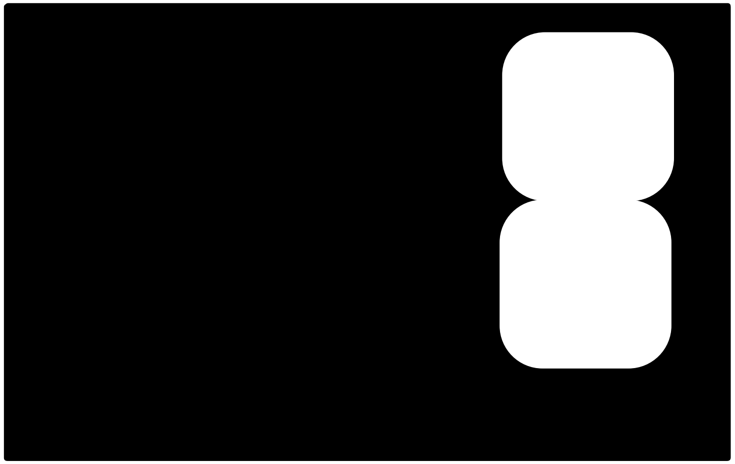
Object Tracking and Segmentation Based on MAST

Multiple-kernel **A**daptive **S**egmentation and **T**racking

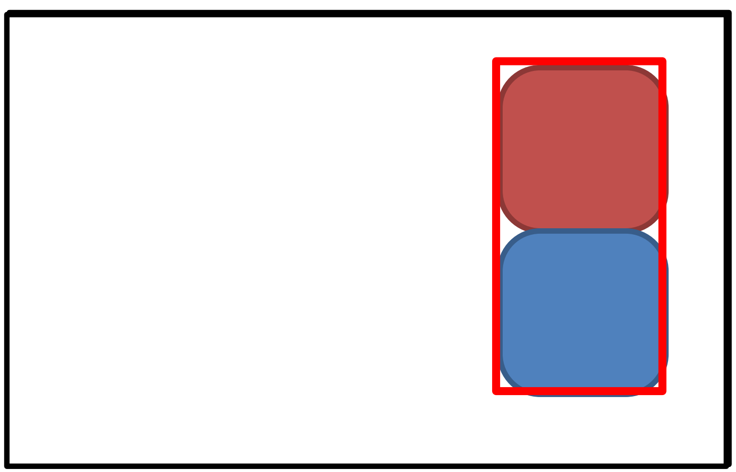
- Motivation



Segmentation result

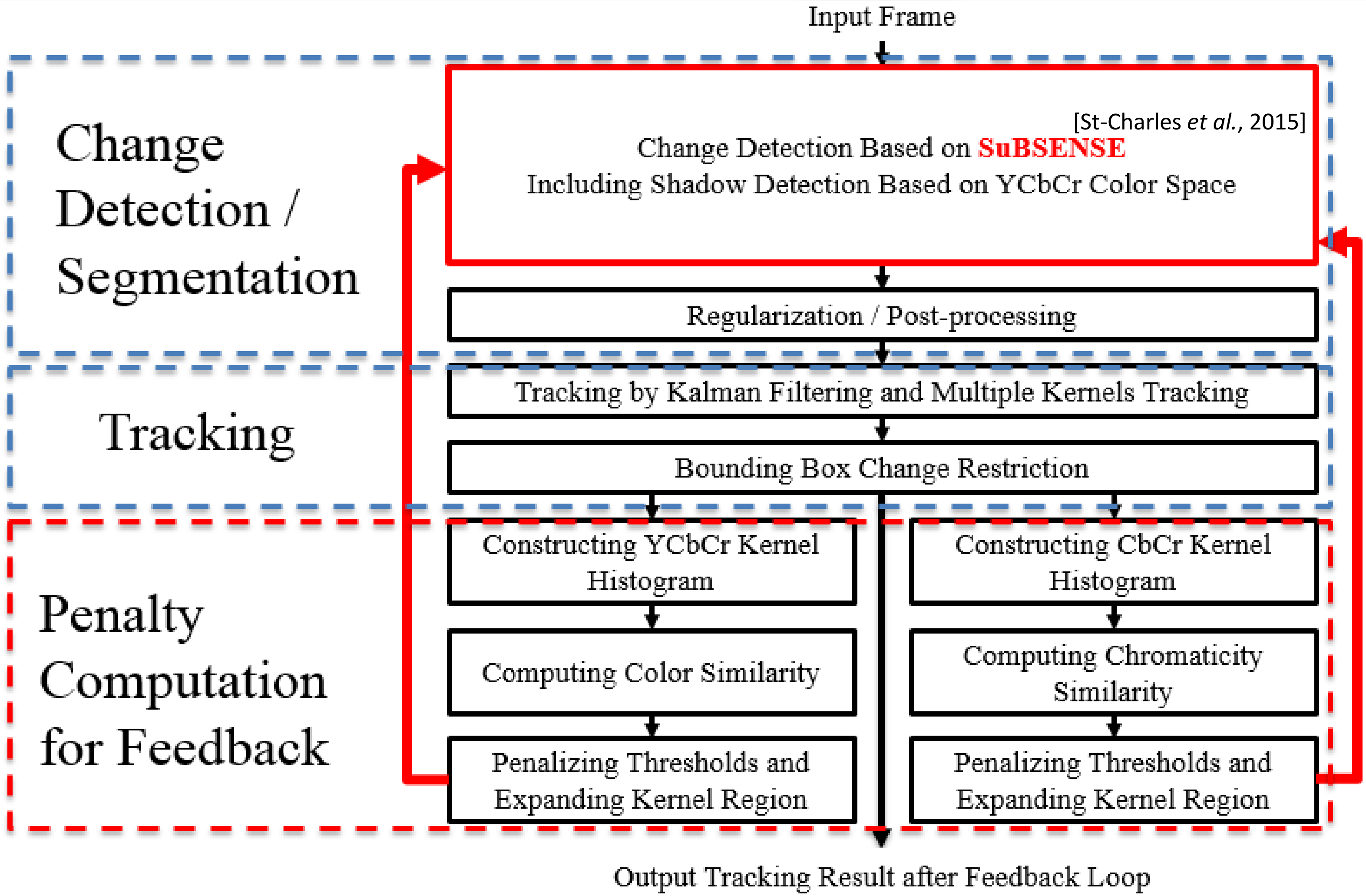


Tracking result

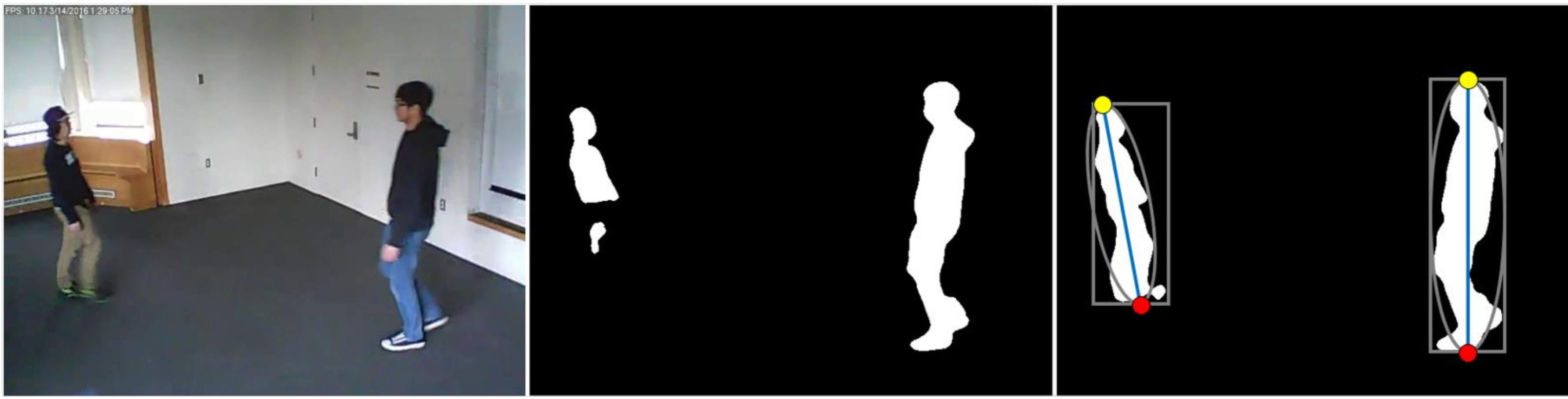


Object Tracking and Segmentation Based on MAST

Multiple-kernel **A**daptive **S**egmentation and **T**racking [Tang *et al.*, 2016]



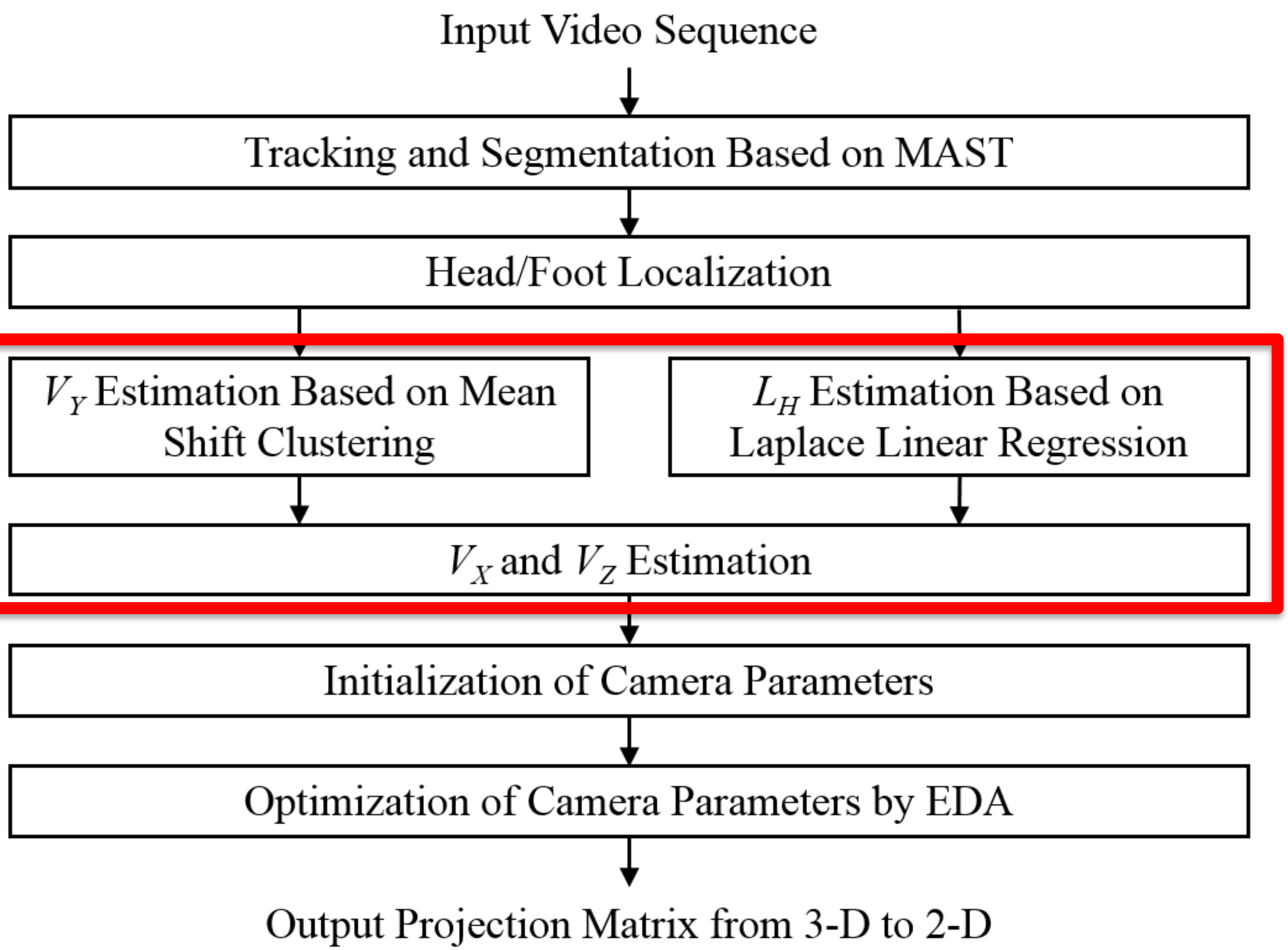
Head/Foot Localization



Agenda

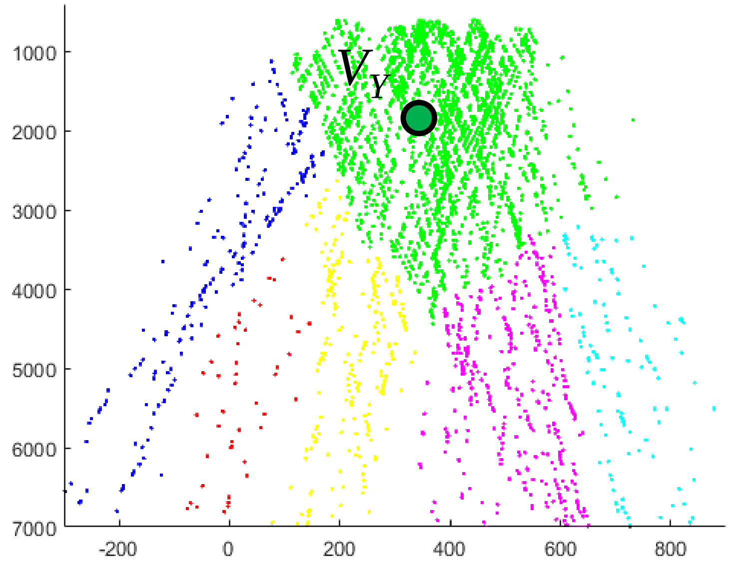
- Introduction
- System Overview
- Computation of Vanishing Points
  - Object Tracking and Head/Foot Localization
  - **Vanishing Points Estimation**
- Self-Calibration by Optimization
- Experimental Results
- Conclusion

System Overview



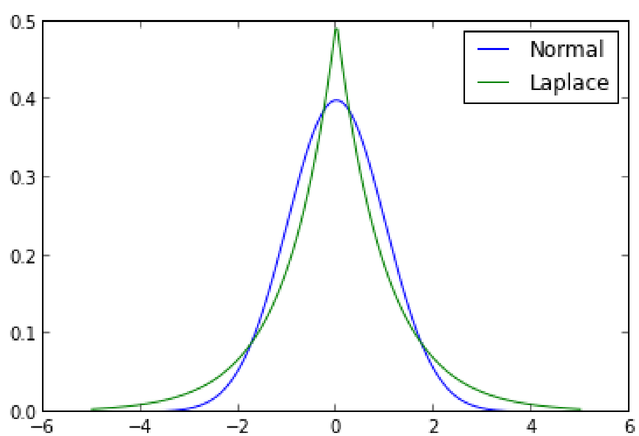
Vertical Vanishing Point (V_Y) Estimation Based on Mean Shift Clustering

- Disadvantage of RANSAC
  - Failure when the number of outliers is significantly large
- Proposed method
  - Mean shift clustering among all the candidate points of V_Y
  - Choosing the mean point of the largest cluster as the estimated V_Y



Horizon Line (L_H) Estimation Based on Laplace Linear Regression

- Disadvantage of RANSAC
  - Setting of threshold parameter for inliers
- Proposed method
  - Formulating as convex optimization by Laplace linear regression



$$p(\mathbf{y}|\mathbf{x},\mathbf{w}) = \mathrm{Laplace}\left(\mathbf{y}|\mathbf{w}^T\mathbf{x}\right)$$
$$\propto \exp\left(-\left|\mathbf{y}-\mathbf{w}^T\mathbf{x}\right|\right)$$

$$\min_{\mathbf{w},\mathbf{r}} \sum_i r_i = \min_{\mathbf{w},\mathbf{r}^+,\mathbf{r}^-} \sum_i (r_i^+ + r_i^-)$$
$$\text{s.t. } r_i^+ \geq 0, r_i^- \geq 0, \mathbf{w}^T\mathbf{x}_i + r_i^+ - r_i^- = y_i$$

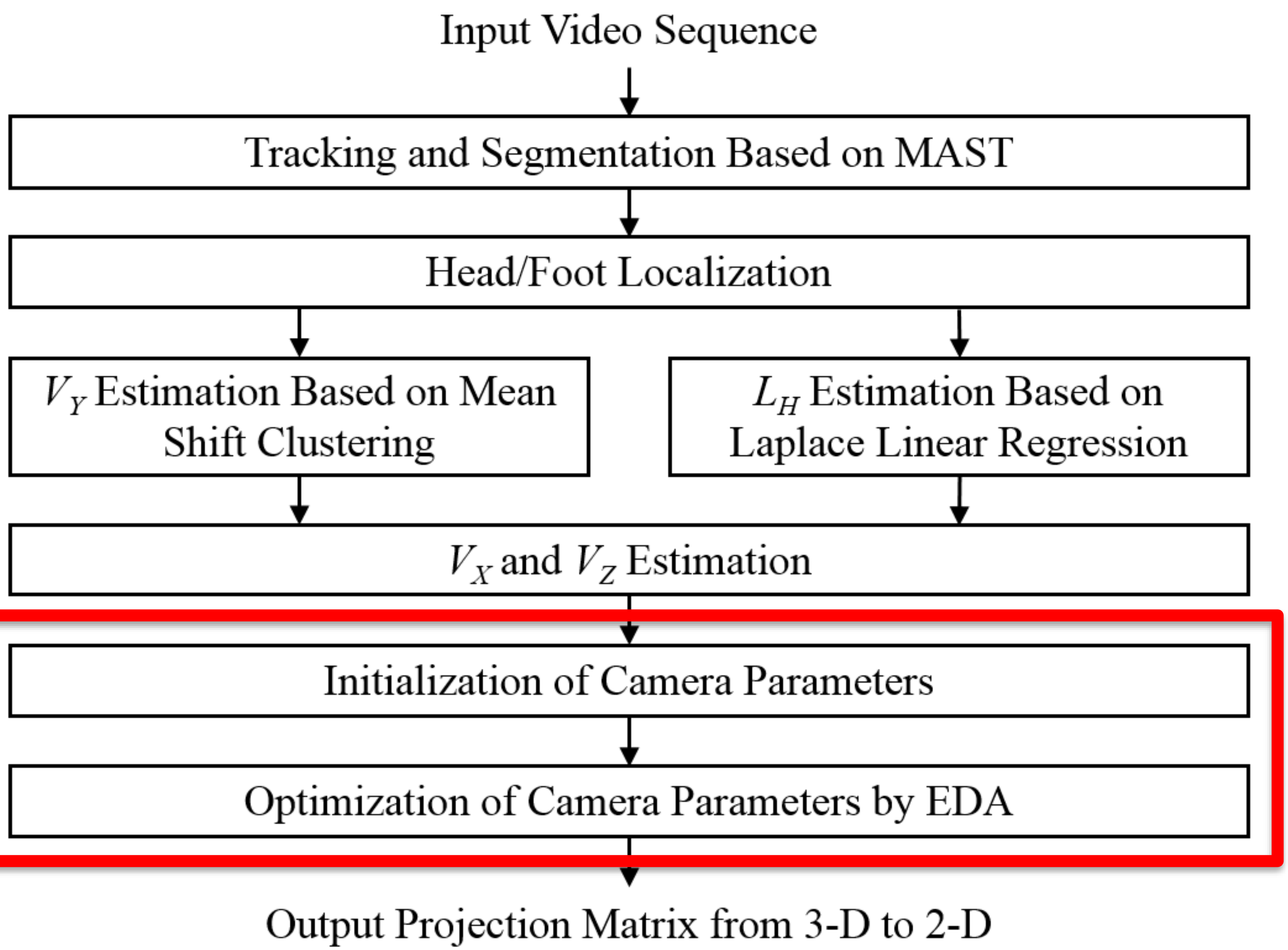
$$\min_{\boldsymbol{\theta}} \mathbf{f}^T\boldsymbol{\theta} \ \text{ s.t. } \mathbf{A}\boldsymbol{\theta} \leq \mathbf{b}, \mathbf{A}_{eq}\boldsymbol{\theta} = \mathbf{b}_{eq}, \mathbf{l} \leq \boldsymbol{\theta} \leq \mathbf{u}$$

in which $\boldsymbol{\theta} = (\mathbf{w},\mathbf{r}^+,\mathbf{r}^-)$, $\mathbf{f} = [\mathbf{0},\mathbf{1},\mathbf{1}]$, $\mathbf{A} = []$, $\mathbf{b} = []$, $\mathbf{A}_{eq} = [\mathbf{x},\mathbf{I},-\mathbf{I}]$, $\mathbf{b}_{eq} = \mathbf{y}$, $\mathbf{l} = [-\infty\mathbf{1},\mathbf{0},\mathbf{0}]$ and $\mathbf{u} = []$.

Agenda

- Introduction
- System Overview
- Computation of Vanishing Points
  - Object Tracking and Head/Foot Localization
  - Vanishing Points Estimation
- **Self-Calibration by Optimization**
- Experimental Results
- Conclusion

System Overview

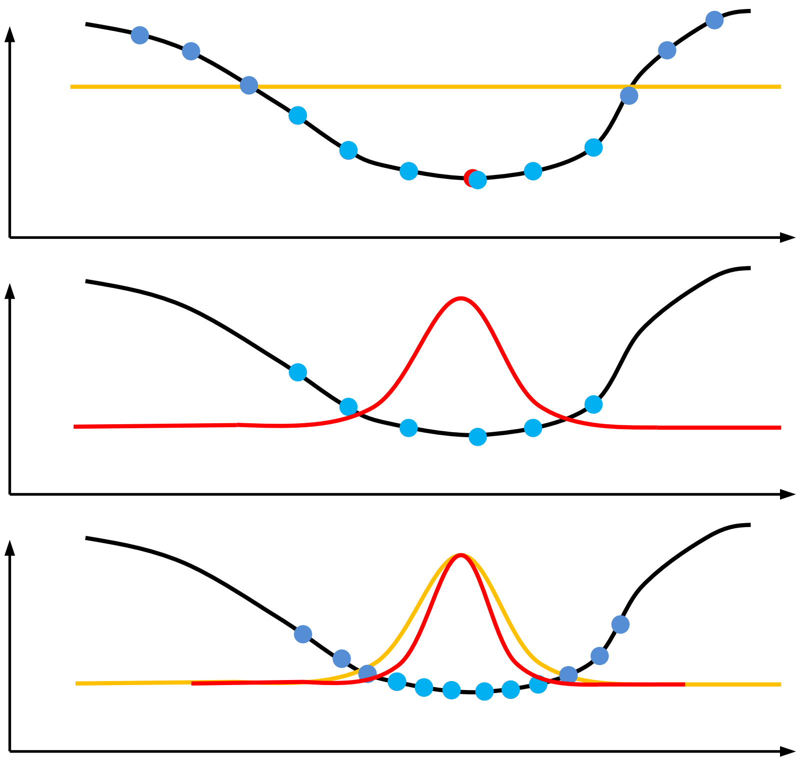


Estimation of Distribution Algorithm (EDA)

Consider $\arg\min_{x} f(x)$

1. Randomly generate *R* samples.
2. Calculate $f(x_i)$ of each sample, and sort the results.
3. Use the best *N* results to generate a *pdf* with normal distribution.
4. If stopping criterion is not met, use the *pdf* to generate new *R* samples, jump to 2.

In this example, *R* = 12, *N* = 6

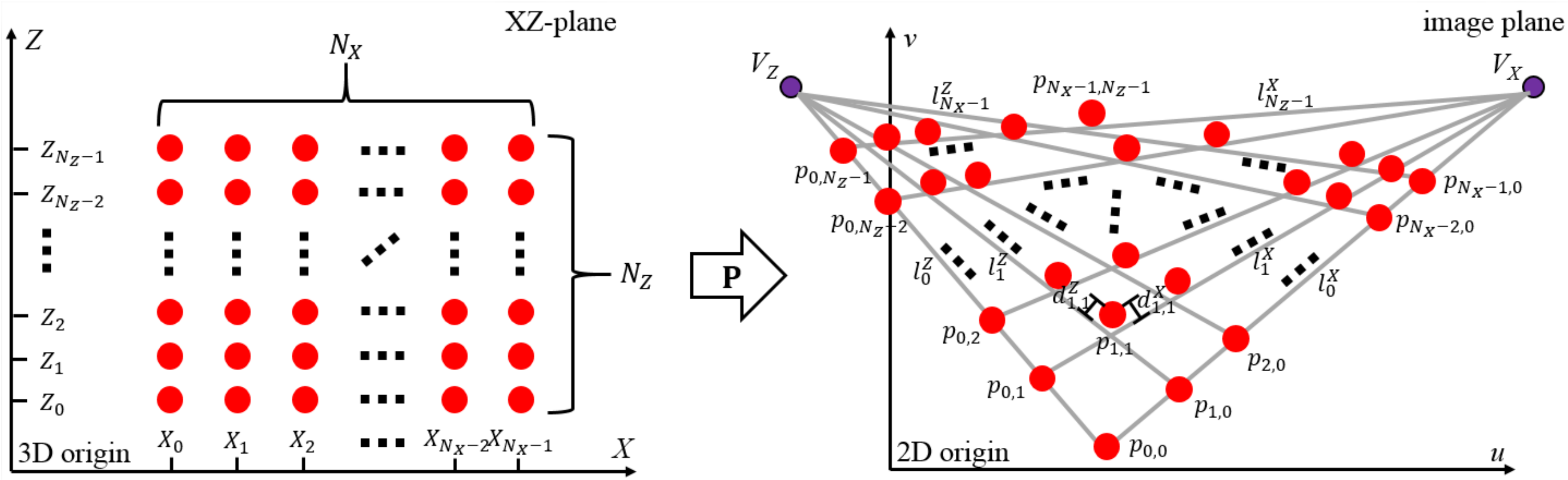


⋮

until stopping criterion is met

Optimization of Camera Parameters by EDA

Estimation of **M**ultivariate **N**ormal **A**lgorithm–global [Larrauaga *et al.*, 2002]



- Extending from univariate EDA, for multivariate scenario (8 parameters)
    - Using multivariate normal density function as *pdf*
- Each projection matrix formed by a set of camera parameters is regarded as a sample.
- Using reprojection error on the ground plane as the evaluation
- Stopping criterion
    - Change of reprojection error between generations is small enough
    - Number of generations is too large

Advantages of the Proposed Formulation

- Optimizing all camera parameters simultaneously by EMNA_global
- Relaxing original assumptions on intrinsic camera parameters by allowing them to be optimized within given ranges
- Advantages of EDA [Hauschild *et al.*, 2011]
  - Ability to adapt their operators to the structure of the problem
  - Prior knowledge exploitation
  - Reduced memory requirements
  - Implementation of parallel computation

Agenda

- Introduction
- System Overview
- Computation of Vanishing Points
  - Object Tracking and Head/Foot Localization
  - Vanishing Points Estimation
- Self-Calibration by Optimization
- **Experimental Results**
- Conclusion

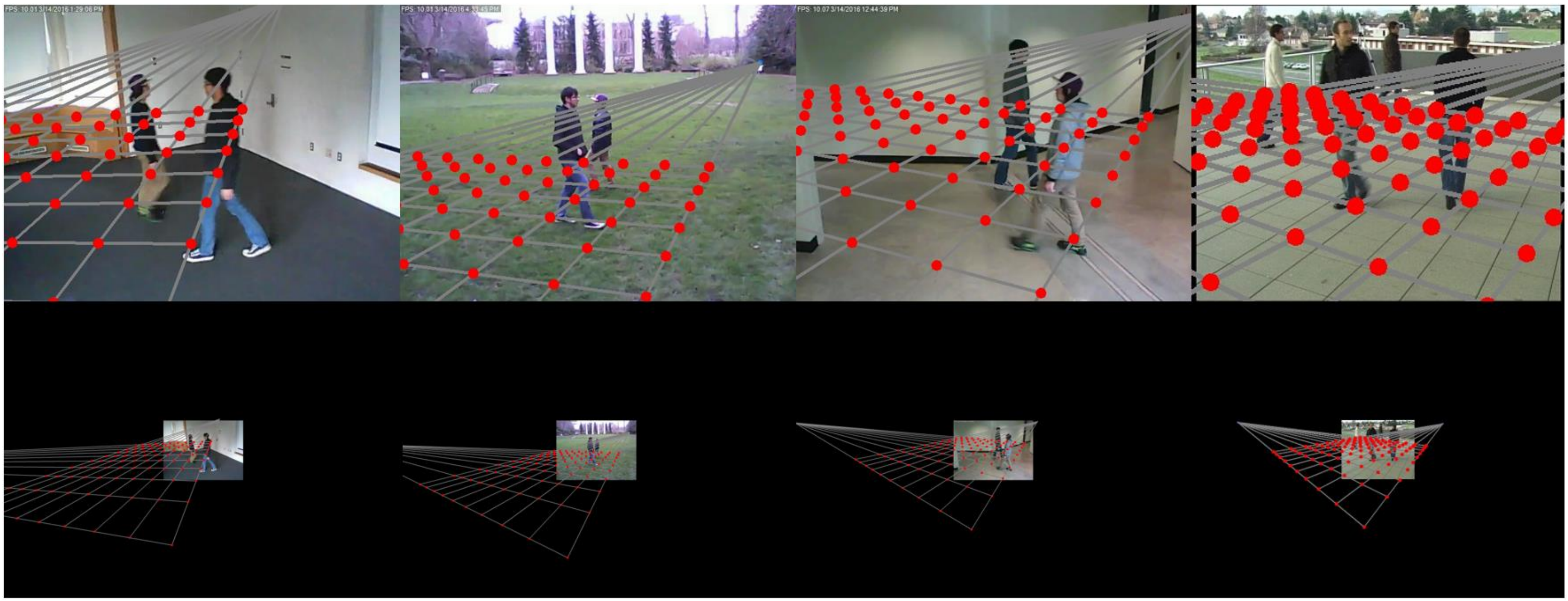
Experimental Results

- 3 captured video sequences
  - Length: ~ 1 min 30 sec
  - Resolution: 640 * 480
  - Frame rate: 10 fps
  - Ground truth: Extracted using linear method based on 52, 52, and 38 measured 3-D points
- 1 video sequence from EPFL dataset
  - Length: 3 min
  - Resolution: 360 * 288
  - Frame rate: 25 fps
  - Ground truth: Extracted using Tsai's method

Experimental Results

Seq. #	f_x (pix.)	f_y (pix.)	c_x (pix.)	c_y (pix.)	roll (deg.)	pitch (deg.)	yaw (deg.)	μ_{err} (pix.)
1. Ground Truth	731.3880	728.2518	322.1298	237.2676	-3.1371	16.2676	-78.3065	N/A
1. Method in [6]	611.5239	611.5239	320.0000	**240.0000**	5.7439	22.4758	-64.9974	11.7954
1. Method in [10]	638.2676	638.2676	320.0000	**240.0000**	**3.8800**	23.2010	-71.8167	8.7750
1. Proposed w/o EDA	**738.7650**	**738.7650**	**320.0000**	**240.0000**	5.0689	**17.6076**	**-79.0154**	**6.0133**
1. Proposed	**730.9167**	**735.9371**	**322.9955**	**236.1948**	**-5.0345**	**17.4224**	**-79.1491**	**2.50E-5**
2. Ground Truth	731.3880	728.2518	322.1298	237.2676	-1.8887	11.0081	-68.7126	N/A
2. Method in [6]	618.7858	618.7858	320.0000	**240.0000**	2.3671	8.7161	-71.5302	4.9334
2. Method in [10]	647.4640	647.4640	320.0000	**240.0000**	1.8874	9.8994	-71.7033	5.0624
2. Proposed w/o EDA	**679.6617**	**679.6617**	**320.0000**	**240.0000**	**1.7928**	**10.7818**	**-70.3027**	**4.6445**
2. Proposed	**727.6335**	**728.1606**	**321.4372**	**241.1506**	**-2.2546**	**10.3345**	**-70.3032**	**3.12E-5**
3. Ground Truth	731.3880	728.2518	322.1298	237.2676	-0.3459	18.3846	-63.8778	N/A
3. Method in [6]	606.8088	606.8088	**320.0000**	**240.0000**	-0.8635	13.2525	-67.1697	2.1670
3. Method in [10]	662.9474	662.9474	**320.0000**	**240.0000**	**-0.2164**	22.4663	-57.6830	0.5403
3. Proposed w/o EDA	**719.8882**	**719.8882**	**320.0000**	**240.0000**	0.2693	**17.4219**	**-64.7125**	**0.3398**
3. Proposed	**720.6649**	**729.5090**	**319.8556**	**240.6065**	**-0.2658**	**17.2493**	**-64.7081**	**1.17E-4**
4. Ground Truth	437.2689	437.8792	173.7693	142.7878	1.5466	14.1153	-54.5257	N/A
4. Method in [6]	406.8041	406.8041	**180.0000**	**144.0000**	-0.2633	22.4482	-63.5813	0.5051
4. Method in [10]	**432.0973**	432.0973	**180.0000**	**144.0000**	**-0.2062**	20.8494	-45.6322	0.4321
4. Proposed w/o EDA	**440.5366**	**440.5366**	**180.0000**	**144.0000**	-0.4297	**16.2182**	**-55.8775**	**0.1858**
4. Proposed	442.4795	**440.9664**	**176.2516**	**142.1498**	**0.4313**	**15.9846**	**-55.6434**	**2.74E-5**

Experimental Results



Agenda

- Introduction
- System Overview
- Computation of Vanishing Points
  - Object Tracking and Head/Foot Localization
  - Vanishing Points Estimation
- Self-Calibration by Optimization
- Experimental Results
- **Conclusion**

Conclusion

- We proposed a robust single camera self-calibration method based on moving persons tracking.
- **Contribution (1)**: Combining the state-of-the-art change detection (SuBSENSE) and tracking (MAST) to generate accurate head/foot localization
- **Contribution (2)**: Introducing mean shift clustering and Laplace linear regression to the estimation of vanishing points
- **Contribution (3)**: formulating the problem of camera parameters optimization by EDA that can relax the assumptions on unknown intrinsic parameters.

Thank you!

Q&A