ELECTRICAL & COMPUTER ENGINEERING

Robust Video Object Tracking via Camera Self-calibration

Ph.D. Dissertation Defense

Zheng (Thomas) Tang

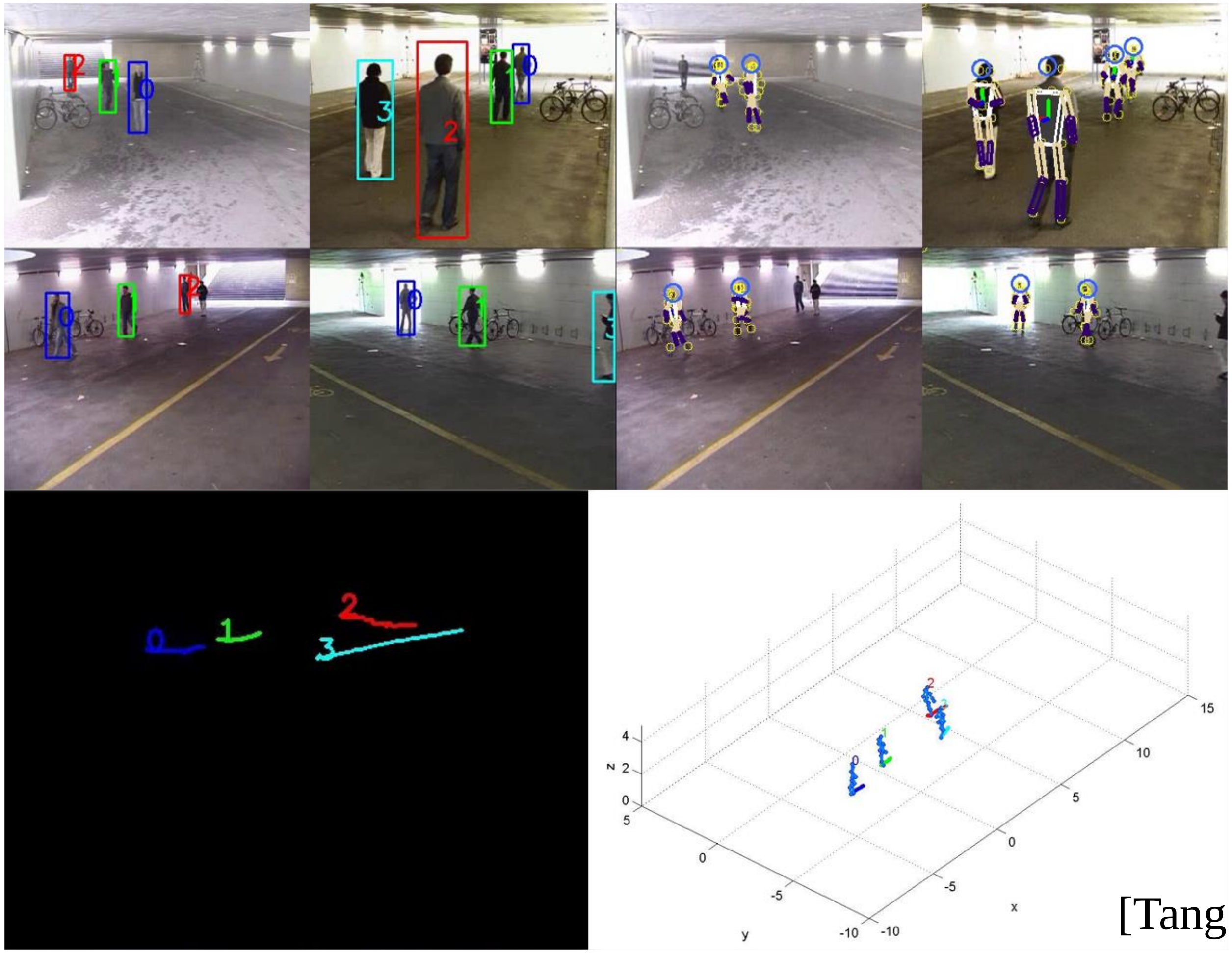
Committee:

Prof. Jenq-Neng Hwang (Chair, ECE)
Prof. Kenneth P. Bube (GSR, Mathematics)
Prof. Linda G. Shapiro (CSE & ECE)
Prof. Ming-Ting Sun (ECE)
Prof. Fa-Long Luo (ECE & Micron Technology)

Introduction

Multi-view 2D tracking

2D pose estimation

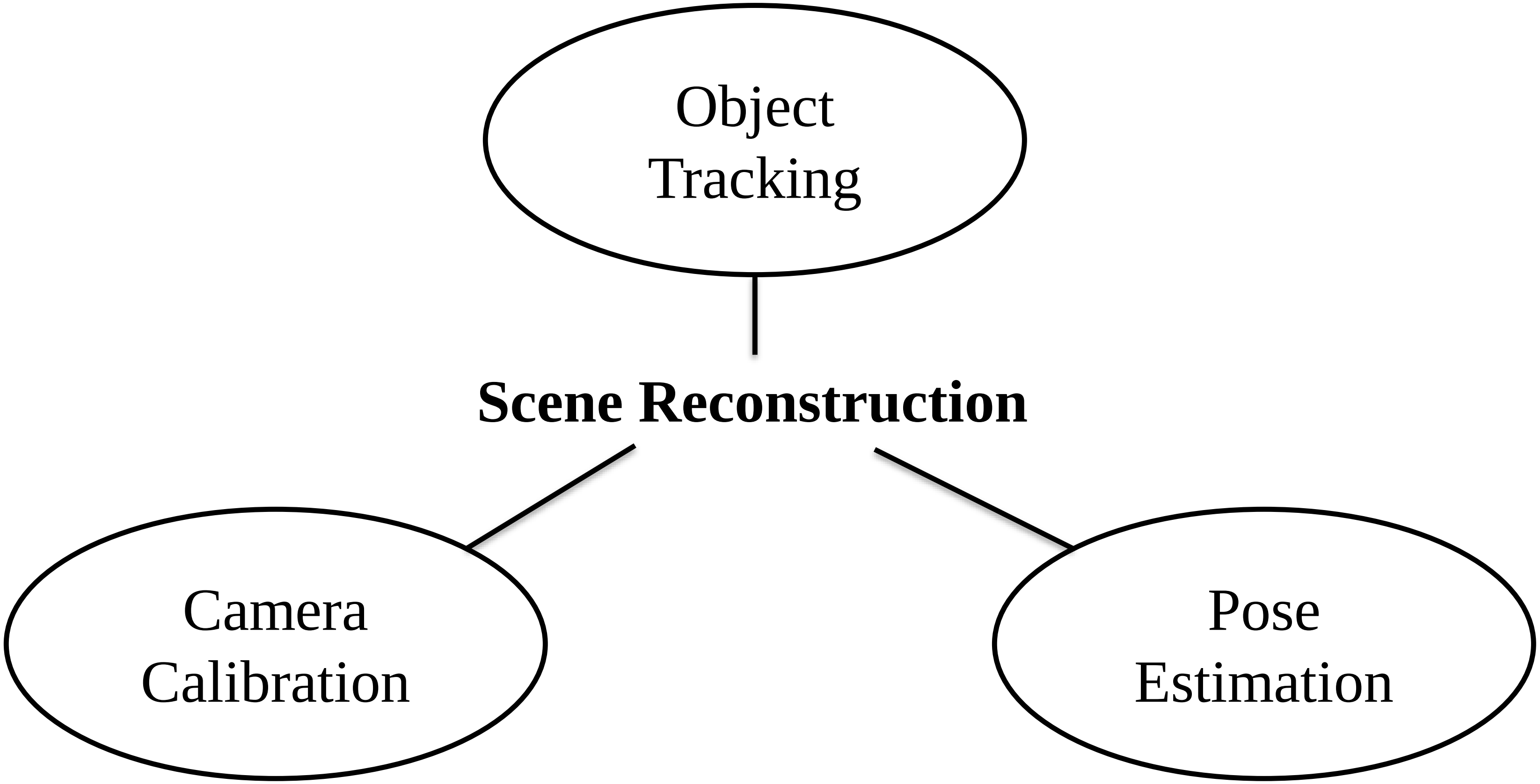


3D tracking (top view)

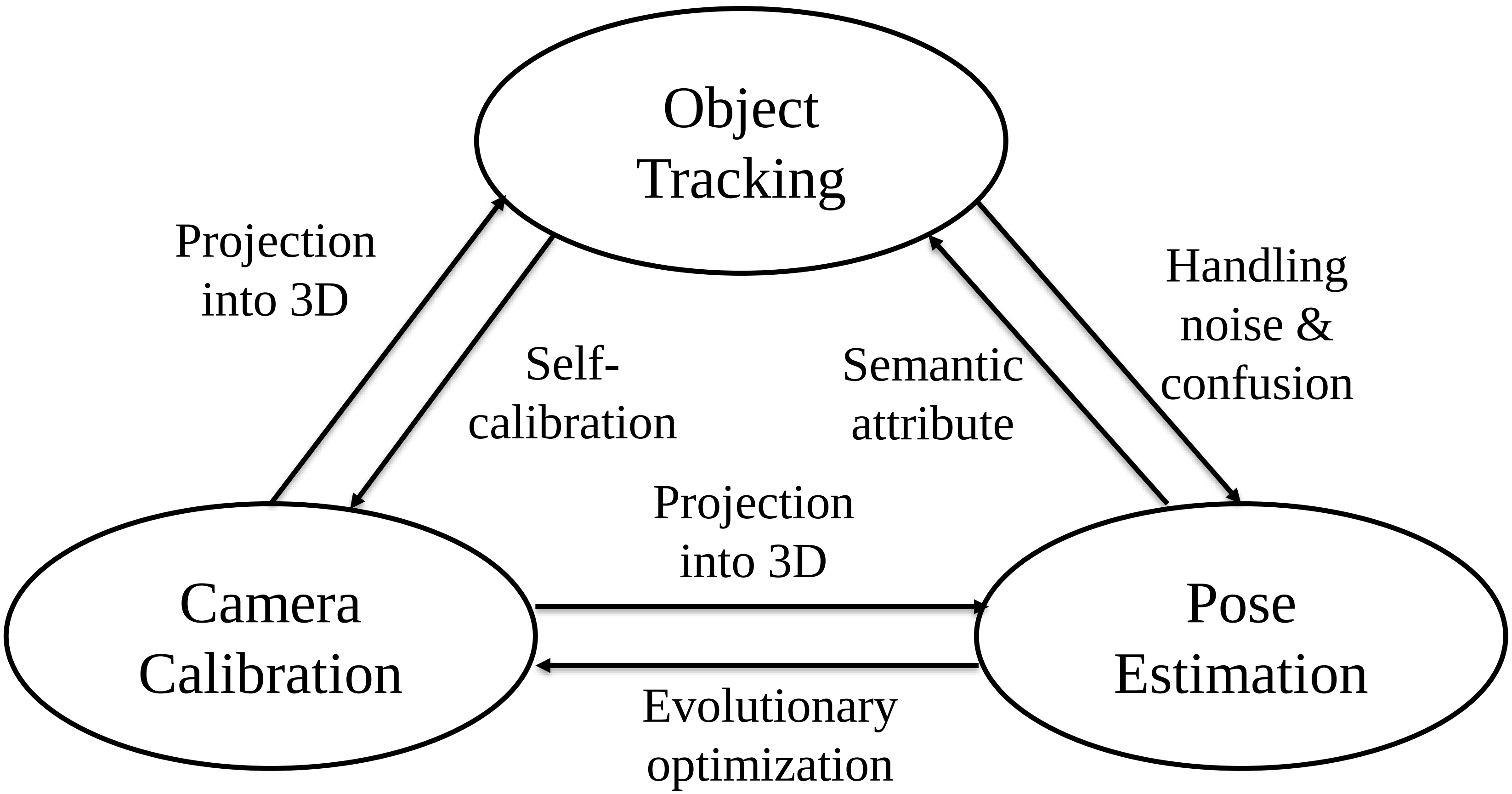
3D scene reconstruction

[Tang *et al.*, ICME’18]

Introduction



Introduction



Object Tracking

- Single-target / visual object tracking (VOT)

[Chu *et al.*, TMM'13]

Object Tracking

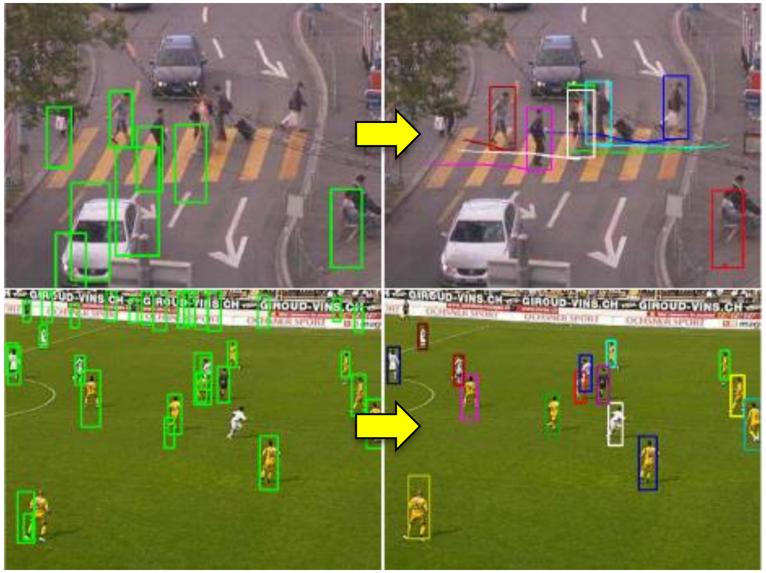
- Single-target / visual object tracking (VOT)
- Multiple object tracking (MOT)

[Chu *et al.*, TMM’13]

[Tang *et al.*, IEEE Access’19]

Object Tracking

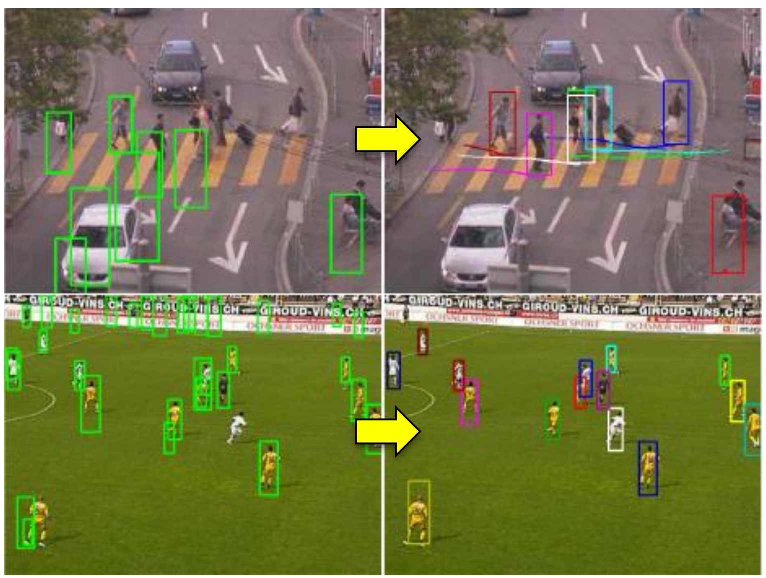
- Tracking by detection



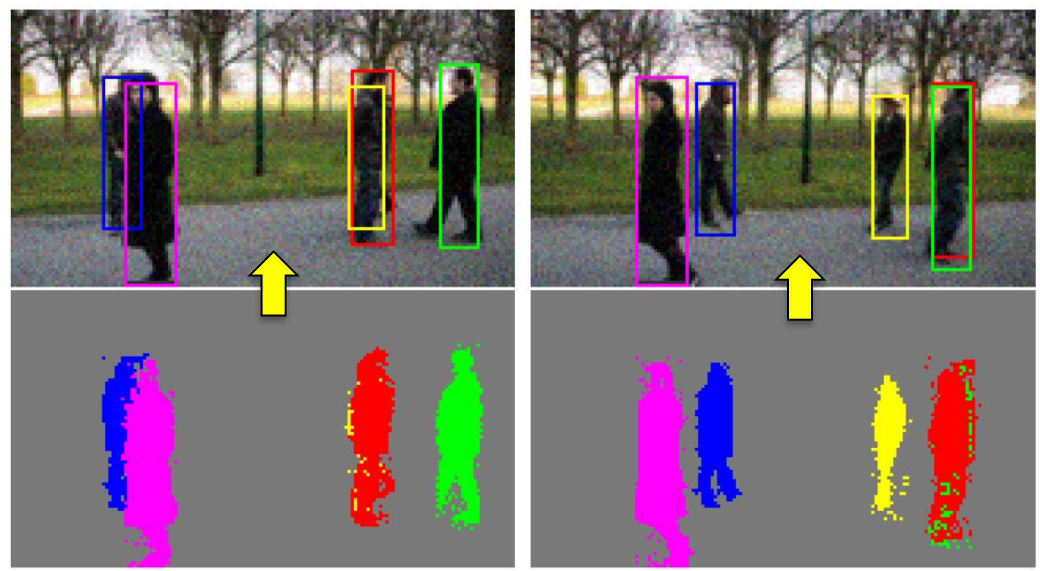
[Breitenstein *et al.*, ICCV'09]

Object Tracking

- Tracking by detection
- Tracking by segmentation



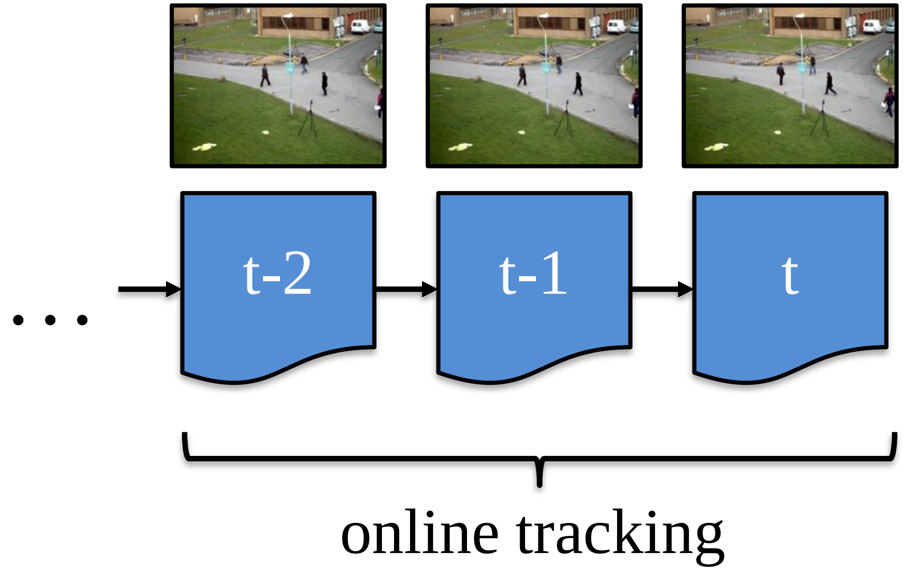
[Breitenstein *et al.*, ICCV'09]



[Wang *et al.*, ICCV'09]

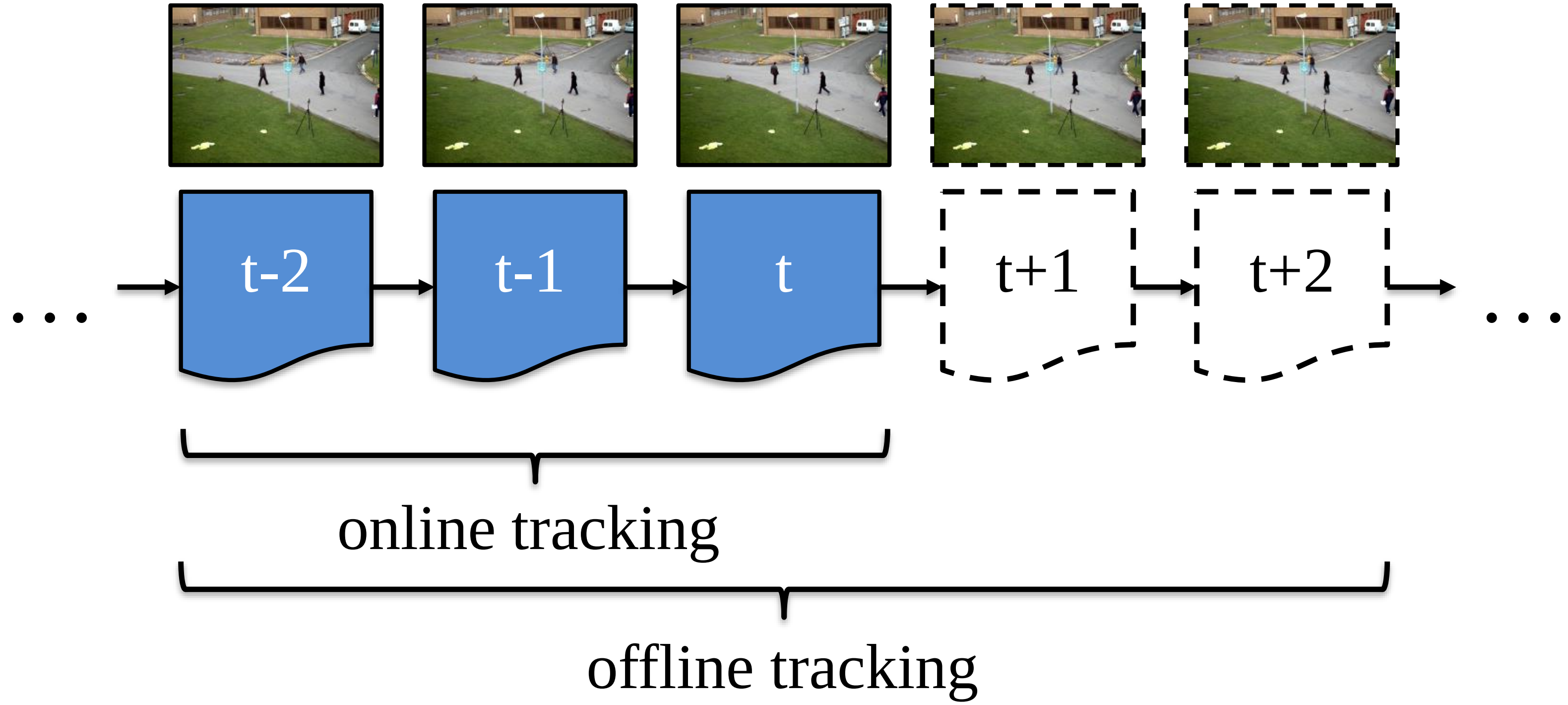
Object Tracking

- Online tracking



Object Tracking

- Online tracking
- Offline tracking



Object Tracking

- Human-based tracking

[Tang *et al.*, IEEE Access'19]

Object Tracking

- Human-based tracking
- Vehicle-based tracking

[Tang *et al.*, IEEE Access'19]

[Tang *et al.*, CVPR'19]

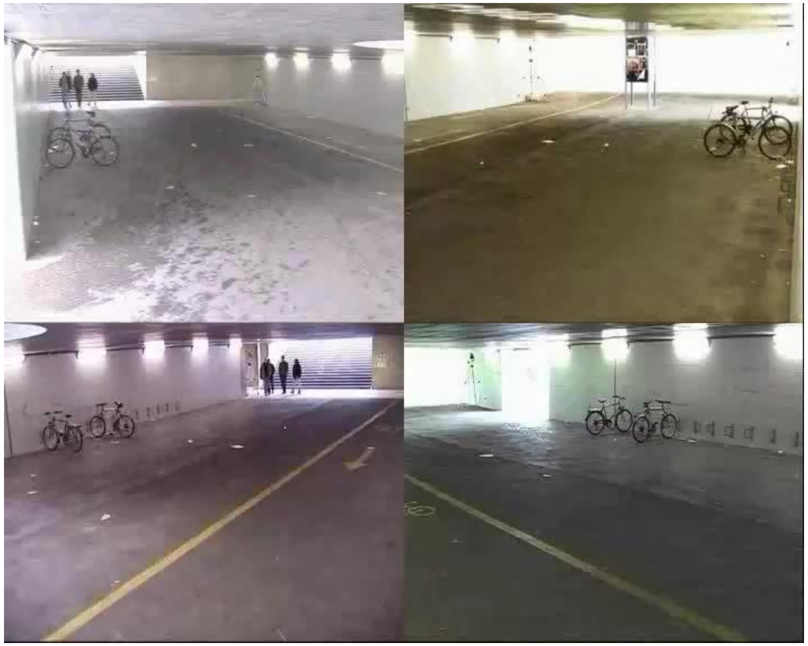
Object Tracking

- Single-view object tracking

[Tang *et al.*, ICME'18]

Object Tracking

- Single-view object tracking
- Multi-view / cross-view object tracking

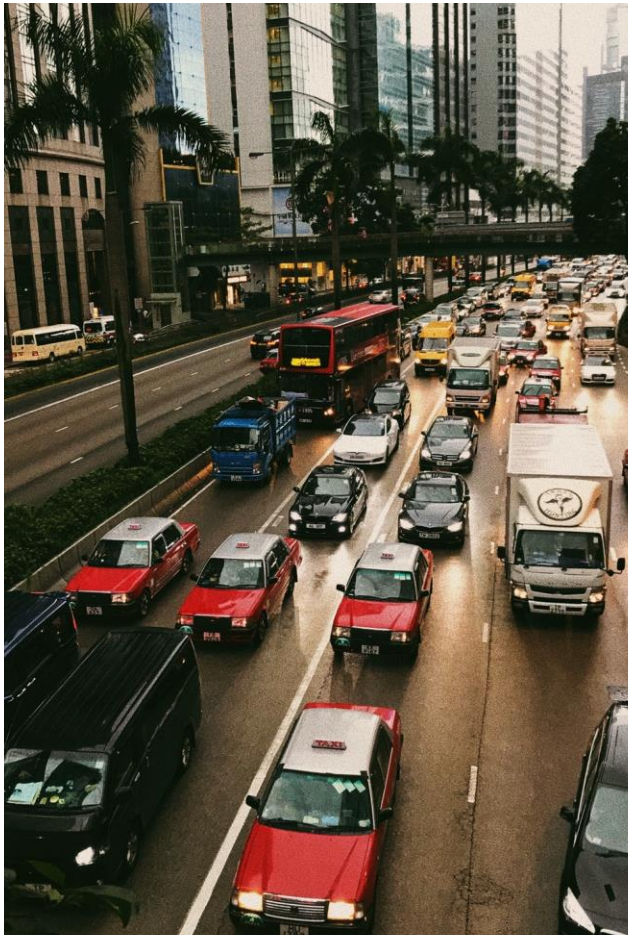


[Tang *et al.*, ICME’18]

Object Tracking

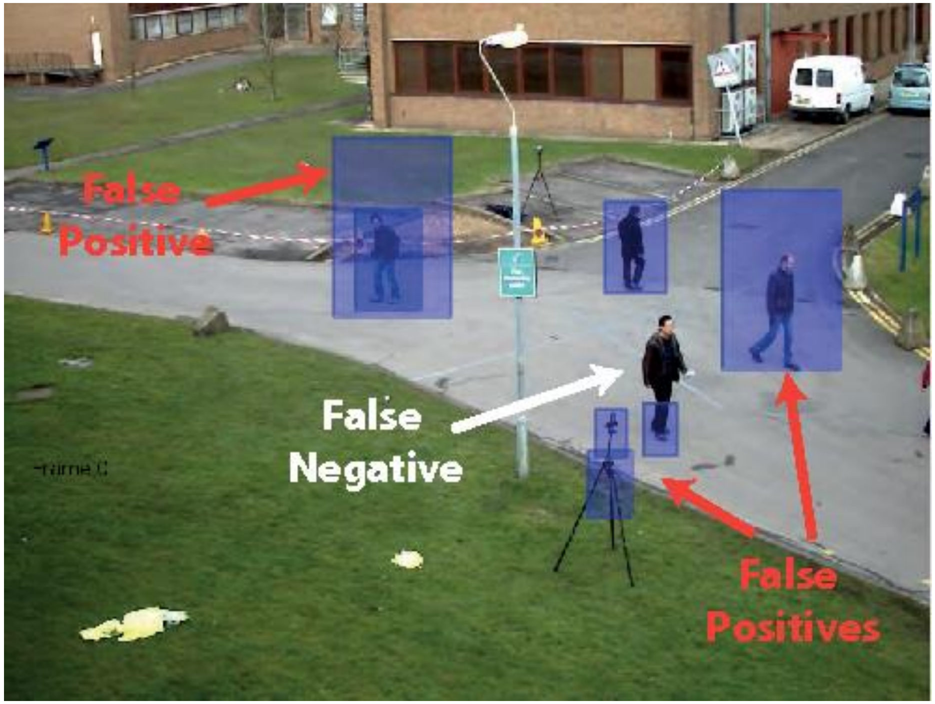
[Unsplash]

- Challenges
  - Object occlusion
  - Grouping of objects



Object Tracking

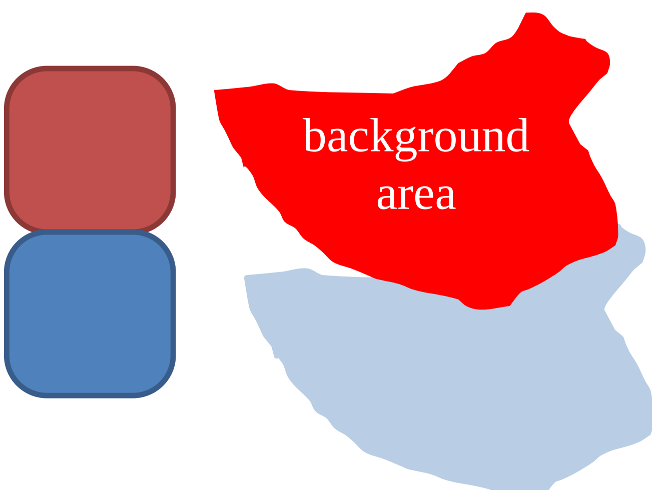
- Challenges
  - False negatives in detection (tracking by detection)
  - False positives in detection (tracking by detection)



[Yao *et al.*, CVPR'12]

Object Tracking

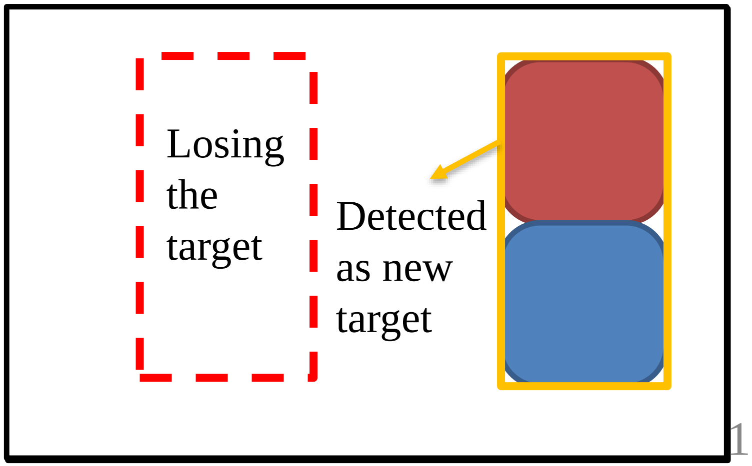
- Challenges
  - Object merging (tracking by segmentation)



[Tang *et al.*, ICASSP'16]

Segmentation results

Tracking results



Object Tracking

- Tracking in 2D
- Tracking in 3D

[Tang *et al.*, ICPR’16]

Camera Calibration

projection matrix

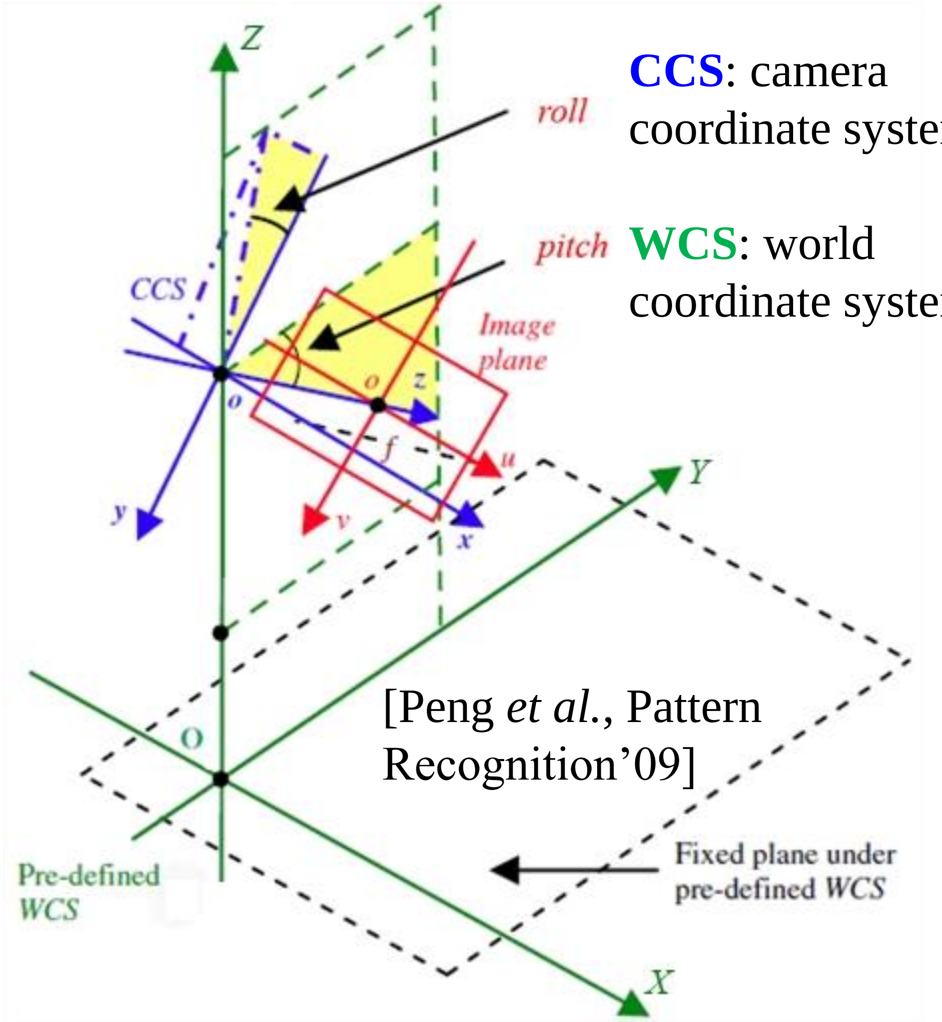
$$[u, v, 1]^T \sim \mathbf{P} \times [X, Y, Z, 1]^T$$

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

Image plane

CCS: camera coordinate system

WCS: world coordinate system



[Peng *et al.*, Pattern Recognition'09]

Camera Calibration

projection matrix

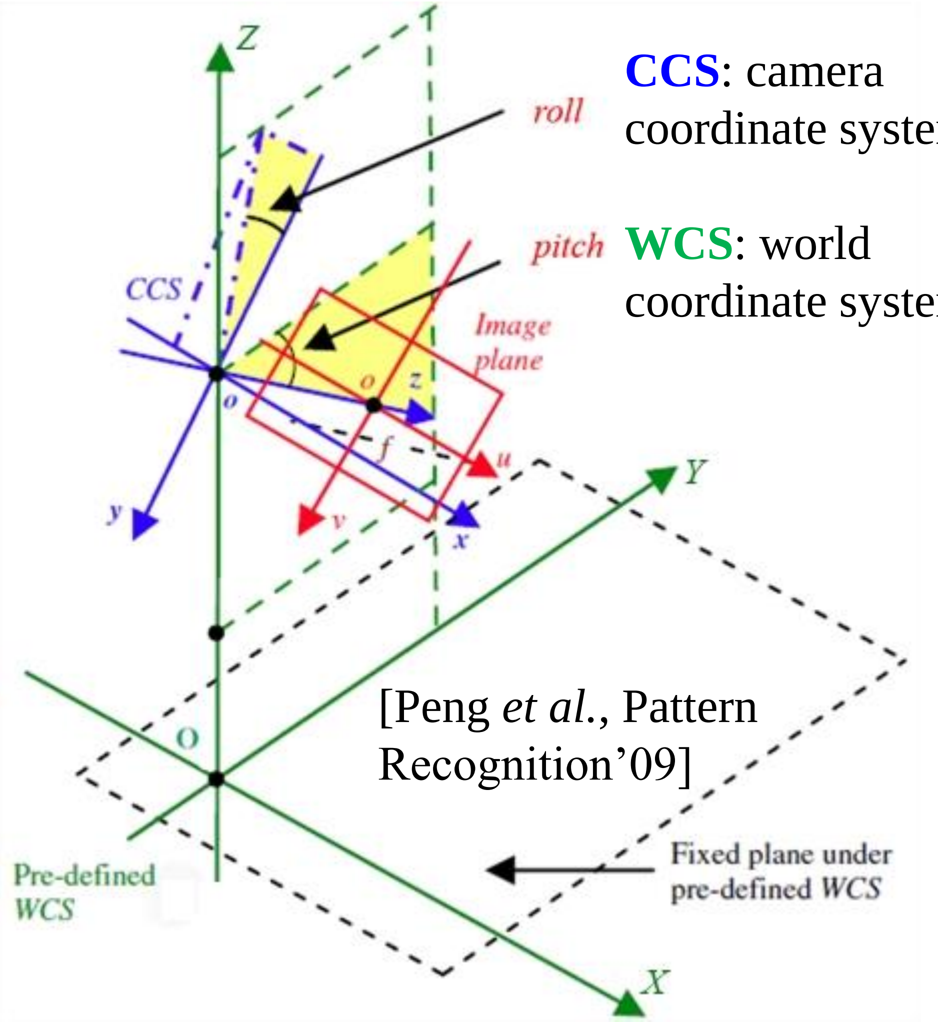
$$[u, v, 1]^T \sim \mathbf{P} \times [X, Y, Z, 1]^T$$

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

intrinsic parameter matrix

$$\mathbf{K} = \begin{bmatrix} f_u & s & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix}$$

Image plane



CCS: camera coordinate system

WCS: world coordinate system

[Peng *et al.*, Pattern Recognition'09]

Camera Calibration

projection matrix

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

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

intrinsic parameter matrix

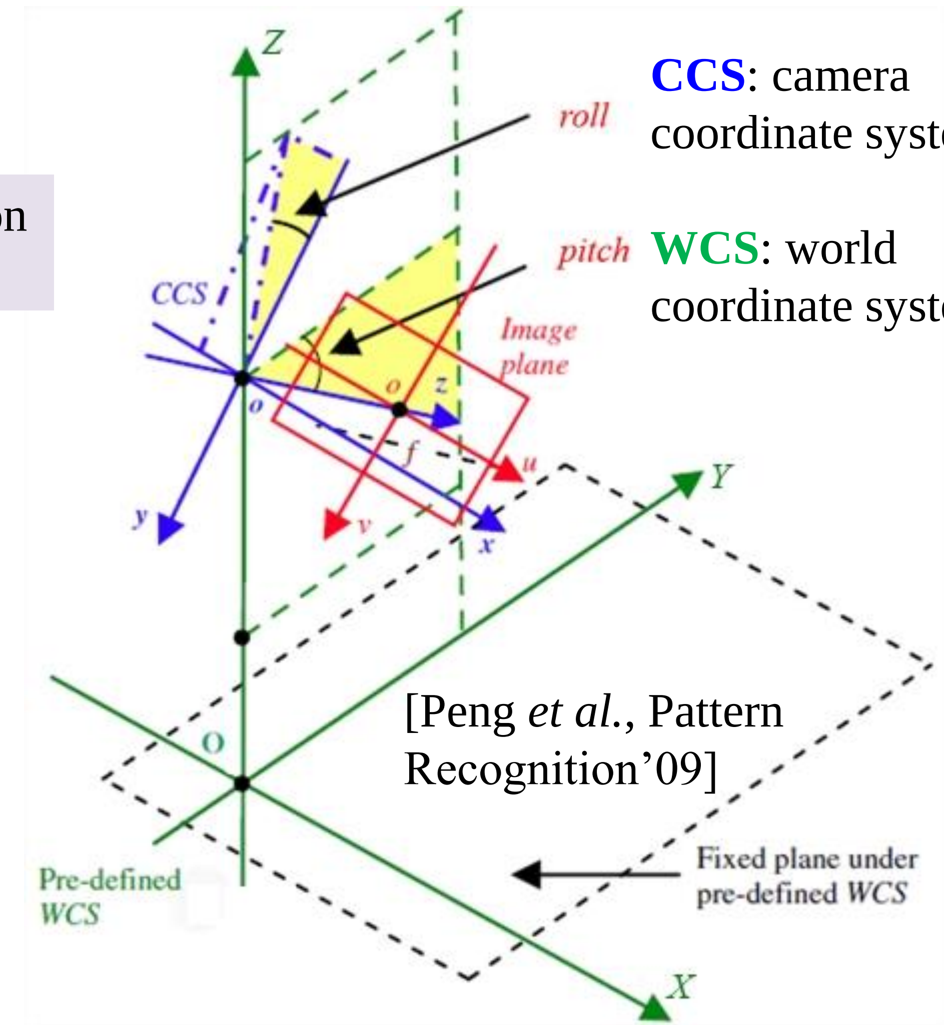
translation matrix

$$\mathbf{K} = \begin{bmatrix} f_u & s & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{t} = \begin{bmatrix} t_X \\ t_Y \\ t_Z \end{bmatrix}$$

Image plane

CCS: camera coordinate system

WCS: world coordinate system



[Peng *et al.*, Pattern Recognition'09]

Camera Calibration

projection matrix

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

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

intrinsic parameter matrix

translation matrix

$$\mathbf{K} = \begin{bmatrix} f_u & s & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{t} = \begin{bmatrix} t_X \\ t_Y \\ t_Z \end{bmatrix}$$

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

rotation matrix

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

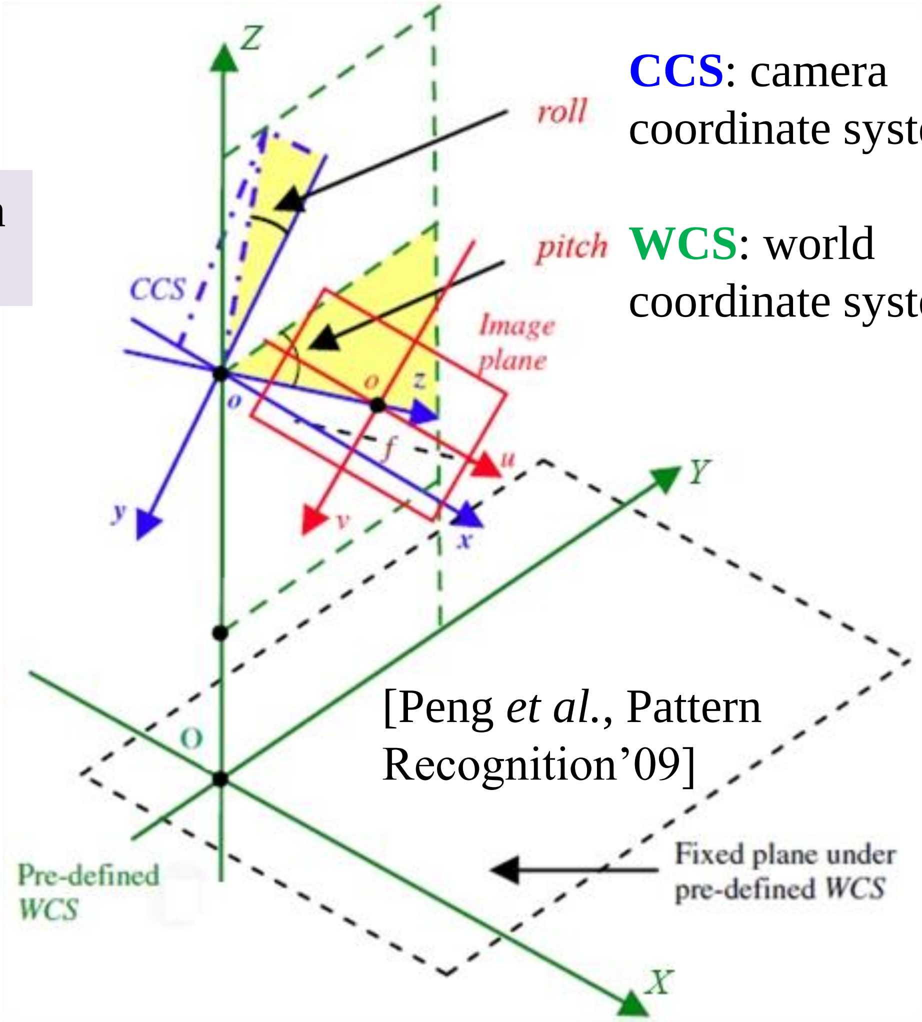
$$\mathbf{R}_Y = \begin{bmatrix} \cos(\alpha) & 0 & -\sin(\alpha) \\ 0 & 1 & 0 \\ \sin(\alpha) & 0 & \cos(\alpha) \end{bmatrix}$$

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

Image plane

CCS: camera coordinate system

WCS: world coordinate system



[Peng *et al.*, Pattern Recognition'09]

Camera Calibration

projection matrix

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

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

intrinsic parameter matrix

translation matrix

$$\mathbf{K} = \begin{bmatrix} f_u & s & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{t} = \begin{bmatrix} t_X \\ t_Y \\ t_Z \end{bmatrix}$$

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

rotation matrix

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

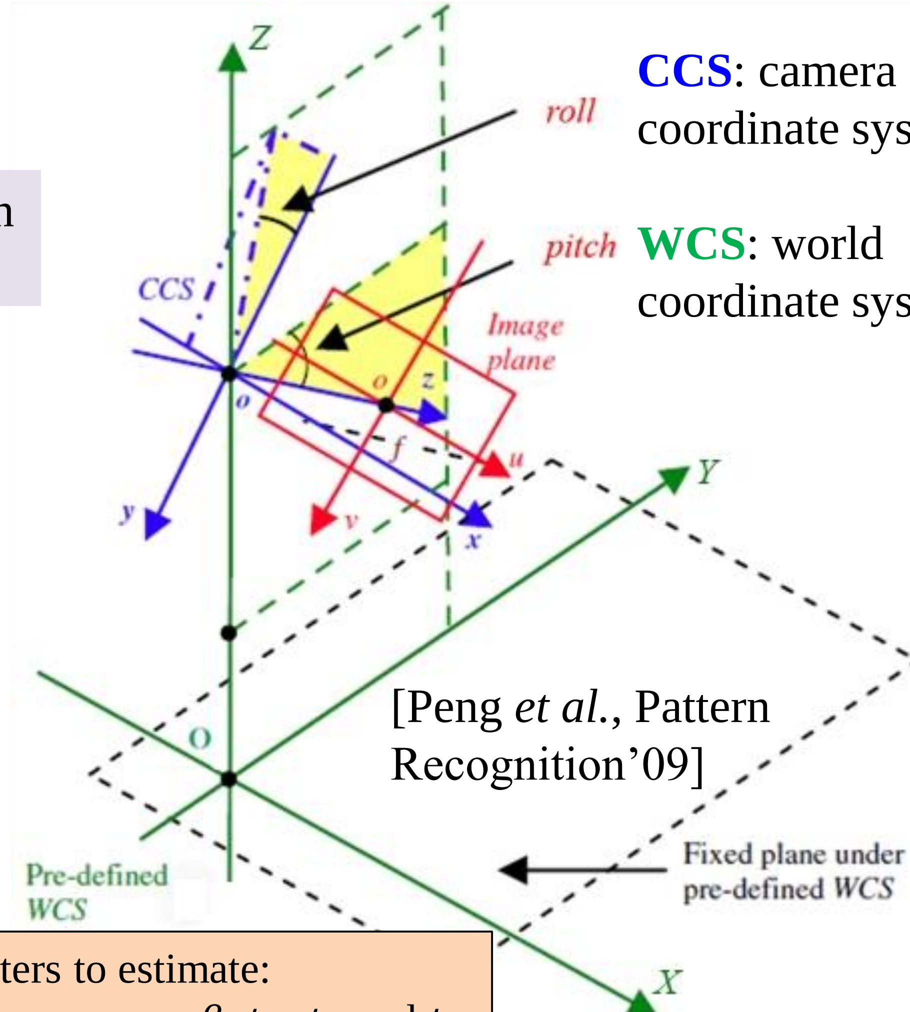
$$\mathbf{R}_Y = \begin{bmatrix} \cos(\alpha) & 0 & -\sin(\alpha) \\ 0 & 1 & 0 \\ \sin(\alpha) & 0 & \cos(\alpha) \end{bmatrix}$$

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

Image plane

CCS: camera coordinate system

WCS: world coordinate system

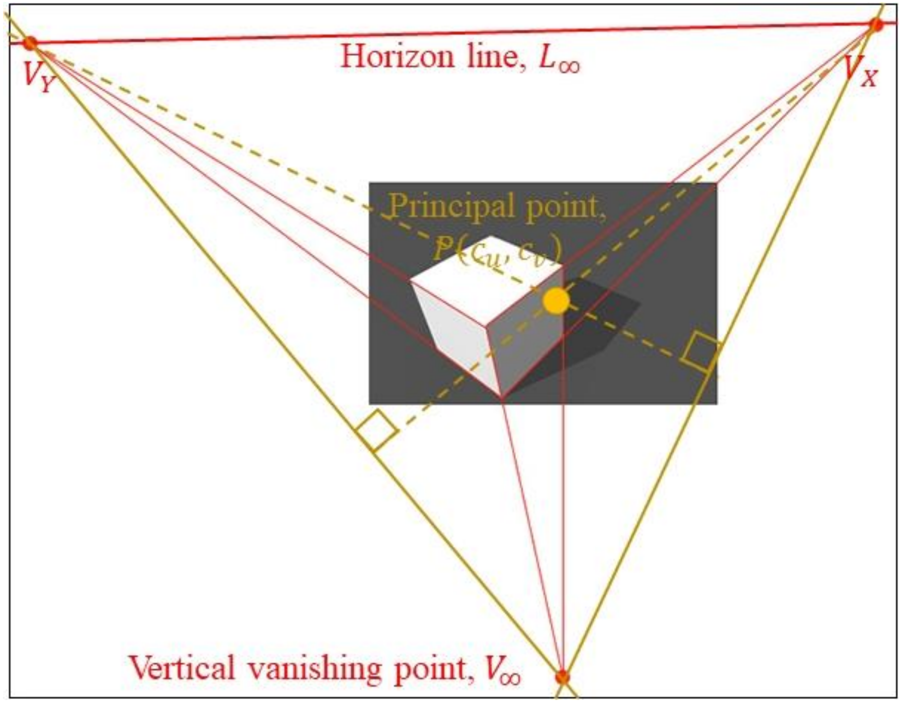


[Peng *et al.*, Pattern Recognition'09]

11 parameters to estimate:
$f_u, f_v, c_u, c_v, s, \gamma, \alpha, \beta, t_X, t_Y$ and t_Z

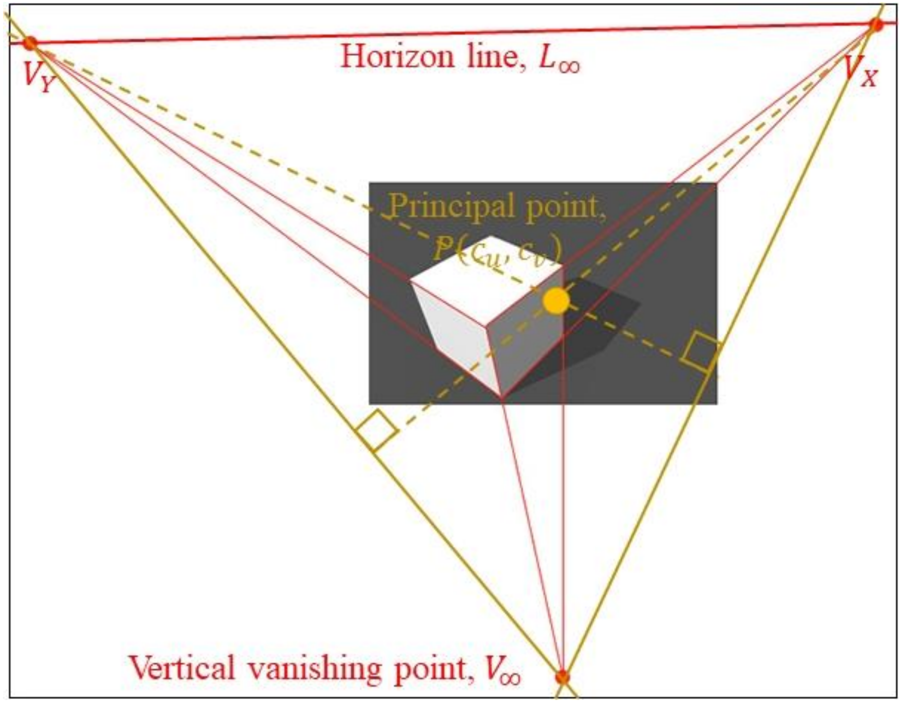
Camera Calibration

- Calibration using calibrated templates
  - Cube



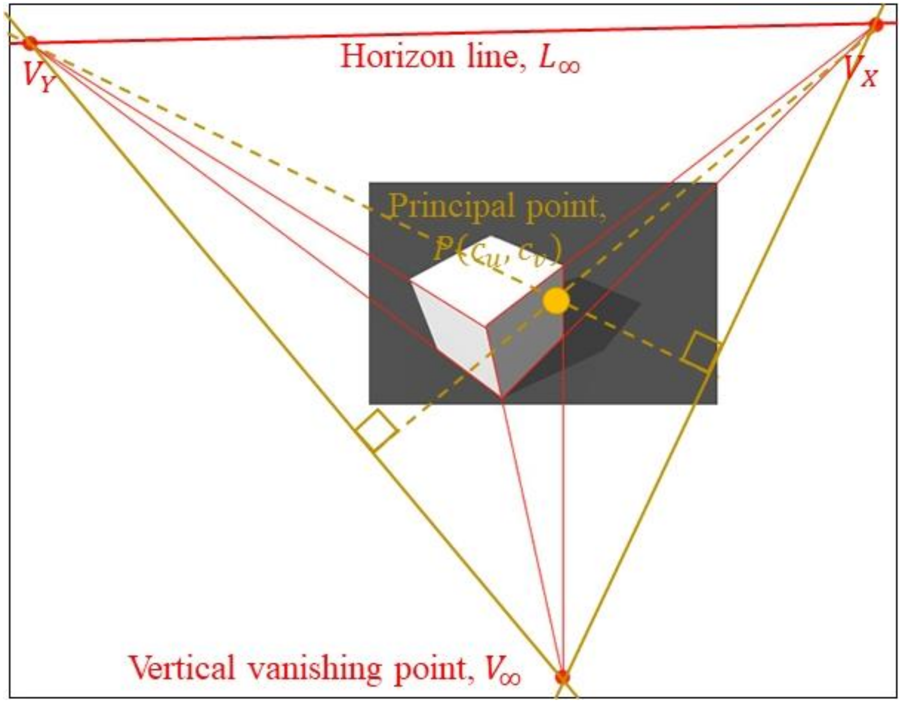
Camera Calibration

- Calibration using calibrated templates
  - Cube
- Self-calibration
  - Static scene structures
    - Manhattan world assumption (MWA)



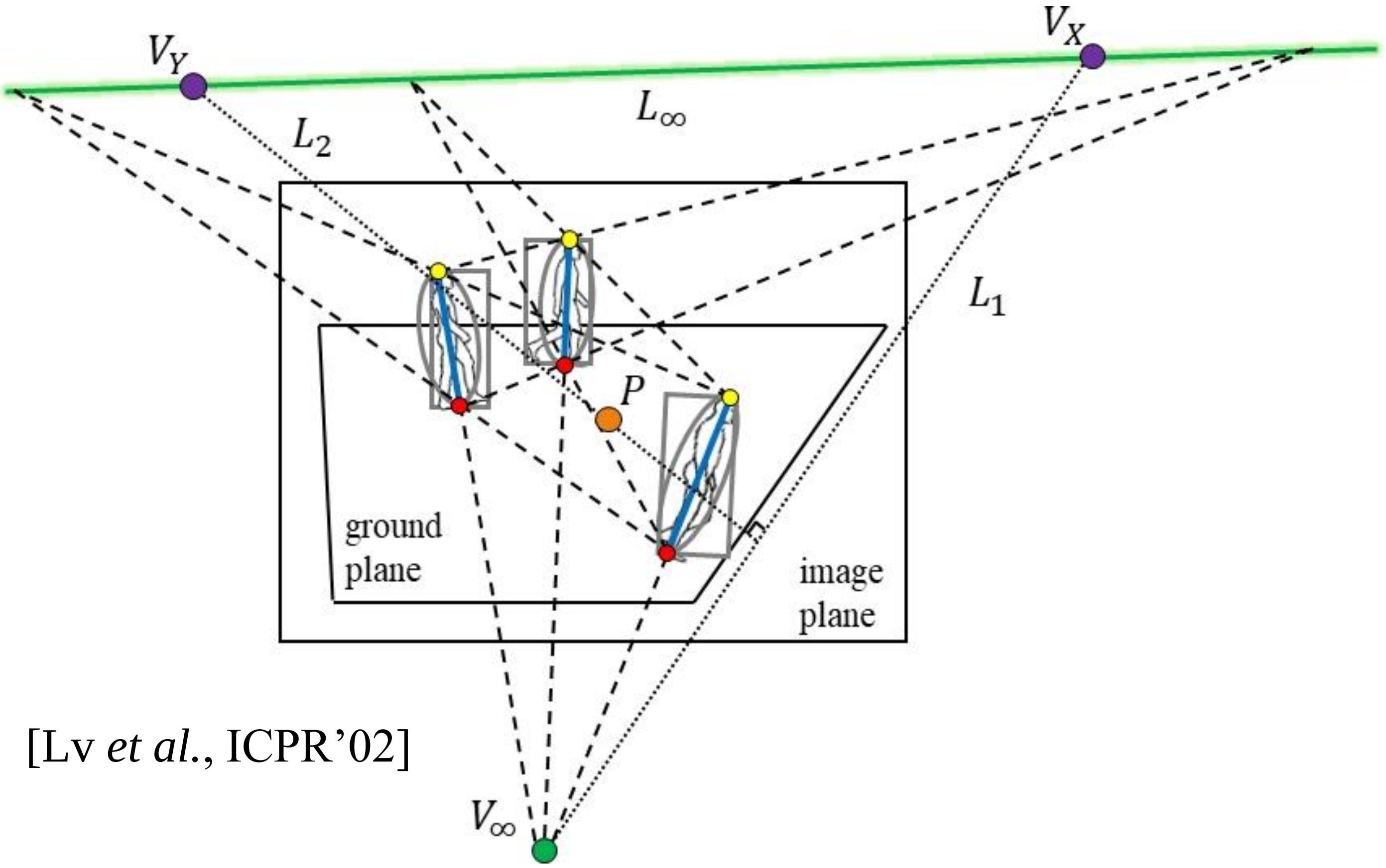
Camera Calibration

- Calibration using calibrated templates
  - Cube
- Self-calibration
  - Static scene structures
    - Manhattan world assumption (MWA)
  - Object motion, *e.g.*, **tracking of walking humans**



Camera Calibration

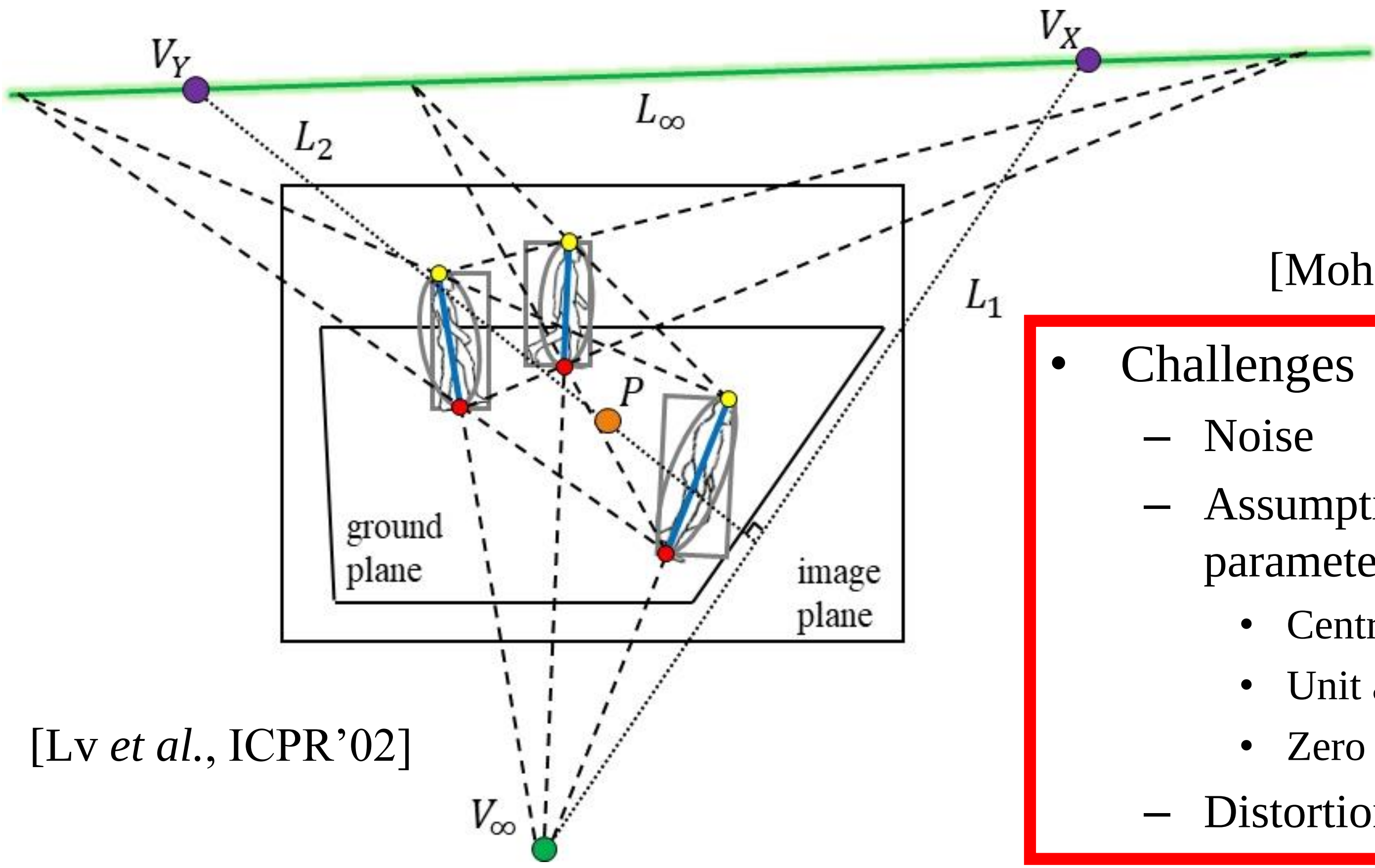
- Self-calibration from human tracking



[Lv *et al.*, ICPR'02]

Camera Calibration

- Self-calibration from human tracking



[Lv *et al.*, ICPR'02]

[Mohedano *et al.*, ICIP'10]

- Challenges
  - Noise
  - Assumptions (only 7 of the 11 parameters can be estimated)
    - Central principal point
    - Unit aspect ratio
    - Zero skew
  - Distortion

Camera Calibration

radial distortion coefficients

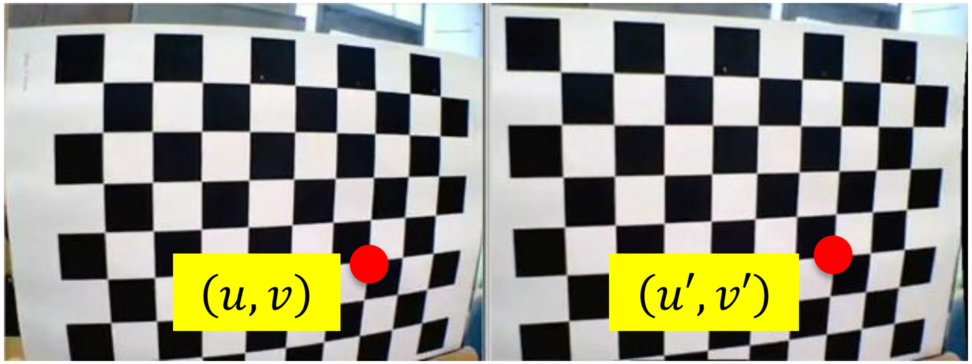
$$(u, v) \xrightarrow{\mathbf{k} = [k_1, k_2, k_3]^T} (u', v')$$

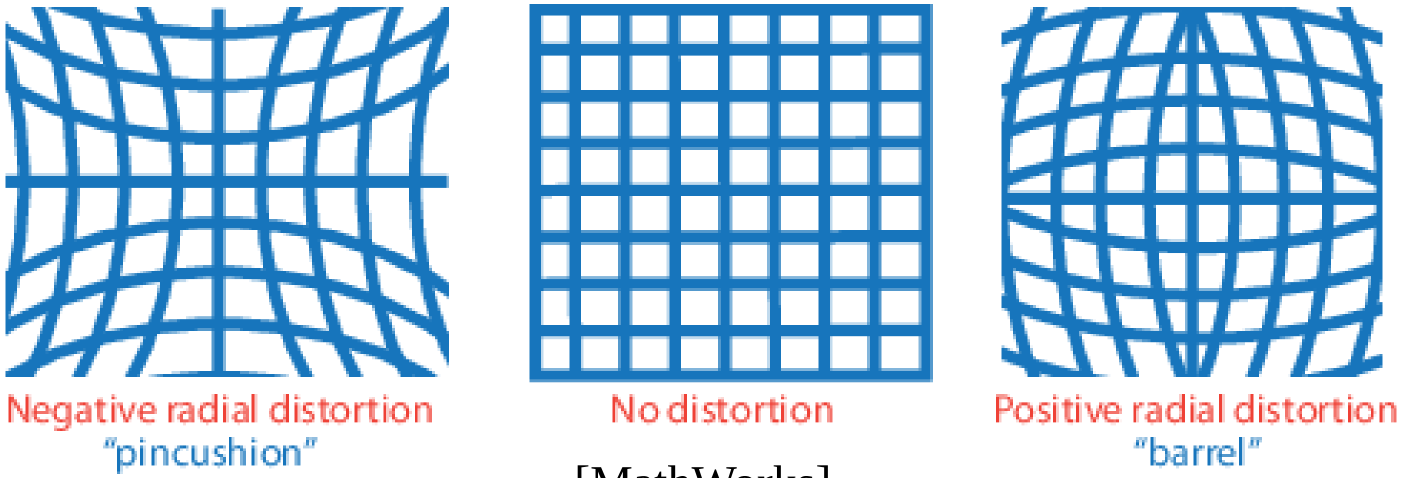
$$u' = u(1 + k_1 r^2 + k_2 r^4 + k_3 r^6)$$

$$v' = v(1 + k_1 r^2 + k_2 r^4 + k_3 r^6)$$

$$\text{s.t.}, r^2 = u^2 + v^2$$

[Bersoft Image Measurement]

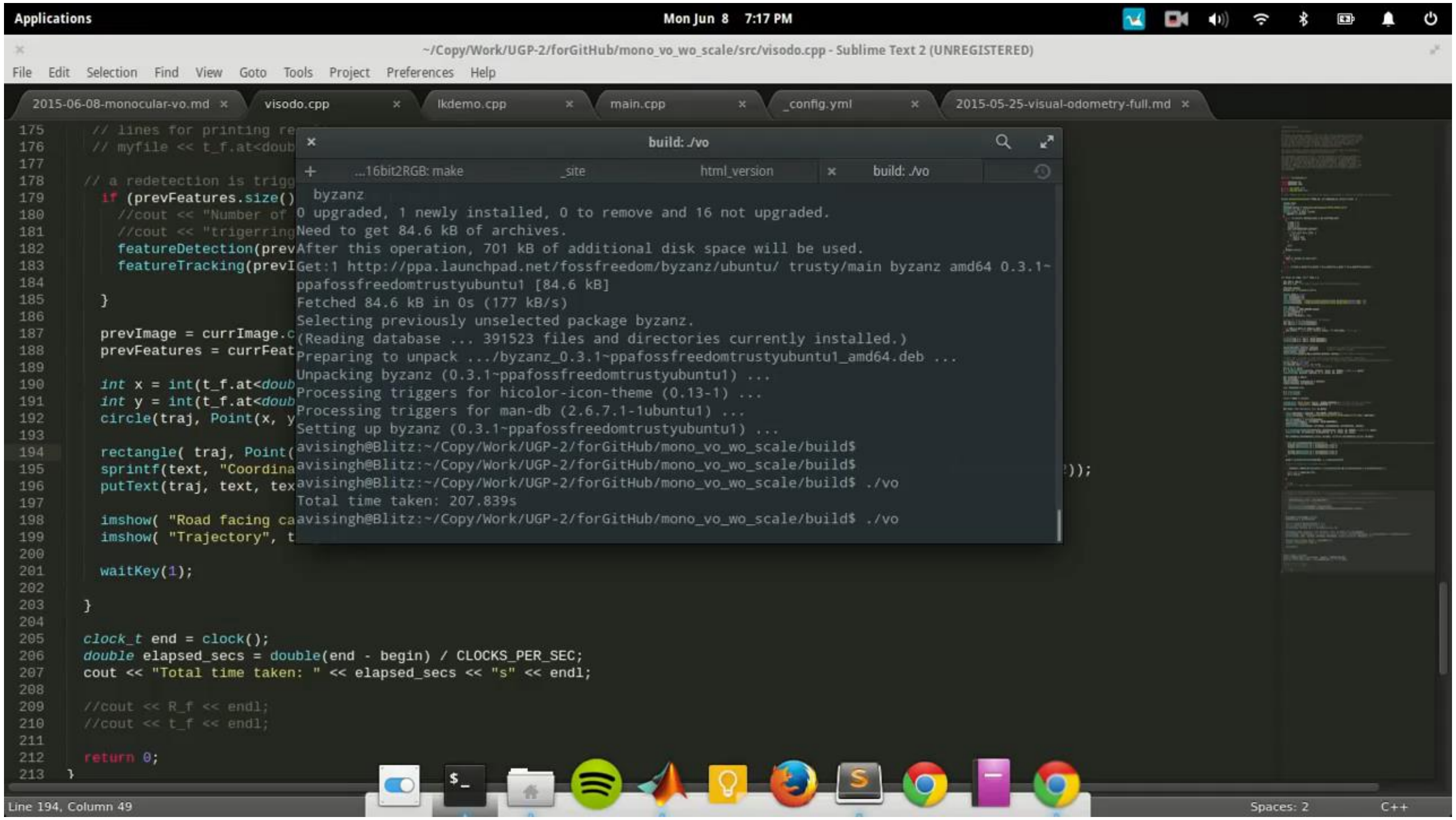




[MathWorks]

Camera Calibration

- Visual odometry for a moving camera

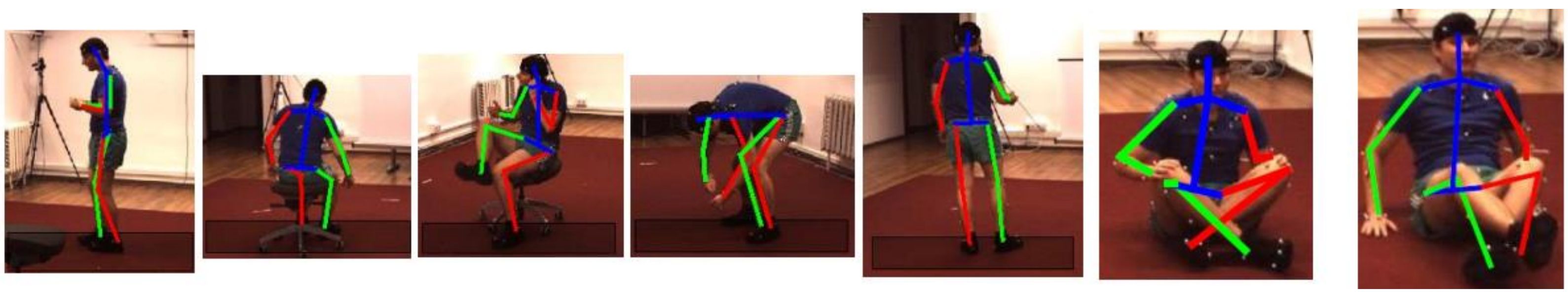


[Avi Singh's blog]

Pose Estimation

- Pose estimation in 2D

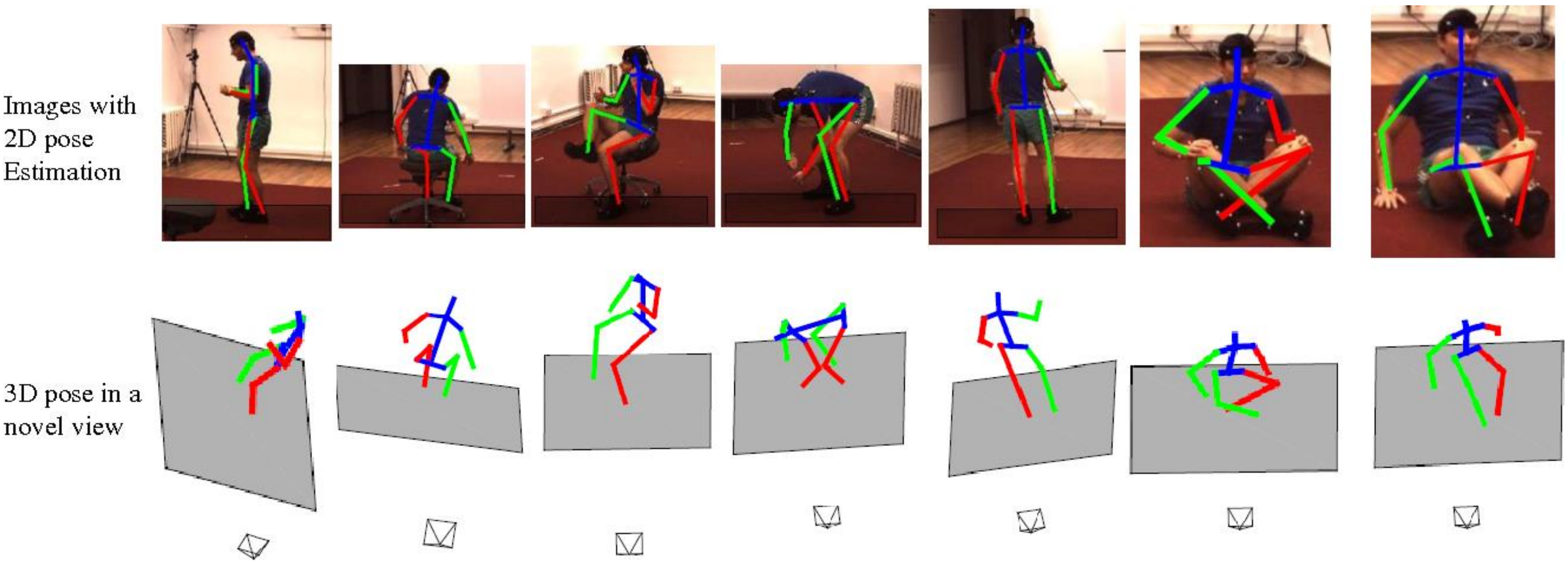
Images with
2D pose
Estimation



[Chen *et al.*, CVPR'17]

Pose Estimation

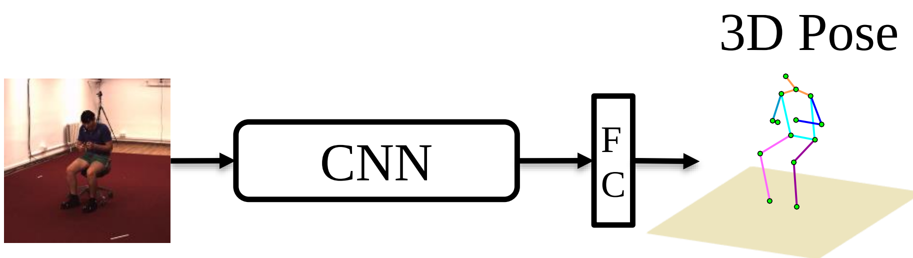
- Pose estimation in 2D
- Pose estimation in 3D



[Chen *et al.*, CVPR'17]

Pose Estimation

- One-stage (end-to-end) 3D pose estimation

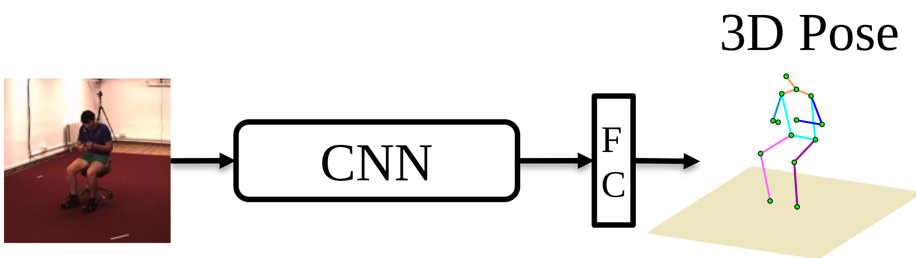


- Limitations
  - Prone to overfitting
  - Relative 3D pose

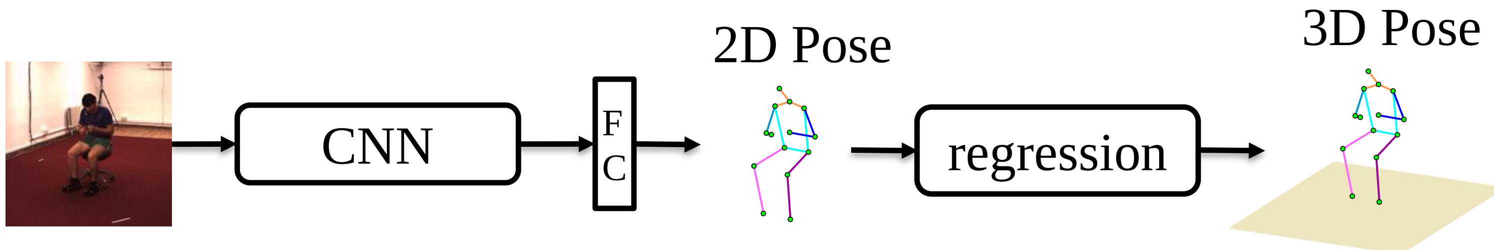
[Iqbal *et al.*, ECCV’18]

Pose Estimation

- One-stage (end-to-end) 3D pose estimation
- Two-stage 3D pose estimation



- Limitations
  - Prone to overfitting
  - Relative 3D pose

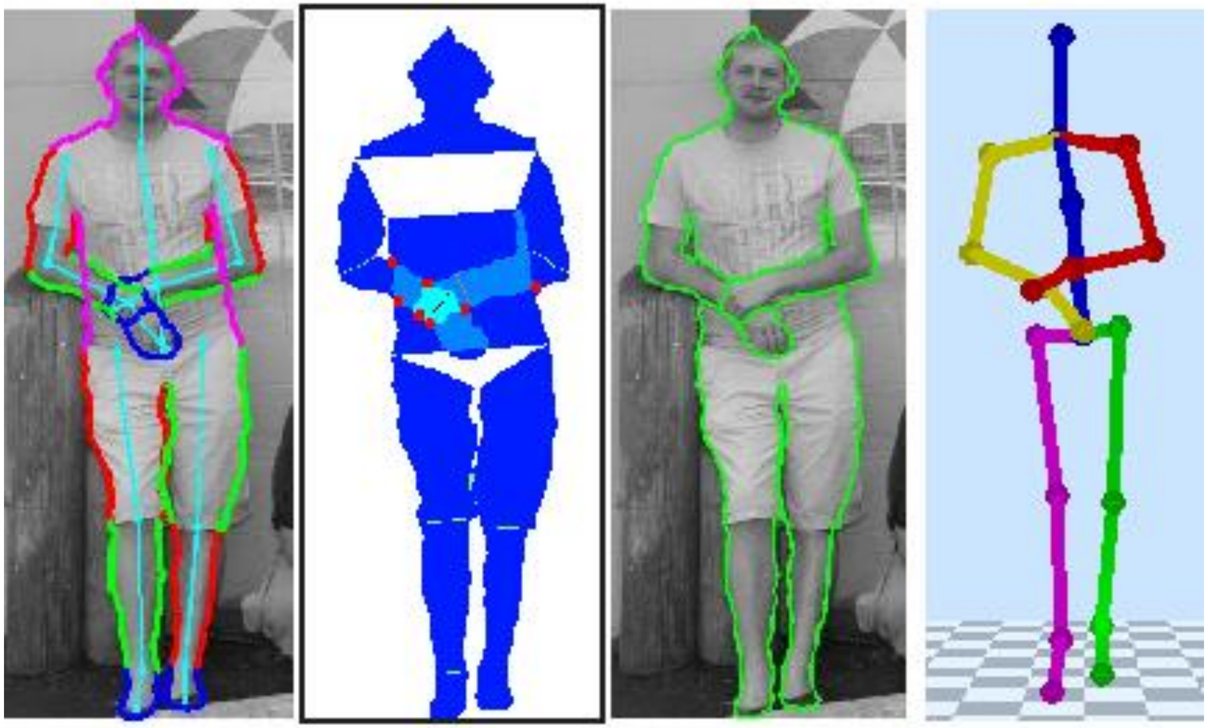


[Iqbal *et al.*, ECCV’18]

Pose Estimation

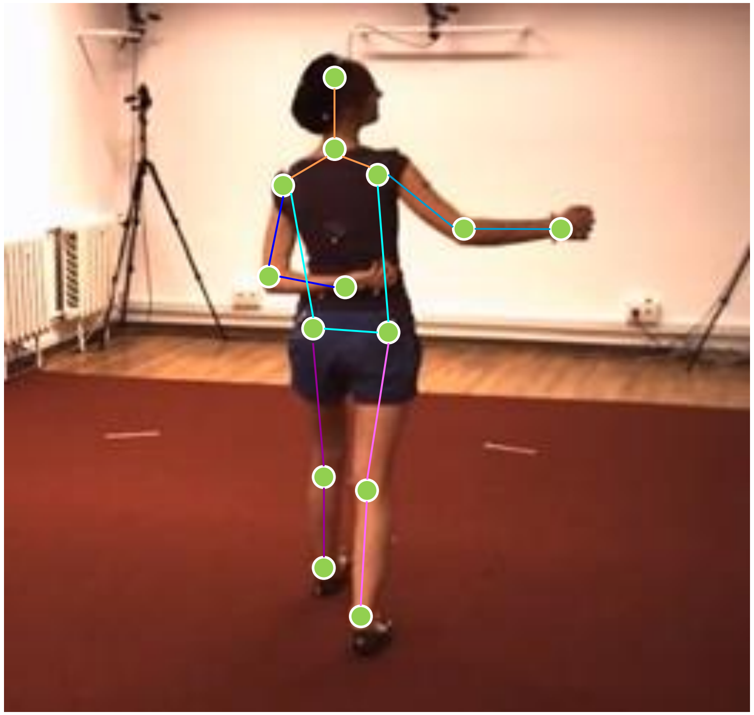
- Challenges
  - Self-occlusion

[Jacques *et al.*, ICIP'13]



Pose Estimation

- Challenges
  - Projection ambiguity



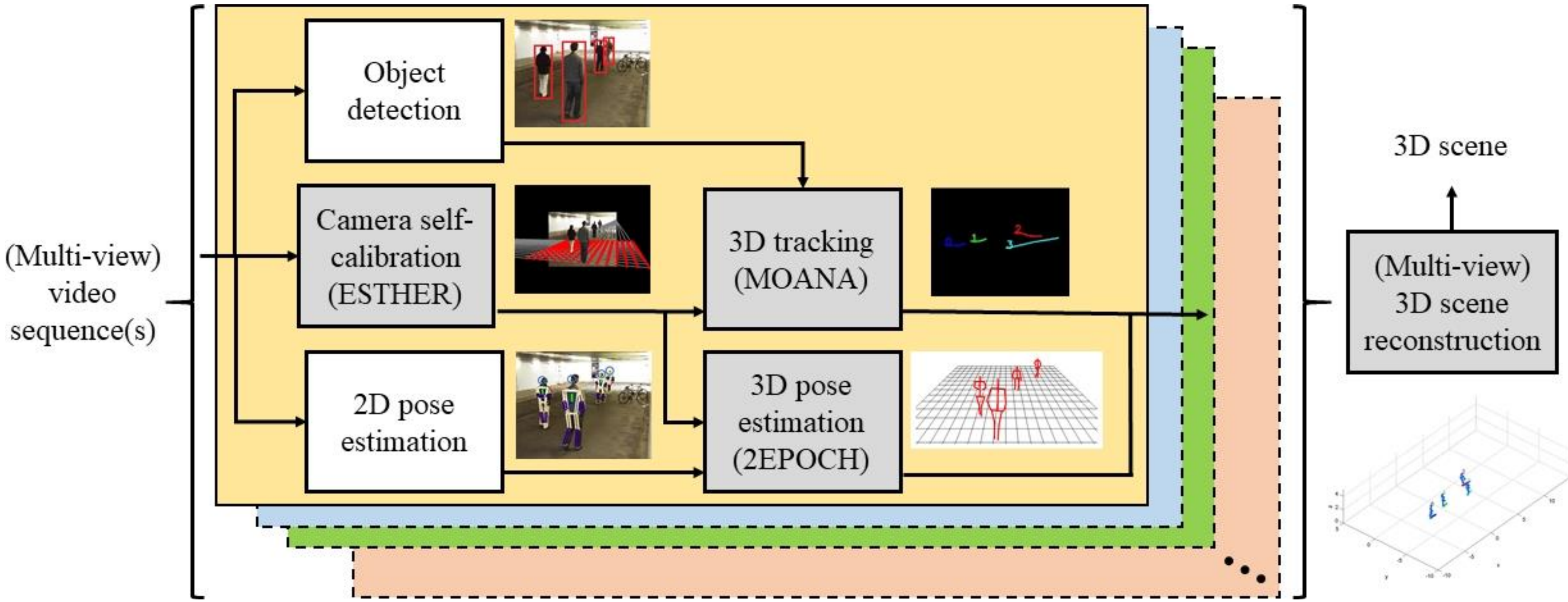
[Iqbal *et al.*, ECCV'18]

Pose Estimation

- Challenges
  - Ambiguity between objects

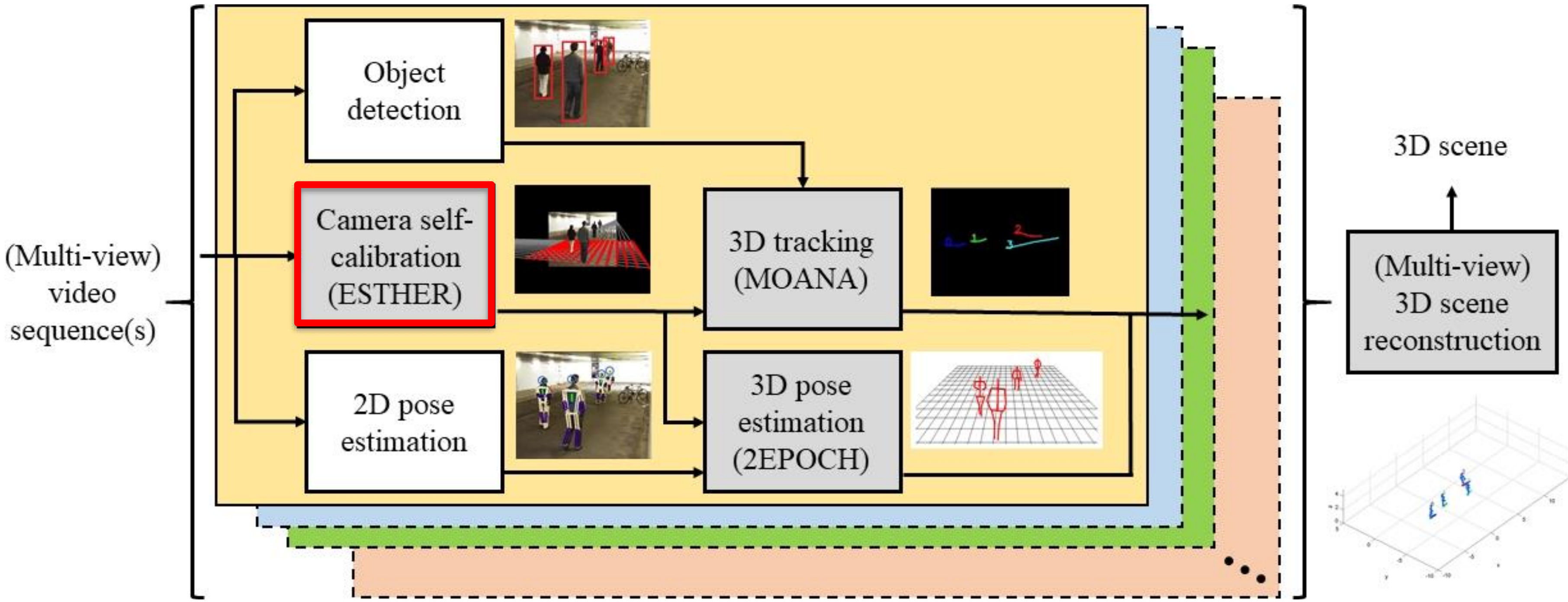
[Pishchulin *et al.*, CVPR'16]

Outline



- **ESTHER**: Evolutionary Self-calibration from Tracking of Humans for Enhancing Robustness
- **MOANA**: Modeling of Object Appearance by Normalized Adaptation
- **2EPOCH**: Two-step Evolutionary Pose Optimization for Camera and Humans
- Extension to multi-view 3D scene reconstruction

Outline

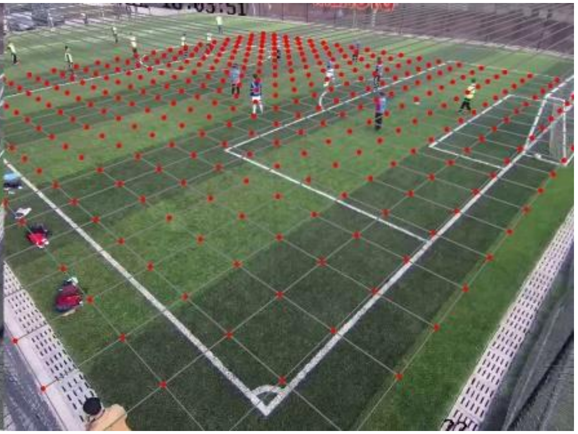


- **ESTHER**: Evolutionary Self-calibration from Tracking of Humans for Enhancing Robustness
- **MOANA**: Modeling of Object Appearance by Normalized Adaptation
- **2EPOCH**: Two-step Evolutionary Pose Optimization for Camera and Humans
- Extension to multi-view 3D scene reconstruction

Camera Self-calibration

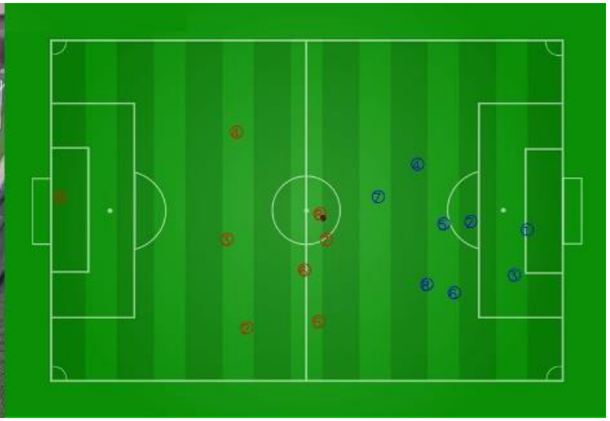
Input video frame

Radial distortion correction



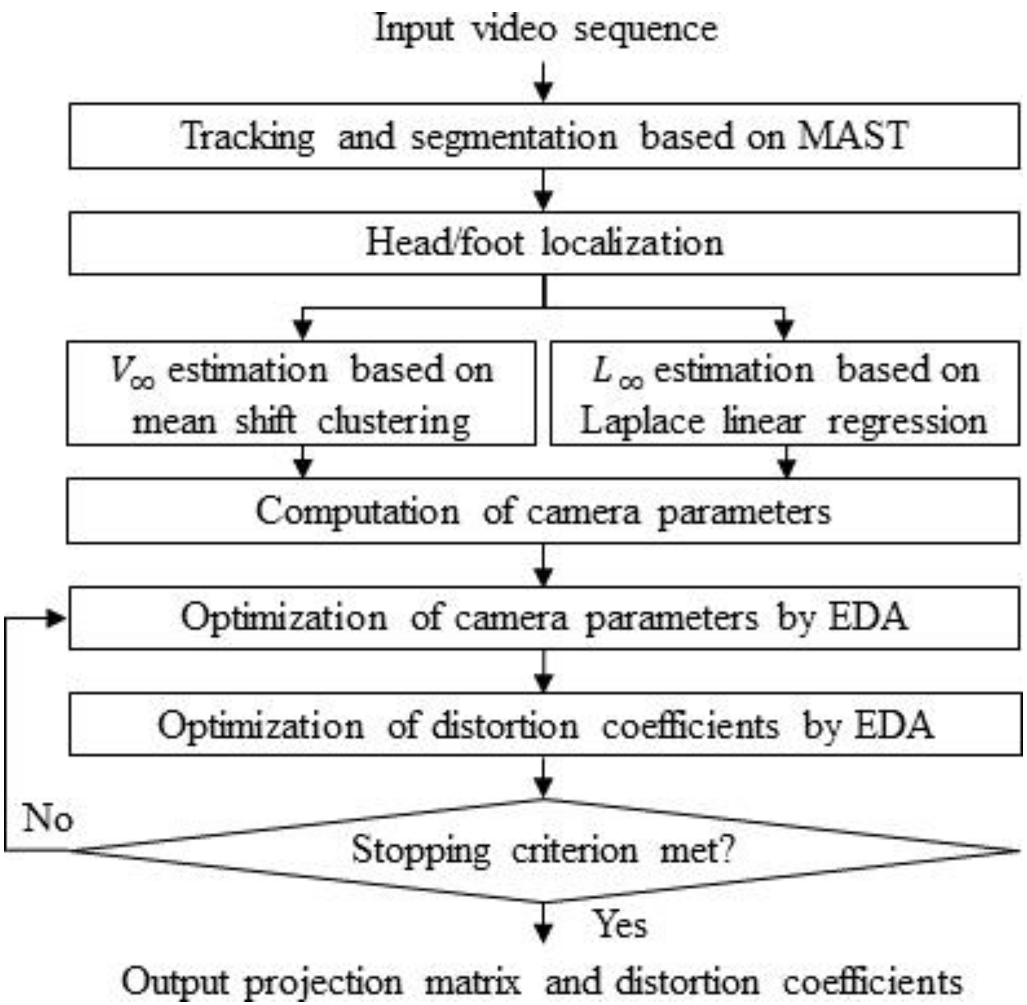
Camera self-calibration

2D tracking

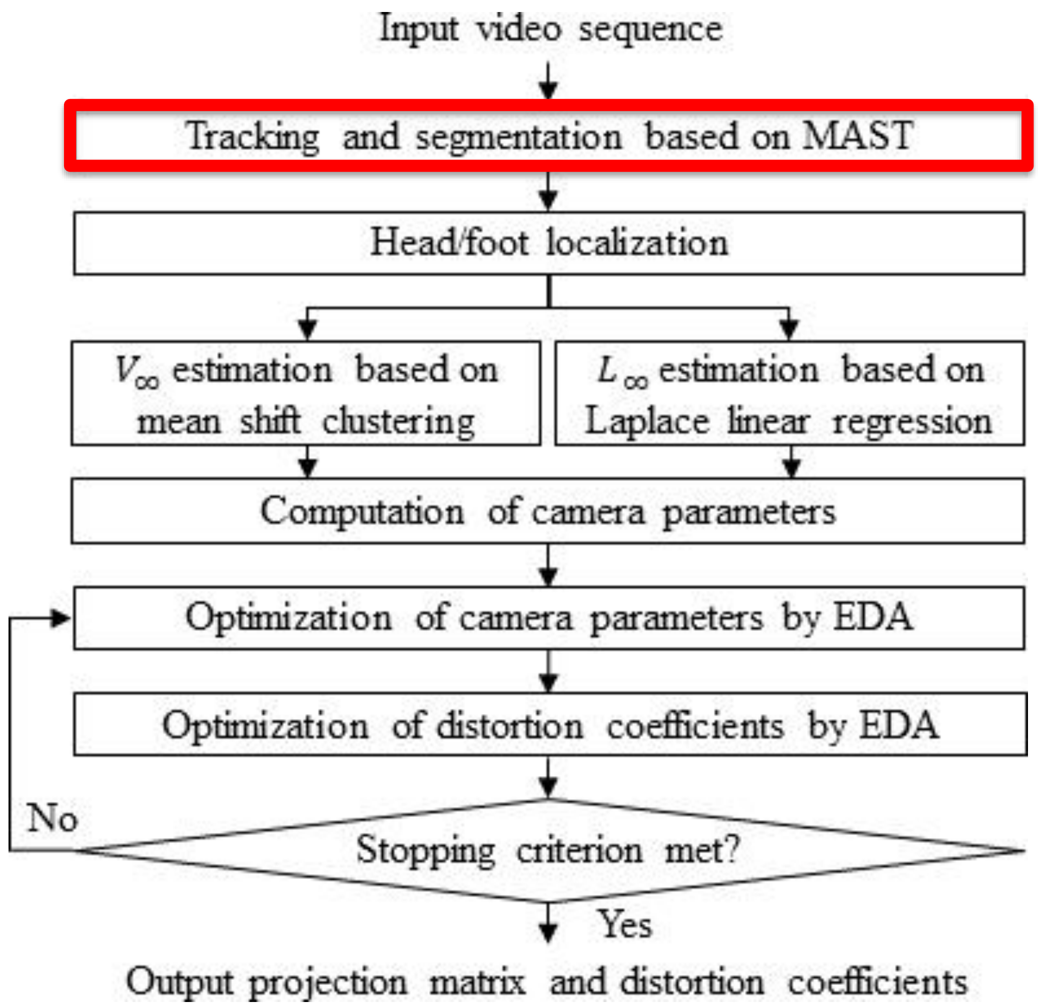


3D tracking based on calibration

Camera Self-calibration

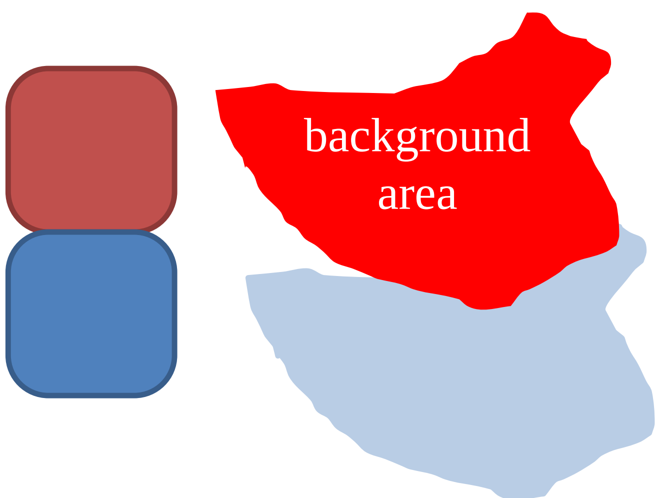


Camera Self-calibration



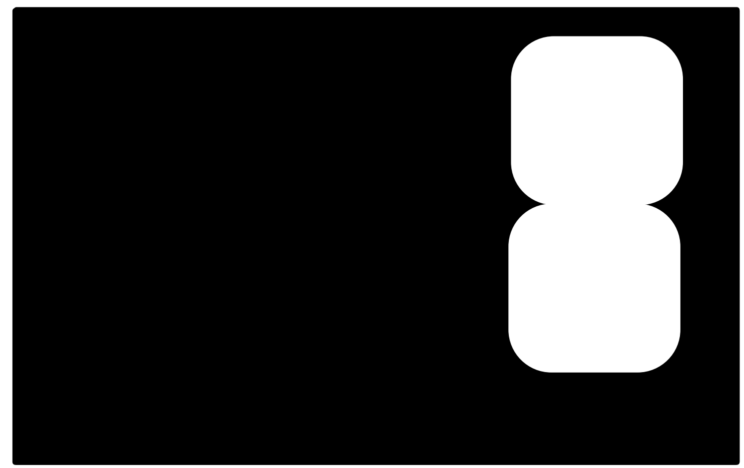
Camera Self-calibration

- **MAST**: Multi-kernel Adaptive Segmentation and Tracking

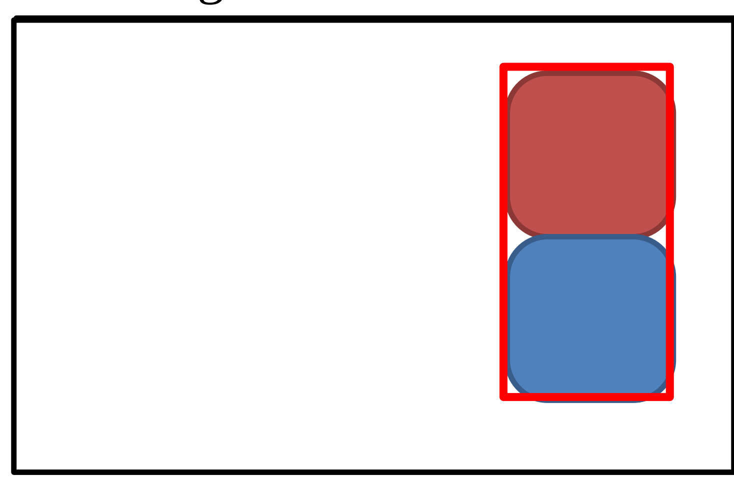


[Tang *et al.*, ICASSP'16]

Segmentation results



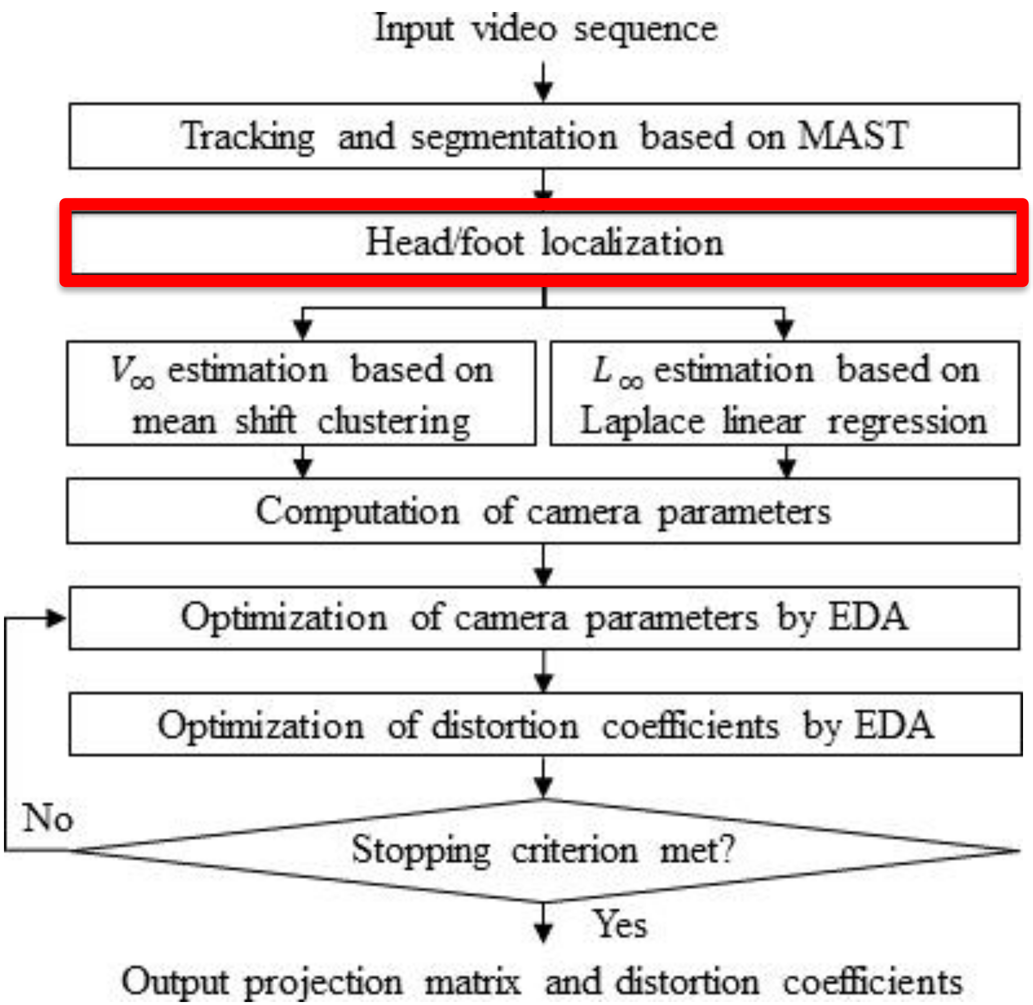
Tracking results



Camera Self-calibration

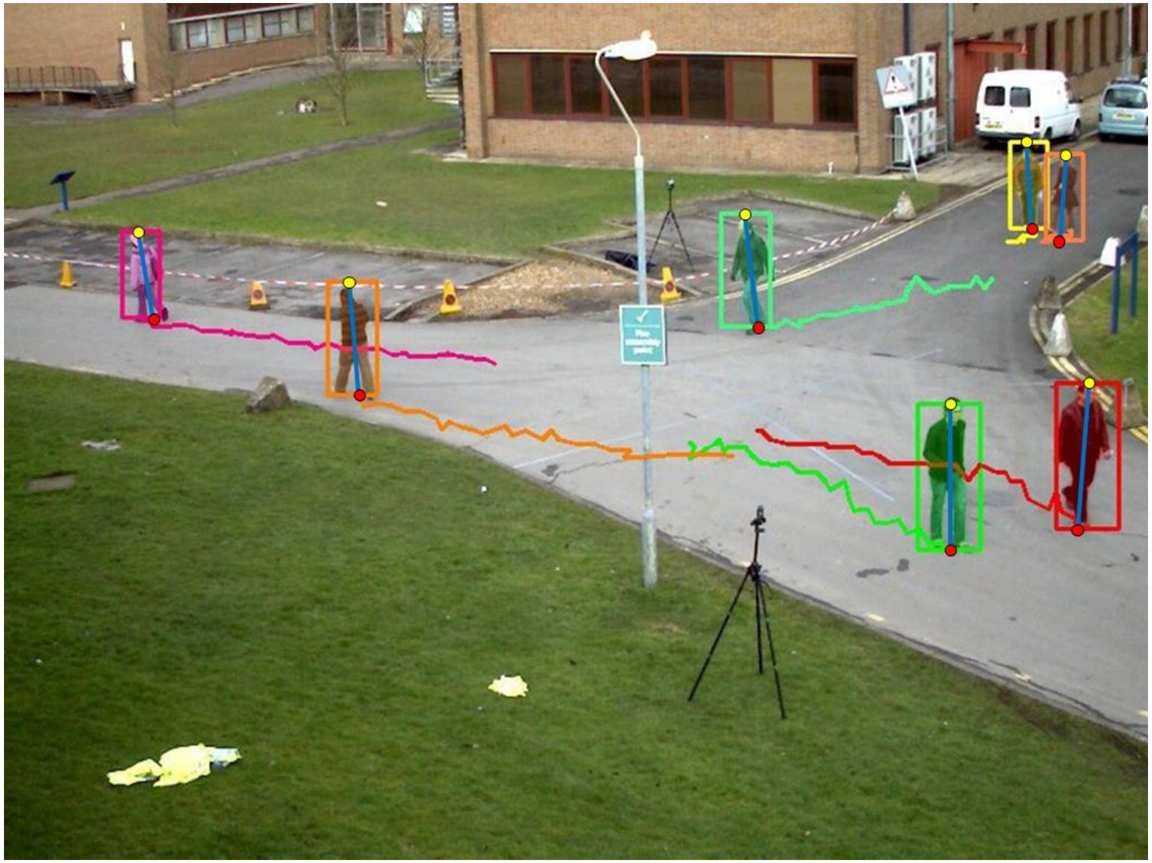
- MAST for tracking by segmentation

Camera Self-calibration

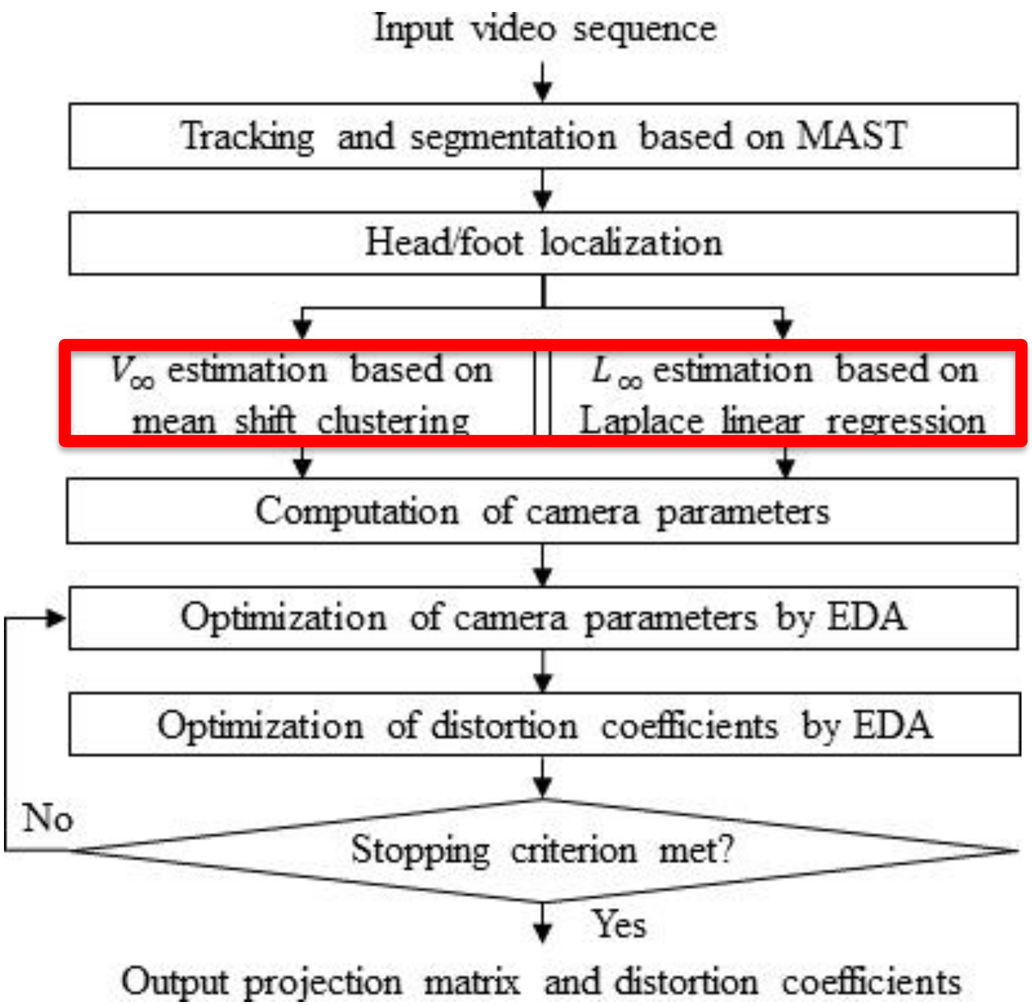


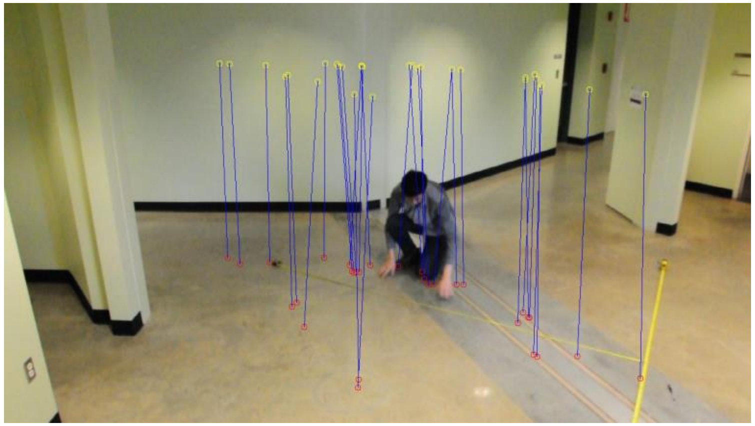
Camera Self-calibration

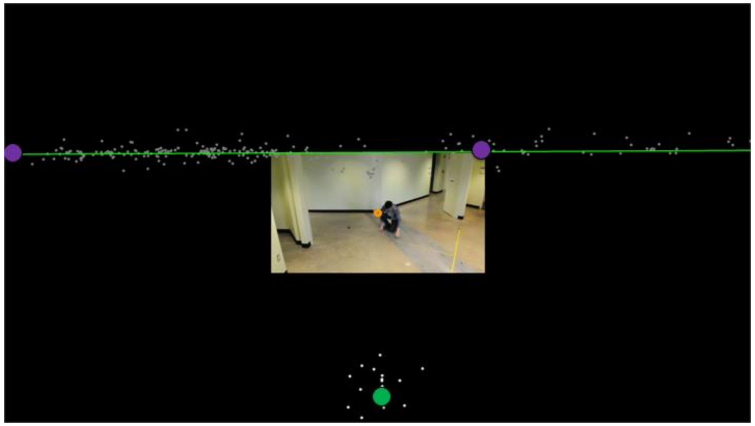
- Head/foot localization



Camera Self-calibration

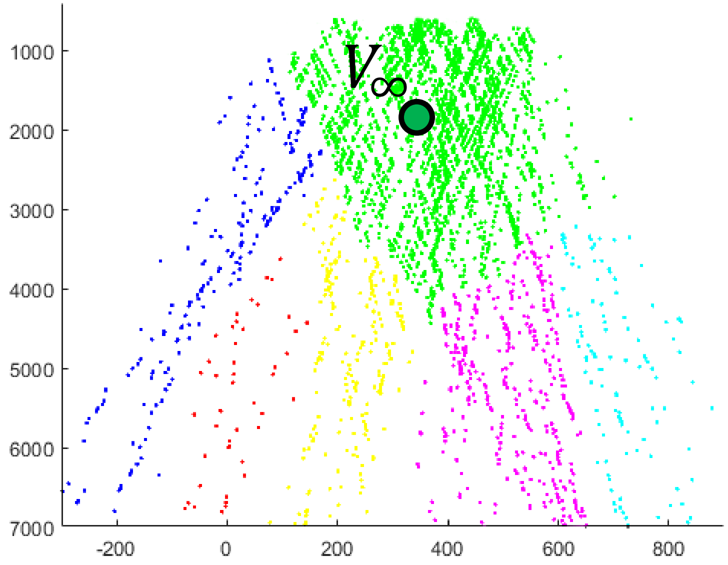






Camera Self-calibration

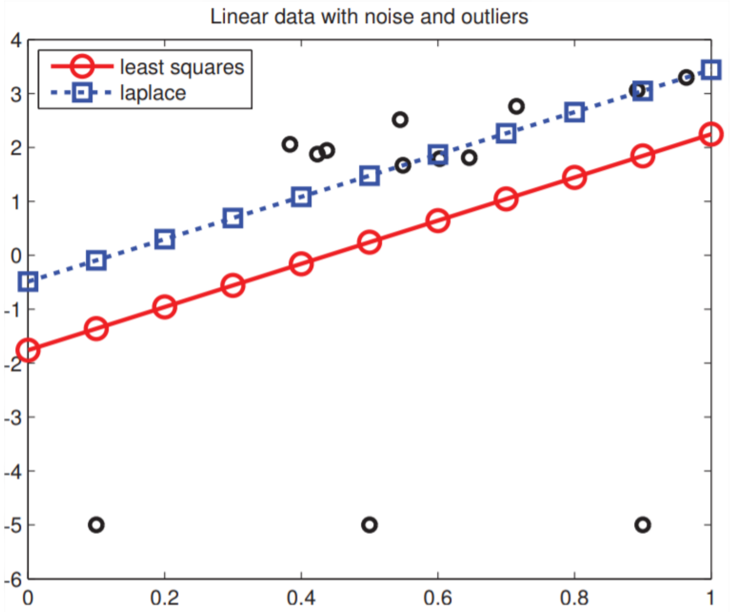
- V_∞ estimation based on mean shift clustering
  - Limitation of RANSAC
    - Cannot handle large number of outliers
  - Proposed method
    - Mean shift clustering for all candidates
    - Locating the mean point of the largest cluster



Camera Self-calibration

- L_∞ estimation based on Laplace linear regression
  - Limitation of RANSAC
    - Threshold parameter for inliers
  - Proposed method
    - Formulation as Laplace linear regression

[Machine Learning: A Probabilistic Perspective]

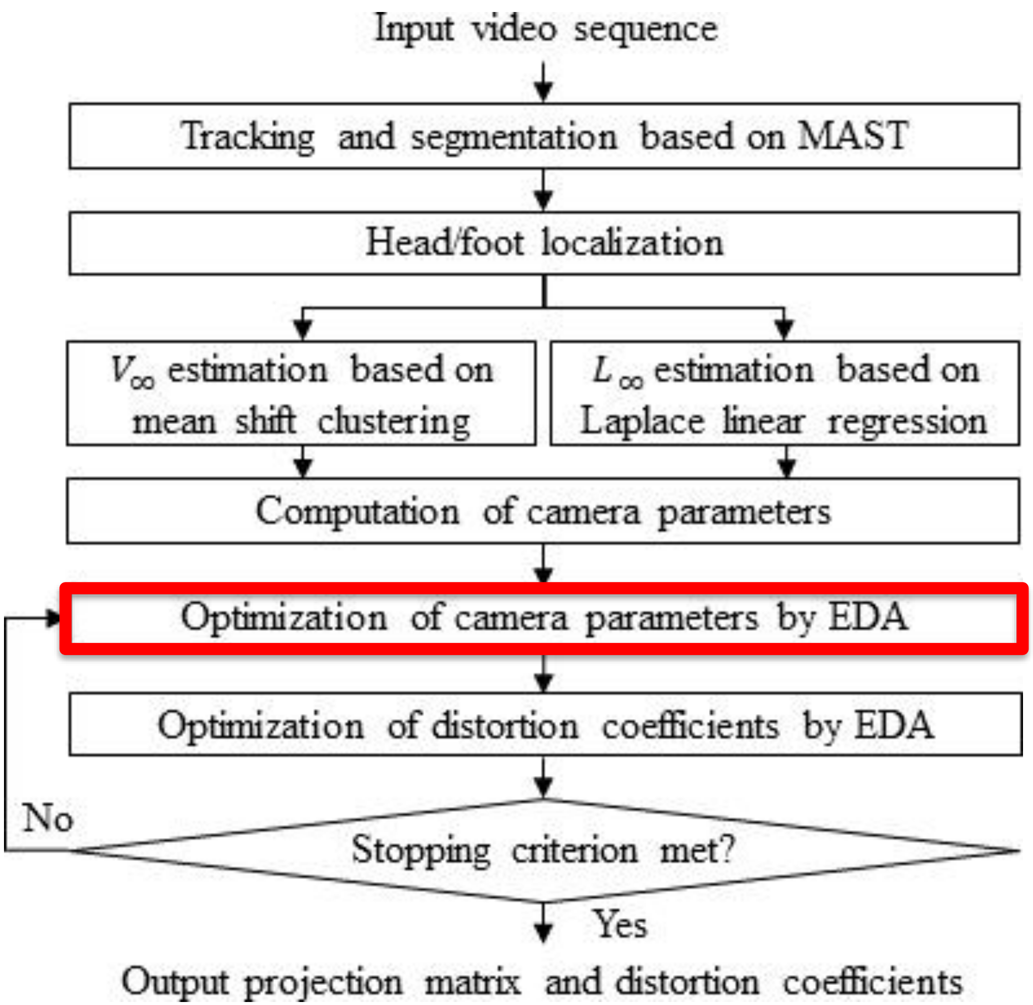


$$\mathrm{Laplace}(\mathbf{v}|\mathbf{w}^T\mathbf{u}) \propto \exp(-|\mathbf{v}-\mathbf{w}^T\mathbf{u}|)$$

$$\mathrm{Gaussian}(\mathbf{v}|\mathbf{w}^T\mathbf{u}) \propto \exp\left(-(\mathbf{v}-\mathbf{w}^T\mathbf{u})^2\right)$$

$(\mathbf{u},\mathbf{v})$: Input candidate points
$\boldsymbol{w}$: Parameters to be estimated

Camera Self-calibration



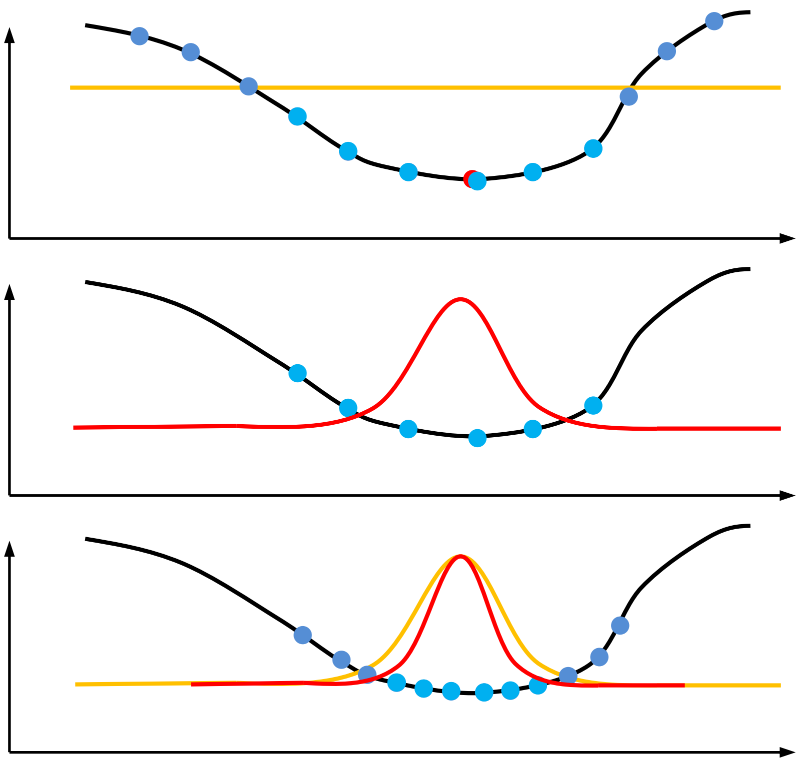
Camera Self-calibration

- Estimation of Distribution Algorithm (EDA)

Objective function: $\arg\min_{\boldsymbol{x}} f(x)$

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.

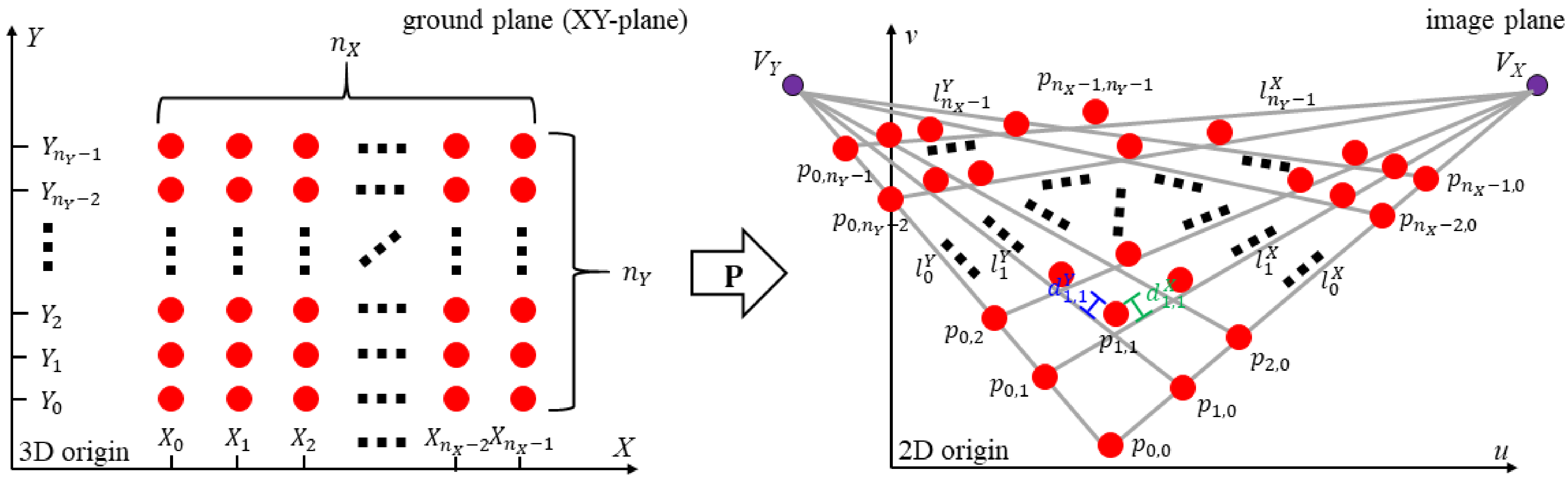
In this example, $R = 12$, $N = 6$



⋮

until stopping criterion is met

Camera Self-calibration



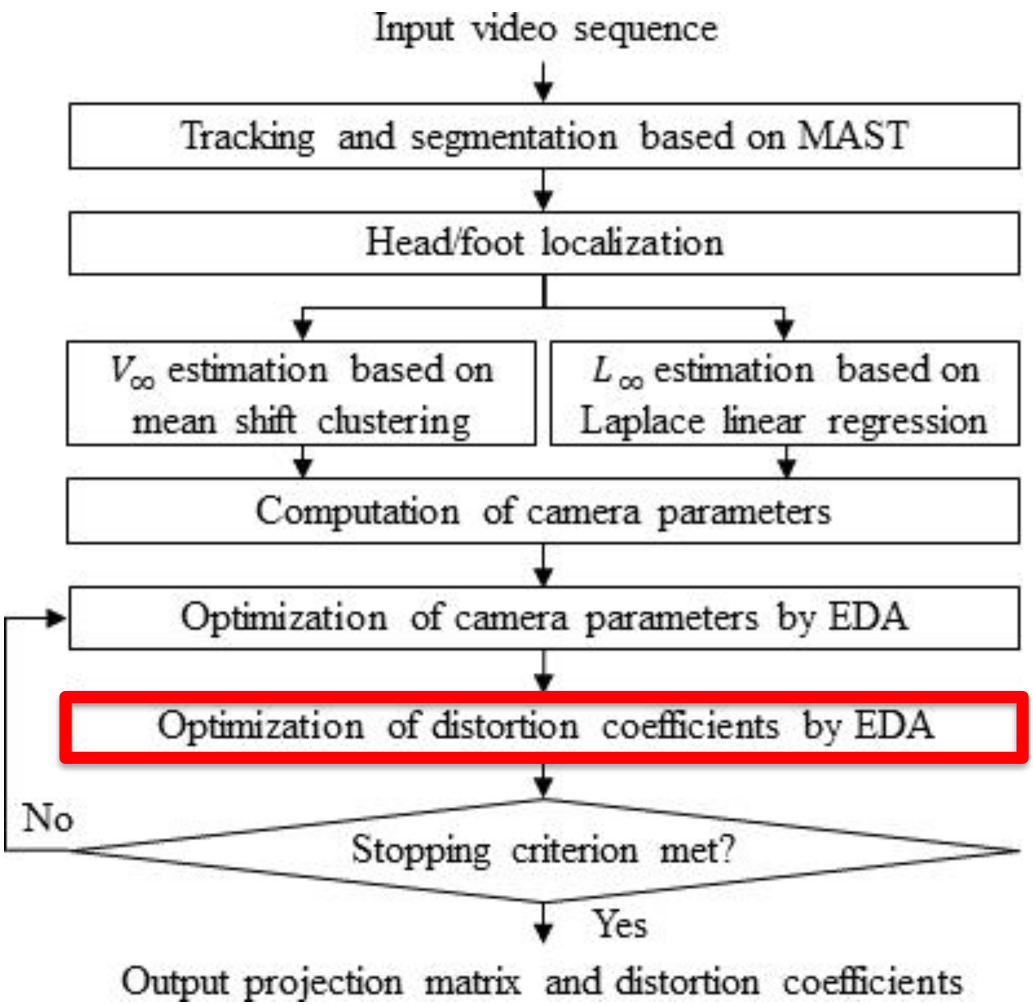
- **Sample**: Projection matrix **P** formed by a set of 11 camera parameters
- **PDF**: 11-variate normal density function
- **Stopping criterion**: Changing ratio between generations smaller than threshold

- **Objective function**: Reprojection error (Distance between projected points and grid lines)

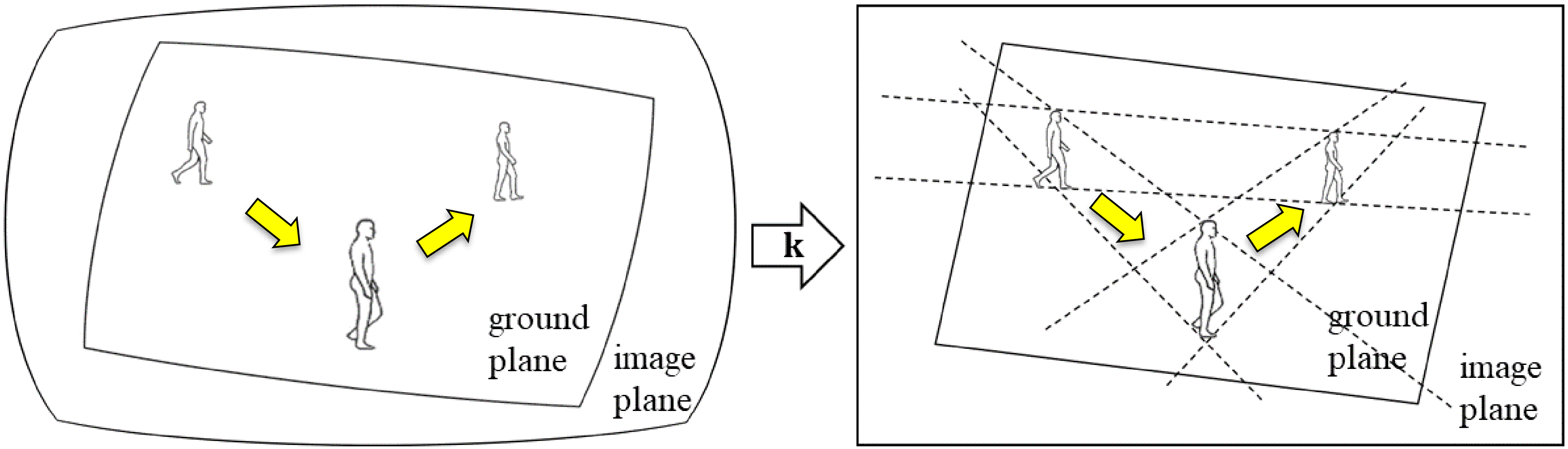
$$\mathbf{P}^* = \underset{\mathbf{P}\in\mathrm{Rng}_\mathbf{P}}{\arg\min}\, \mathrm{E}\left(d^X_{i,j} + d^Y_{i,j}\right)$$

$$\mathrm{s.t.},\ d^X_{i,j} = \left\| l^X_j, p_{i,j} \right\|_2,\ d^Y_{i,j} = \left\| l^Y_i, p_{i,j} \right\|_2$$

Camera Self-calibration



Camera Self-calibration



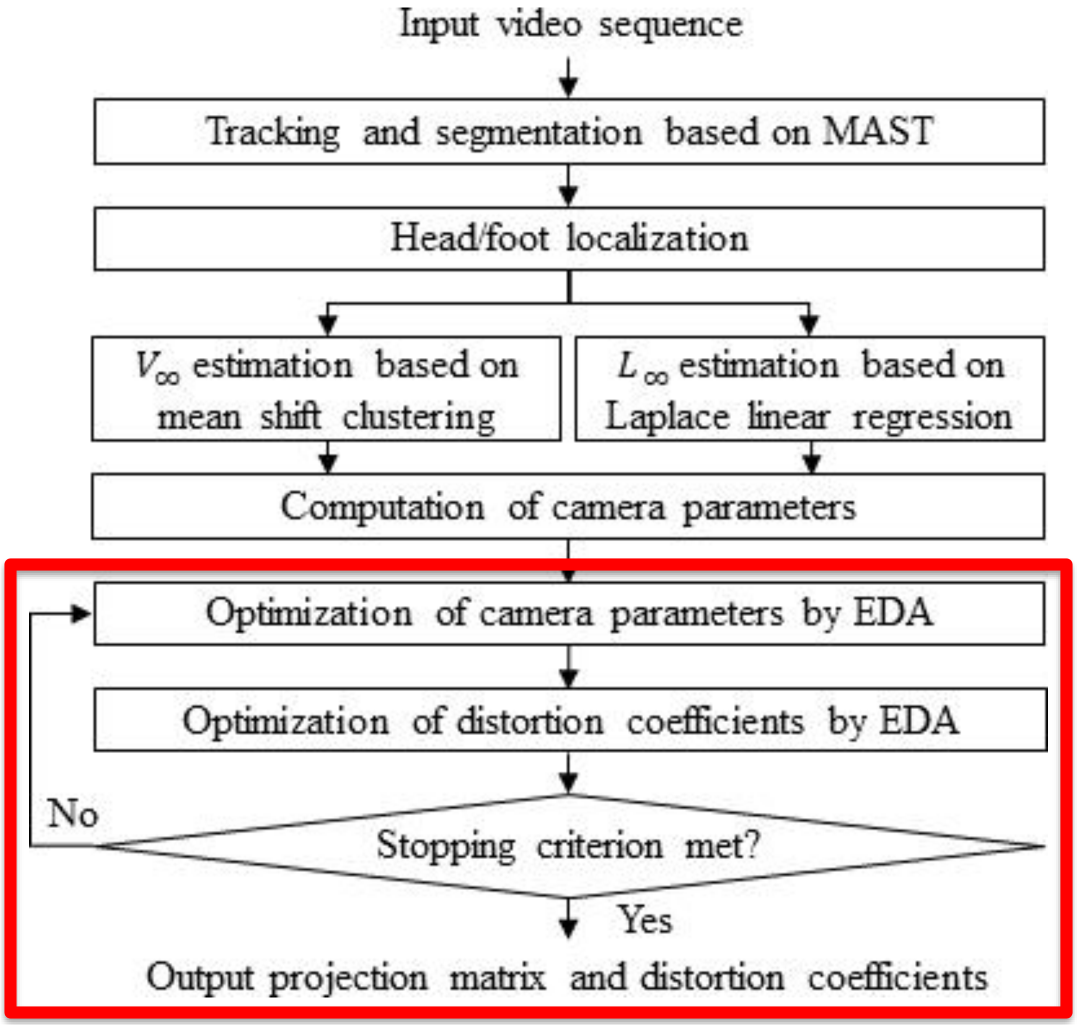
- **Sample** : Vector **k** formed by 3 radial distortion coefficients
- **PDF** : 3-variate normal density function
- **Stopping criterion**: Changing ratio between generations smaller than threshold

- **Objective function**: Relative human height variance

$$\mathbf{k}^* = \underset{\mathbf{k} \in \mathrm{Rng}_{\mathbf{k}}}{\arg\min} \, \mathrm{E}\left(\Delta H_{o,t}{}^2\right) \quad \mathrm{s.t.}, \Delta H_{o,t} = \frac{H_{o,t} - \overline{H_o}}{\overline{H_o}}$$

$H_{o,t}$: Estimated 3D height of object o at time t
$\overline{H_o}$: Average 3D height of object o along time

Camera Self-calibration



Camera Self-calibration

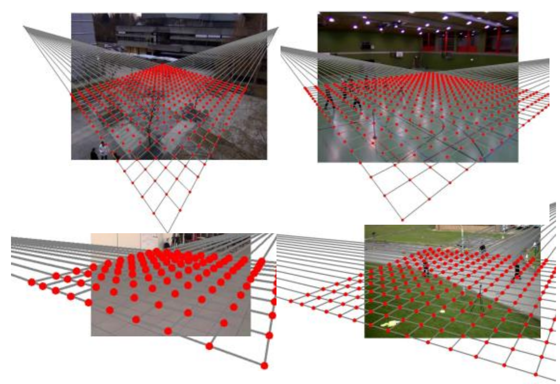
Seq. # & Method	Δf (pix.)	Δc_u (pix.)	Δc_v (pix.)	$\Delta\gamma$ (deg.)	$\Delta\beta$ (deg.)	Δt_Z (mm)
1 - ESTHER	**121.5**	*23.3*	**12.7**	*1.64*	*0.39*	**50**
1 - Tang *et al.*, ICPR'16	*124.6*	**19.2**	16.0	1.82	1.17	78
1 - Brouwers *et al.*, ECCV'16	179.0	43.9	*14.8*	**1.14**	**0.22**	*62*
1 - Liu *et al.*, BMVC'11	347.0	43.9	*14.8*	N/A	N/A	N/A
1 - Liu *et al.*, WACV'13	229.0	43.9	*14.8*	N/A	N/A	N/A
1 - Wu *et al.*, ISVC'07	251.9	43.9	*14.8*	8.68	3.94	N/A
1 - Lv *et al.*, ICPR'02	382.7	43.9	*14.8*	15.01	5.47	N/A
2 - ESTHER	**126.5**	**15.1**	*13.7*	*2.61*	1.57	**97**
2 - Tang *et al.*, ICPR'16	*126.8*	*19.0*	**11.2**	2.90	*1.18*	*115*
2 - Brouwers *et al.*, ECCV'16	265.0	41.2	18.0	**0.27**	**0.33**	790
2 - Wu *et al.*, ISVC'07	362.0	41.2	18.0	6.45	2.64	N/A
2 - Lv *et al.*, ICPR'02	520.3	41.2	18.0	8.93	3.98	N/A
3 - ESTHER	**11.5**	**4.5**	*2.9*	*2.78*	2.07	*116*
3 - Tang *et al.*, ICPR'16	*13.1*	*5.3*	**2.8**	3.49	*1.75*	**112**
3 - Brouwers *et al.*, ECCV'16	43.0	11.5	9.6	**2.91**	**0.63**	520
3 - Wu *et al.*, ISVC'07	28.6	11.5	9.6	7.30	3.04	N/A
3 - Lv *et al.*, ICPR'02	34.6	11.5	9.6	11.69	2.07	N/A
4 - ESTHER	52.2	*13.8*	6.0	*2.46*	**1.45**	**294**
4 - Tang *et al.*, ICPR'16	**51.8**	**12.0**	7.9	**1.84**	*1.75*	*327*
4 - Führ *et al.*, TCSVT'14	*52.0*	59.8	**5.4**	N/A	N/A	N/A
4 - Wu *et al.*, ISVC'07	60.5	59.8	**5.4**	2.77	1.92	N/A
4 - Lv *et al.*, ICPR'02	89.6	59.8	**5.4**	7.56	3.29	N/A

- Calibration results on VPTZ, EPFL & MOTChallenge

[Possegger *et al.*, CVWW'12]

[Fleuret *et al.*, TPAMI'08]

[Leal-Taixé *et al.*, arXiv'15]



Camera Self-calibration

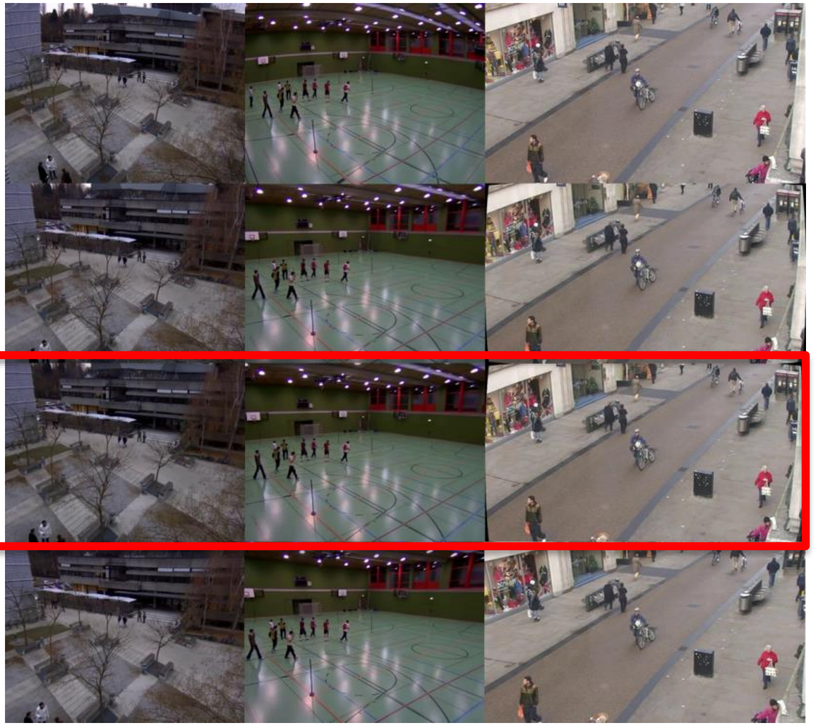
- Radial distortion correction results on VPTZ & MOTChallenge

Orig.

GT

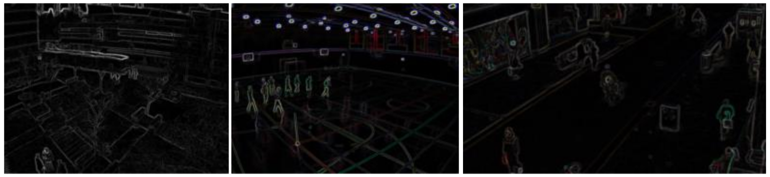
ESTHER

ESTHER (MWA)



Seq. # & Method	k_1	k_2
1 - Ground truth	-0.374	0.159
1 - ESTHER	**-0.383**	**0.176**
1 - ESTHER (MWA)	*-0.346*	*0.119*
2 - Ground truth	-0.365	0.131
2 - ESTHER	**-0.327**	**0.117**
2 - ESTHER (MWA)	*-0.479*	*0.198*
5 - Ground truth	-0.602	4.702
5 - ESTHER	**-0.595**	*4.730*
5 - ESTHER (MWA)	*-0.579*	**4.685**

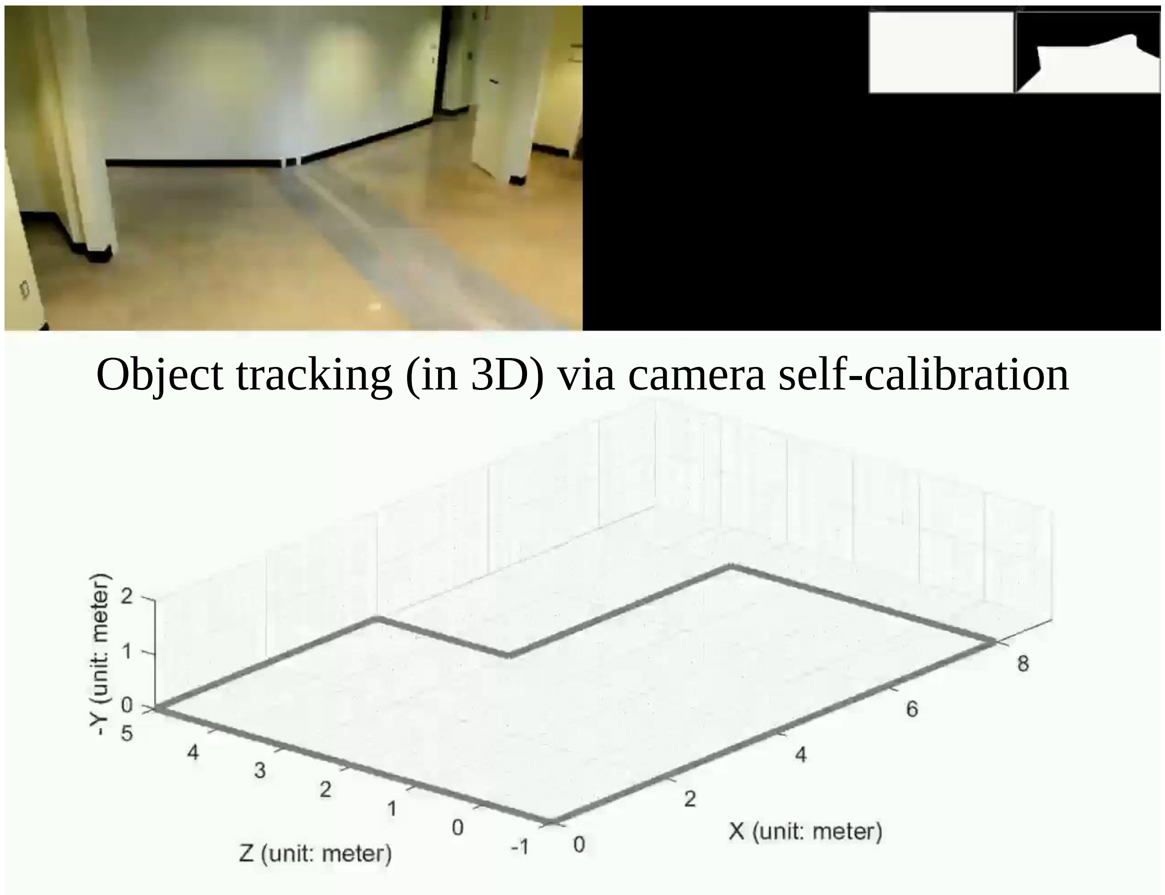
Line segments for MWA



Camera Self-calibration

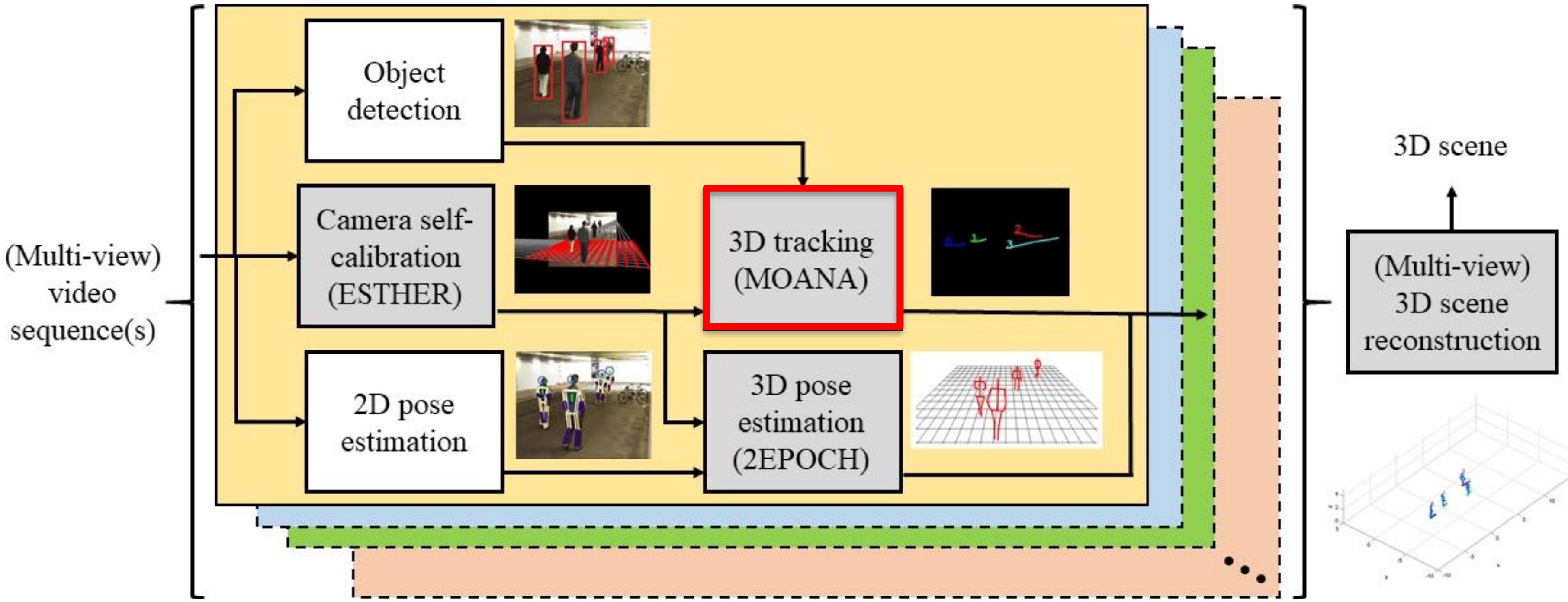
- Demonstration of tracking in 3D

Object tracking (in 2D)



Object segmentation (w/ region of interest)

Outline

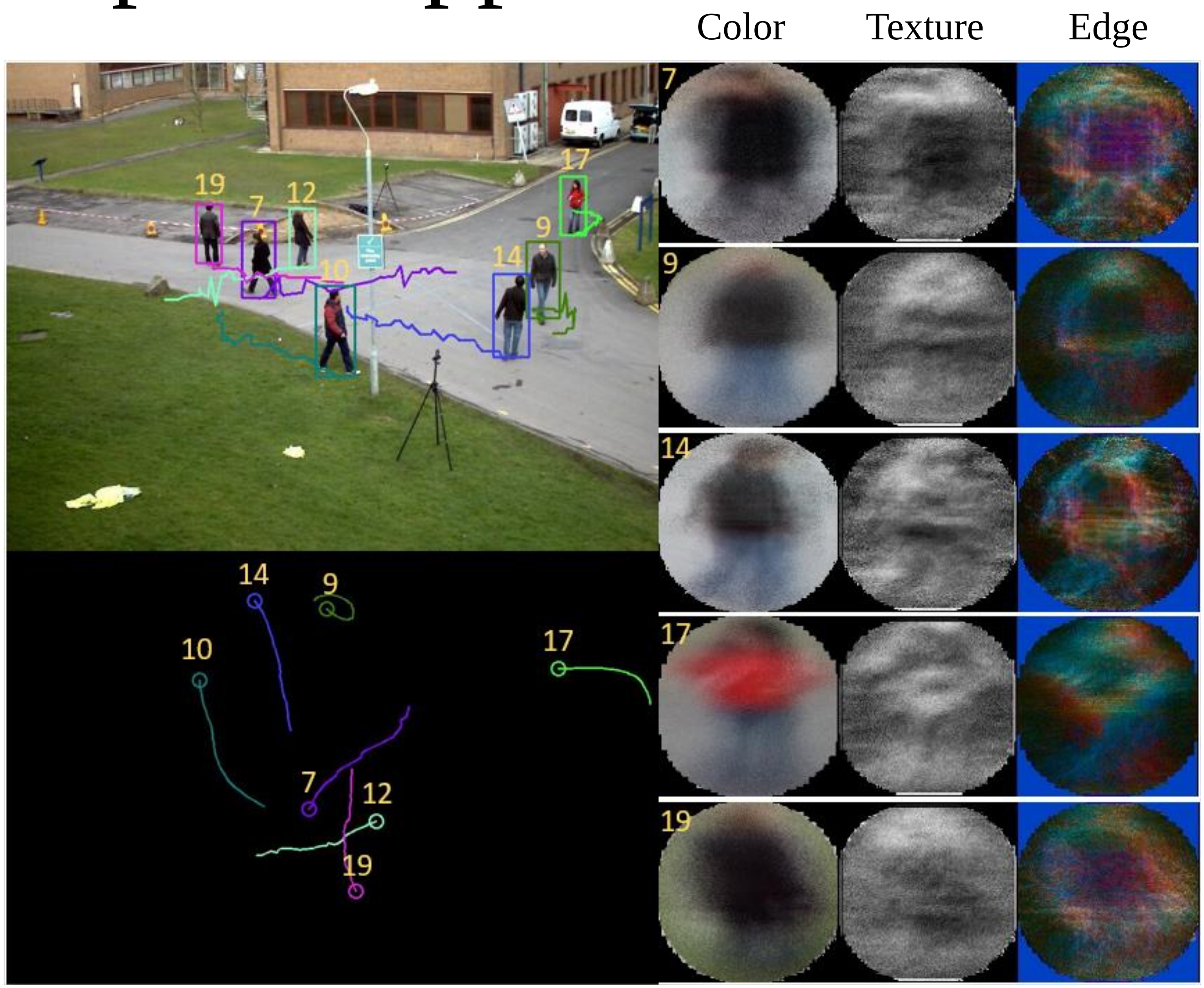


- **ESTHER**: Evolutionary Self-calibration from Tracking of Humans for Enhancing Robustness
- **MOANA**: Modeling of Object Appearance by Normalized Adaptation
- **2EPOCH**: Two-step Evolutionary Pose Optimization for Camera and Humans
- Extension to multi-view 3D scene reconstruction

Adaptive Appearance Modeling

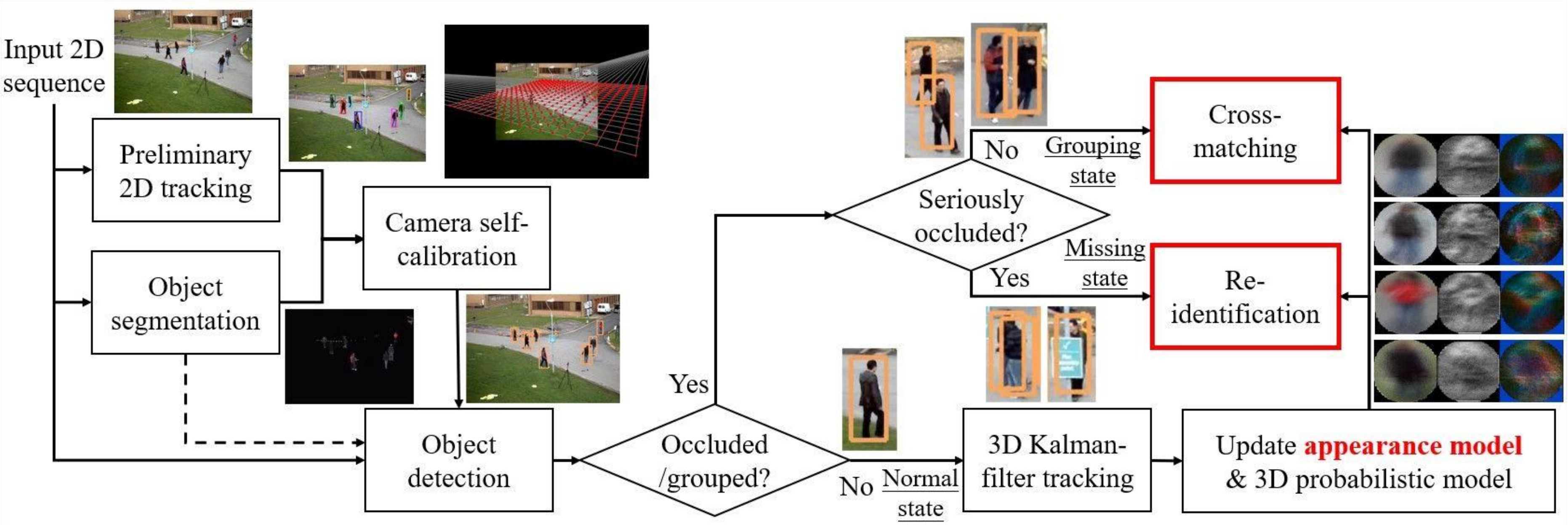
2D tracking

3D tracking (top view)



Adaptive appearance models

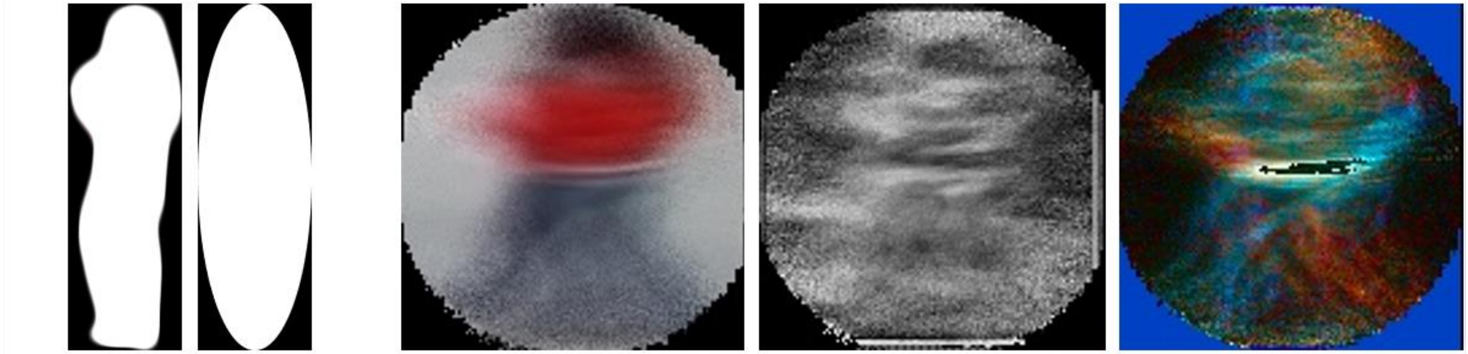
Adaptive Appearance Modeling



Adaptive Appearance Modeling

- Construction of adaptive appearance model

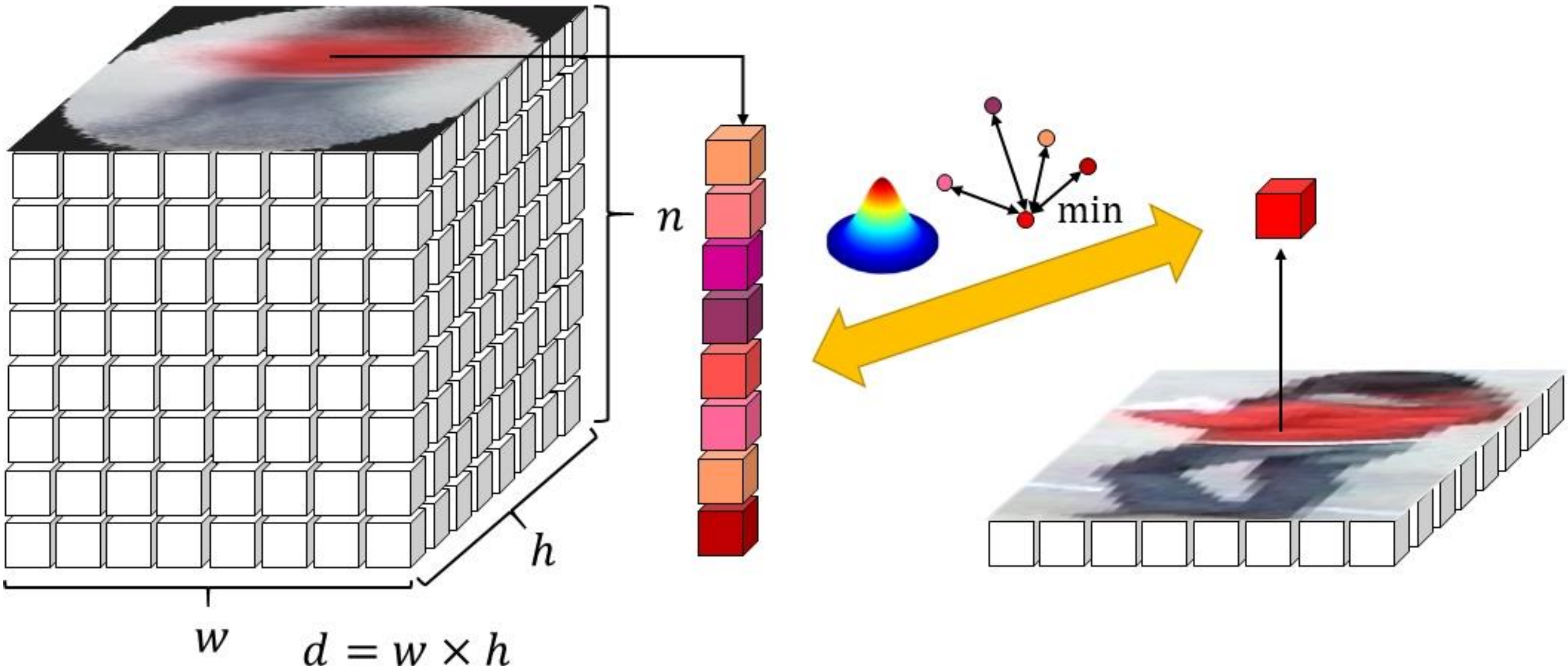
Feature maps Normalized feature maps



Segmentation masks Adaptive appearance models along time

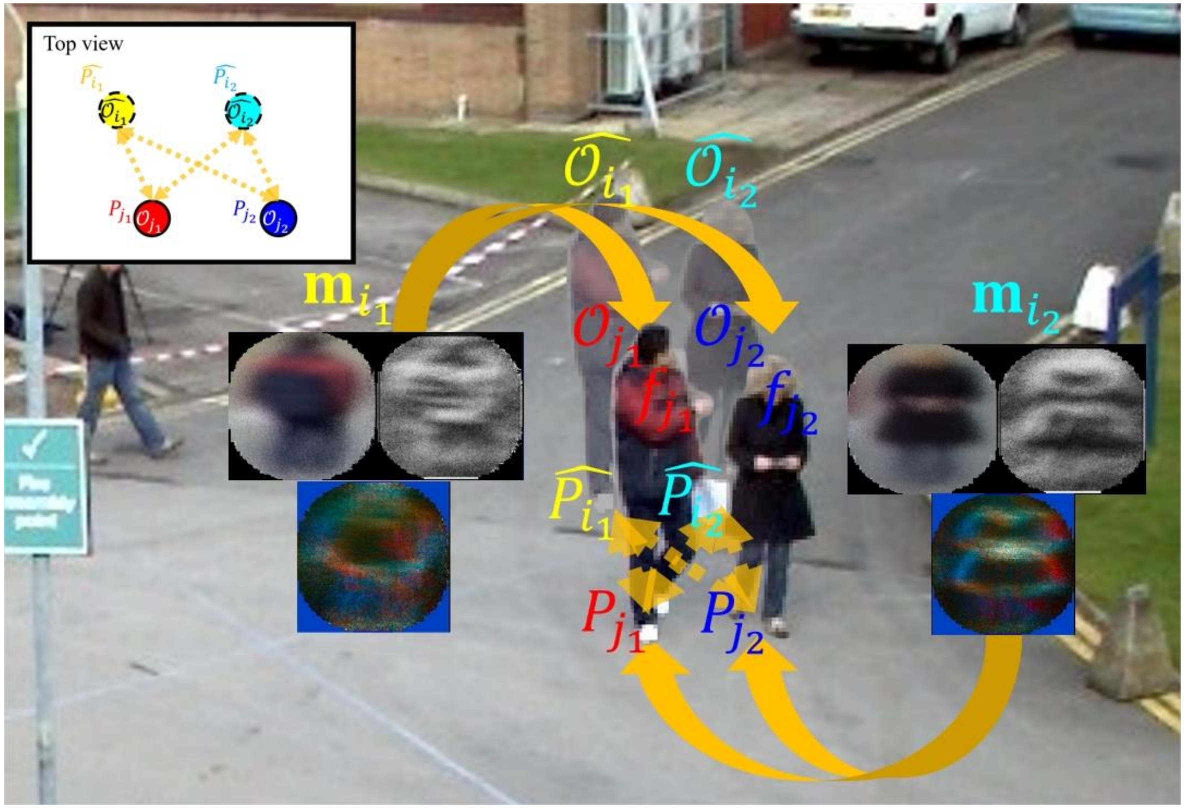
Adaptive Appearance Modeling

- Update of adaptive appearance model



Adaptive Appearance Modeling

- Cross-matching



i: Index for a target

j: Index for an observation

$\mathcal{O}_j$: Observation

$\widehat{\mathcal{O}_i}$: Prediction from target

P_j: Observed 3D location

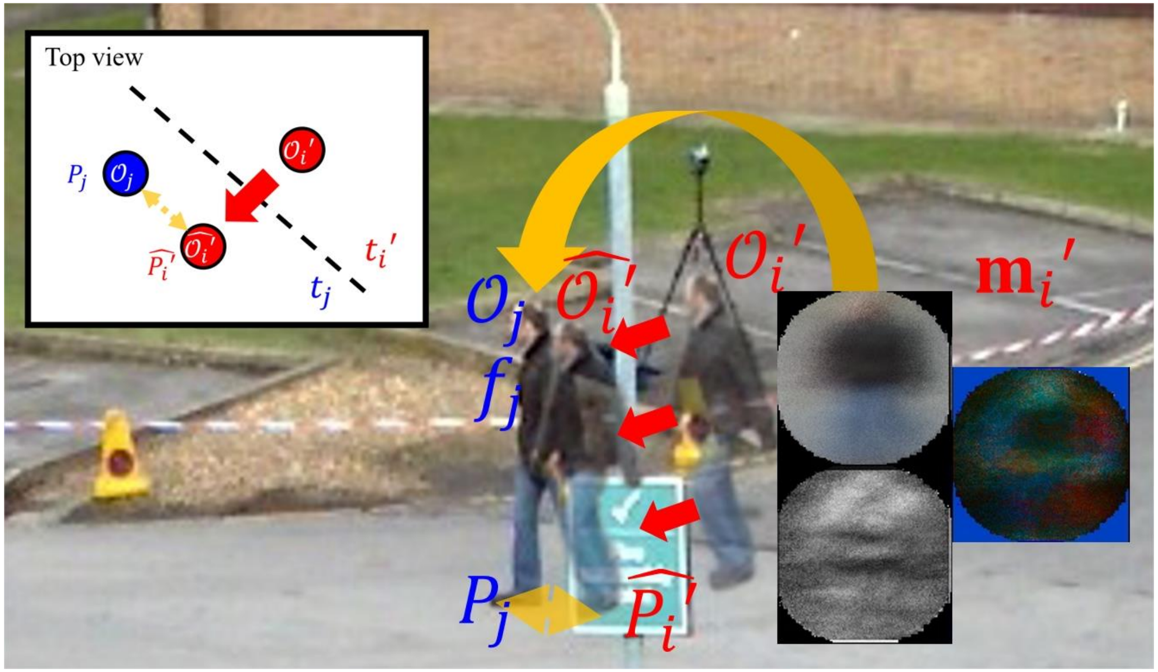
$\widehat{P_i}$: Predicted 3D location

f_j: Appearance features of an observation

$\mathbf{m}_i$: Appearance model of a target

Adaptive Appearance Modeling

- Re-identification



i: Index for a target

j: Index for an observation

t_j: Current time

t_i': Disappeared time

$\mathcal{O}_j$: Observation

$\widehat{\mathcal{O}}_i$: Prediction from target

P_j: Observed 3D location

$\widehat{P}_i$: Predicted 3D location

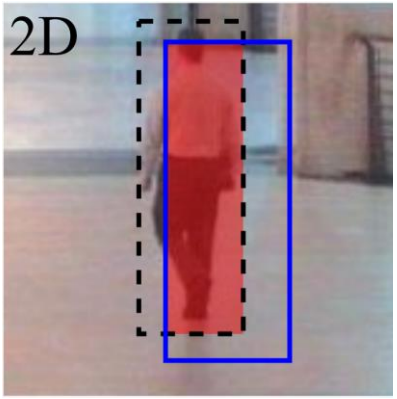
f_j: Appearance features of an observation

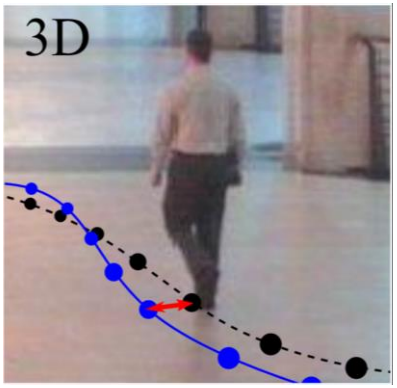
$\mathbf{m}_i$: Appearance model of a target

Adaptive Appearance Modeling

- MOTChallenge 2015 3D benchmark [Leal-Taixé *et al.*, arXiv'15]

Measure	Better	Perfect	Description
Avg Rank	↓	1	This is the rank of each tracker averaged over all present evaluation measures.
MOTA	↑	100 %	Multiple Object Tracking Accuracy. This measure combines three error sources: false positives, missed targets and identity switches.
MOTP	↑	100 %	Multiple Object Tracking Precision. The misalignment between the annotated and the predicted object locations.
MT	↑	100 %	Mostly tracked targets. The ratio of ground-truth trajectories that are covered by a track hypothesis for at least 80% of their respective life span.
ML	↓	0 %	Mostly lost targets. The ratio of ground-truth trajectories that are covered by a track hypothesis for at most 20% of their respective life span.
FP	↓	0	The total number of false positives.
FN	↓	0	The total number of false negatives (missed targets).
ID Sw.	↓	0	The total number of identity switches.
Frag	↓	0	The total number of times a trajectory is fragmented (i.e. interrupted during tracking).
Hz	↑	Inf.	Processing speed (in frames per second) on the benchmark.





Adaptive Appearance Modeling

- MOTChallenge 2015 3D benchmark [Leal-Taixé *et al.*, arXiv'15]

Tracker	Avg Rank	↑MOTA		IDF1	MT	ML	FP	FN	ID Sw.	Frag	Hz	Detector
MOANA 1.	3.2	**52.7**	±14.4	**62.4**	28.4%	22.0%	2,226	**5,551**	**167** (2.5)	586 (8.8)	19.4	Public
	Z. Tang, J. Hwang. MOANA: An online learned adaptive appearance model for robust multiple object tracking in 3D. In IEEE Access, 2019.											
DBN 2.	3.4	51.1	±7.6	0.0	**28.7%**	17.9%	2,077	5,746	380 (5.8)	418 (6.4)	0.1	Public
	T. Klinger, F. Rottensteiner, C. Heipke. Probabilistic Multi-Person Tracking using Dynamic Bayes Networks. In ISPRS Workshop on Image Sequence Analysis (ISA), 2015.											
GPDBN 3.	3.4	49.8	±6.6	0.0	25.7%	17.2%	1,813	6,300	311 (5.0)	**386** (6.2)	0.1	Public
	T. Klinger, F. Rottensteiner, C. Heipke. Probabilistic multi-person localisation and tracking in image sequences. In ISPRS Journal of Photogrammetry and Remote Sensing, 2017.											
GustavHX 4.	3.8	42.5	±0.2	45.0	25.7%	15.7%	2,735	6,623	302 (5.0)	431 (7.1)	0.0	Public
	Anonymous submission											
MCFPHD 5.	4.8	39.9	±12.3	0.0	25.7%	16.8%	3,029	6,700	363 (6.0)	529 (8.8)	17.7	Public
	N. Wojke, D. Paulus. Global data association for the Probability Hypothesis Density filter using network flows. In 2016 IEEE International Conference on Robotics and Automation, ICRA, 2016.											
MCG 6.	6.2	35.9	±7.5	31.9	8.2%	25.7%	**1,600**	8,464	692 (14.0)	1,017 (20.5)	0.1	Public
	Anonymous submission											
LPSFM 7.	5.2	35.9	±6.3	0.0	13.8%	21.6%	2,031	8,206	520 (10.2)	601 (11.8)	8.4	Public
	L. Leal-Taixé, G. Pons-Moll, B. Rosenhahn. Everybody needs somebody: modeling social and grouping behavior on a linear programming multiple people tracker. In IEEE International Conference on Computer Vision Workshops (ICCVW). 1st Workshop on Modeling, Simulation and Visual Analysis of Large Crowds, 2011.											
LP3D 8.	4.9	35.9	±11.1	0.0	20.9%	16.4%	3,588	6,593	580 (9.6)	659 (10.9)	**83.5**	Public
	MOT baseline: Linear programming on 3D image coordinates.											
SVT 9.	6.8	34.2	±15.2	0.0	11.2%	25.4%	3,057	7,454	532 (9.6)	611 (11.0)	1.9	Public
	Longyin Wen, Zhen Lei, Ming-Ching Chang, Honggang Qi, Siwei Lyu. Multi-Camera Multi-Target Tracking with Space-Time-View Hyper-graph. IJCV, 2016.											
AMIR3D 10.	7.1	25.0	±10.8	0.0	3.0%	27.6%	2,038	9,084	1,462 (31.9)	1,647 (35.9)	1.2	Public
	A. Sadeghian, A. Alahi, S. Savarese. Tracking The Untrackable: Learning To Track Multiple Cues with Long-Term Dependencies. In ICCV, 2017.											
KalmanSFM 11.	6.3	25.0	±8.5	0.0	6.7%	**14.6%**	3,161	7,599	1,838 (33.6)	1,686 (30.8)	30.6	Public
	S. Pellegrini, A. Ess, K. Schindler, L. Gool. You'll never walk alone: Modeling social behavior for multi-target tracking.. In ICCV, 2009.											

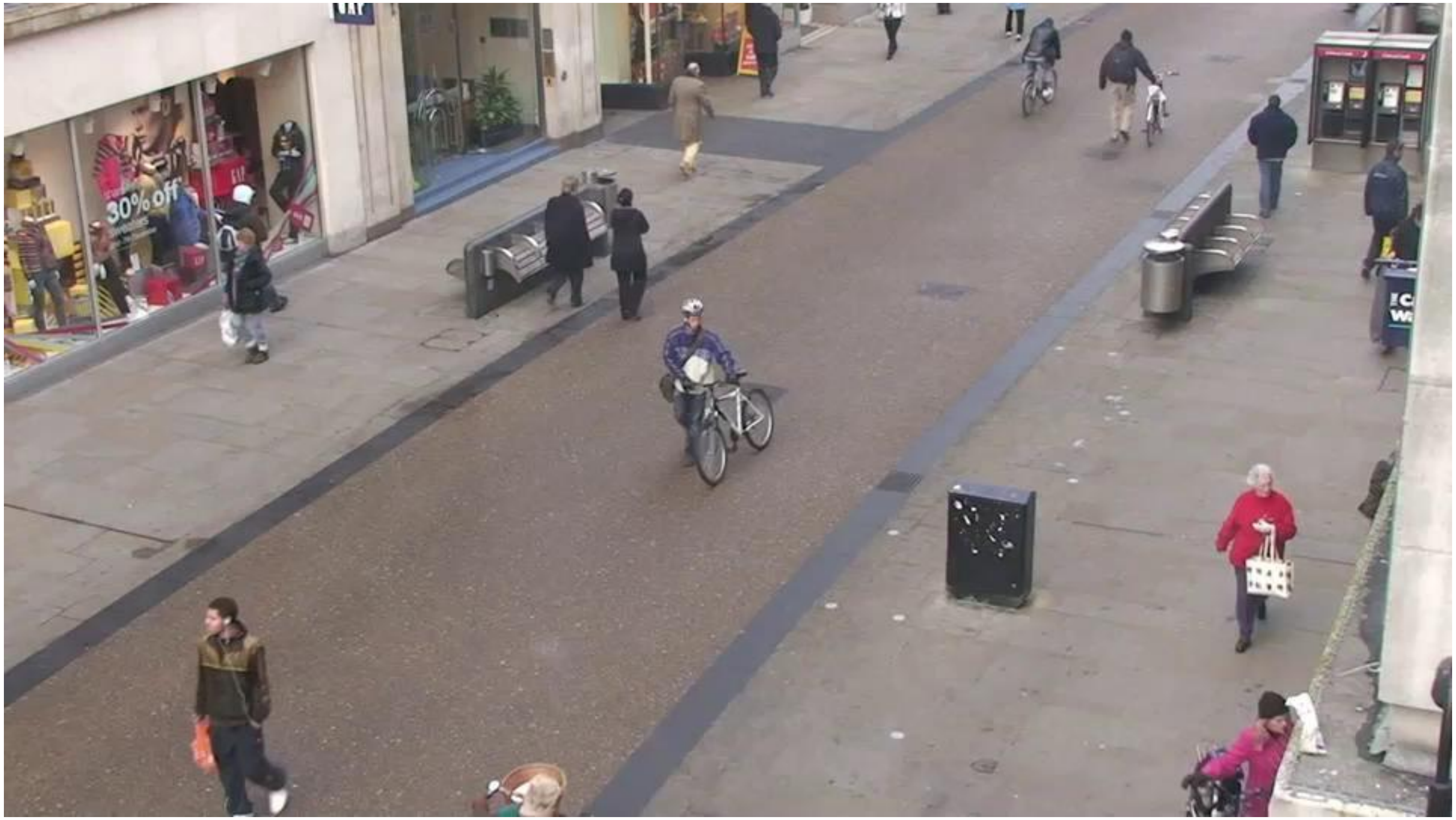
Adaptive Appearance Modeling

- Demo on MOTChallenge 2015 3D benchmark

Public detections from Deformable Part Model [Felzenszwalb *et al.*, CVPR'08]

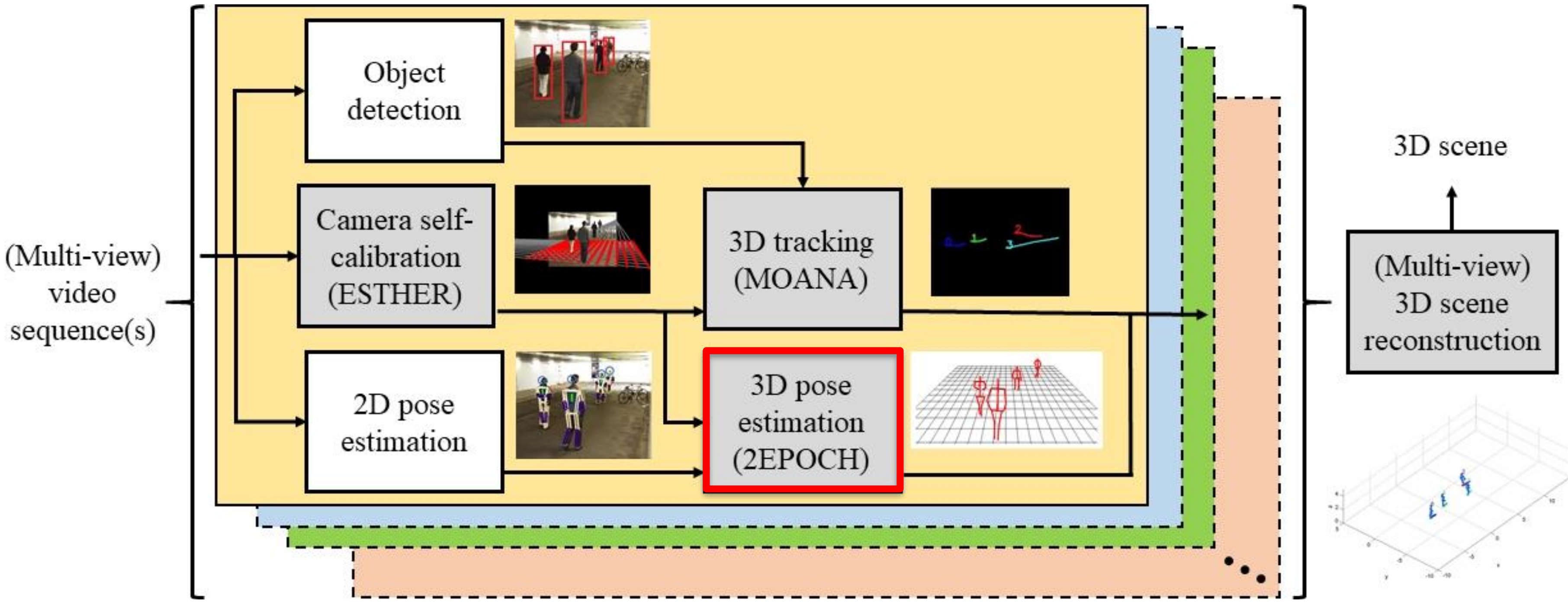
Adaptive Appearance Modeling

- Demo on MOTChallenge 2015 3D benchmark



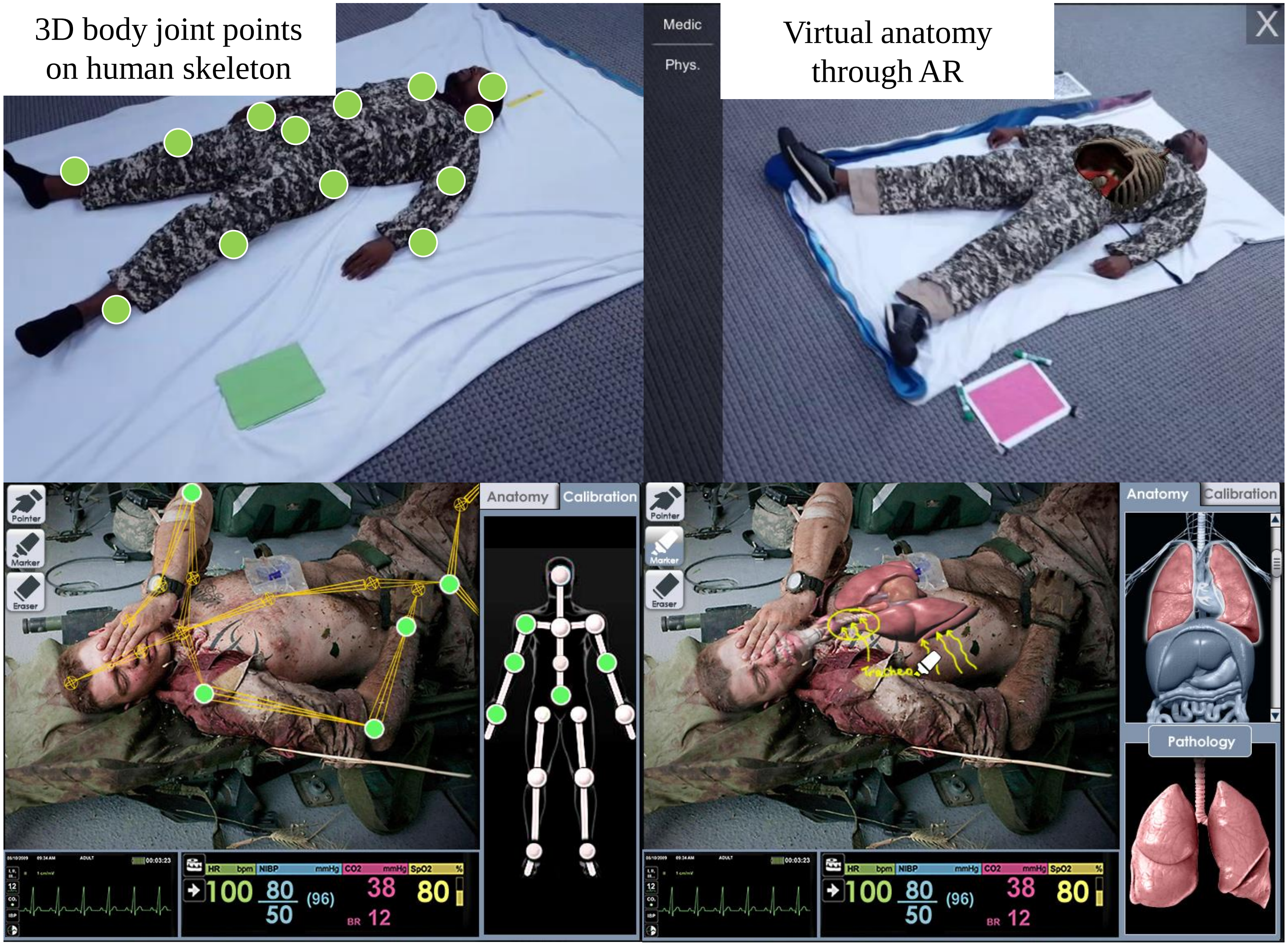
Public detections from Deformable Part Model [Felzenszwalb *et al.*, CVPR'08]

Outline

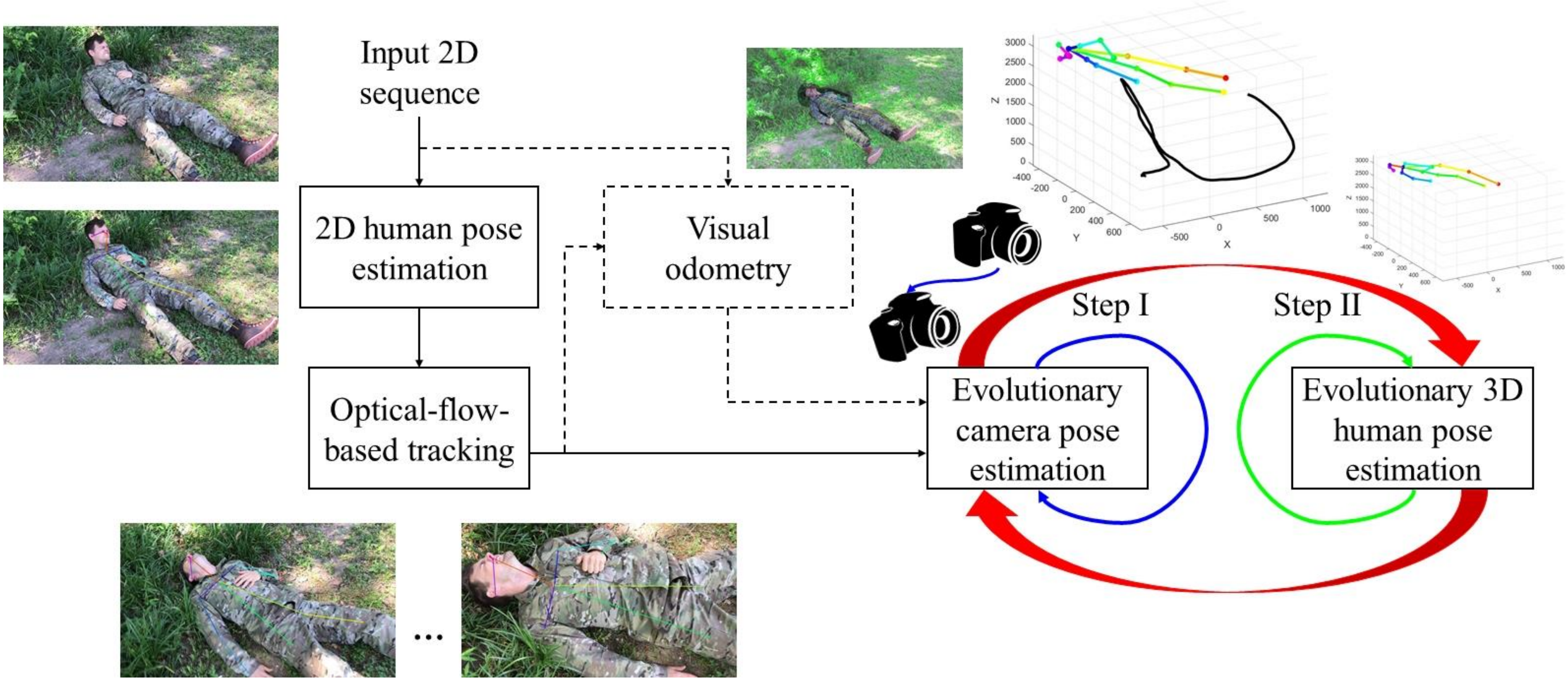


- **ESTHER**: Evolutionary Self-calibration from Tracking of Humans for Enhancing Robustness
- **MOANA**: Modeling of Object Appearance by Normalized Adaptation
- **2EPOCH**: Two-step Evolutionary Pose Optimization for Camera and Humans
- Extension to multi-view 3D scene reconstruction

3D Pose Estimation



3D Pose Estimation



3D Pose Estimation

- 2D human pose estimation [Cao *et al.*, CVPR'17]

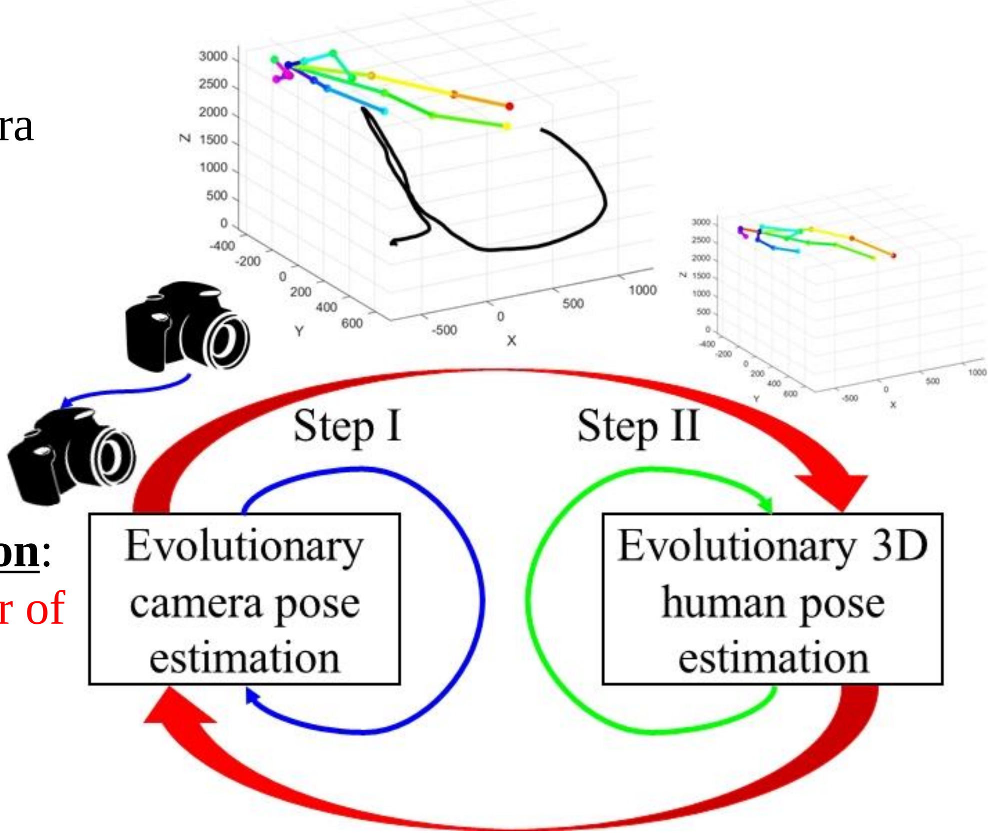
3D Pose Estimation

- Visual odometry [Nistér *et al.*, CVPR'04]

3D Pose Estimation

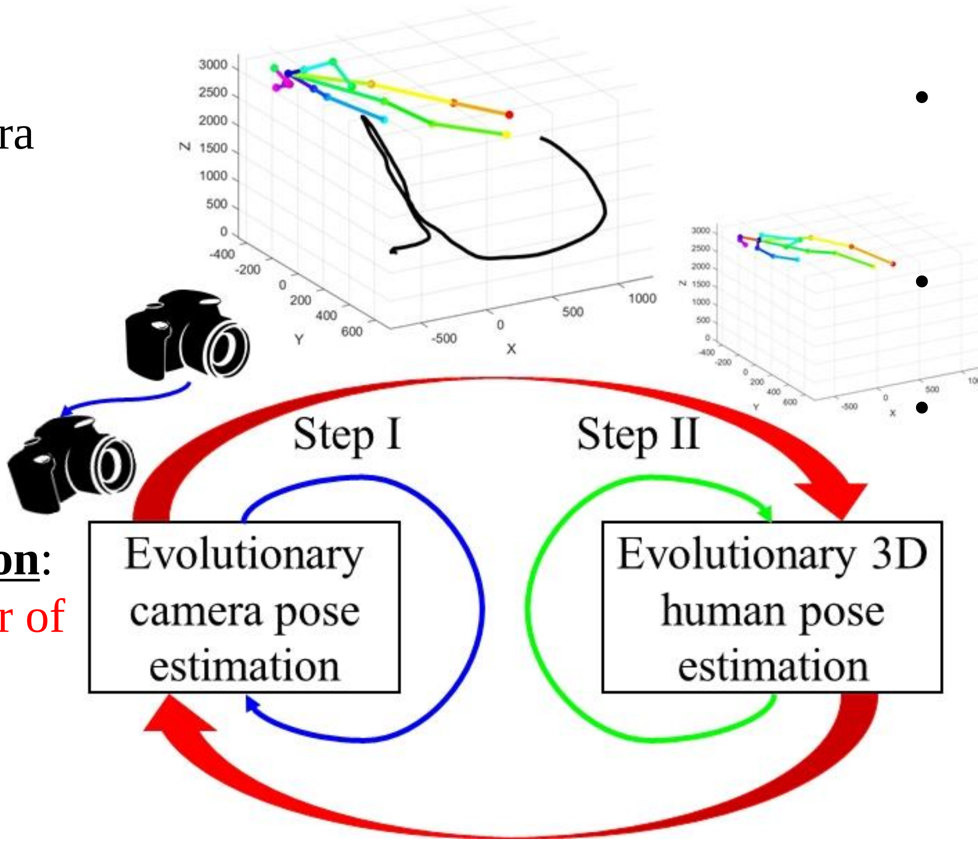
- 3D pose estimation by two-step EDA

- **Sample**：6 camera parameters for rotation and translation
- **PDF**：6-variate normal density function
- **Objective function**: Reprojection error of 18 joint points



3D Pose Estimation

- 3D pose estimation by two-step EDA

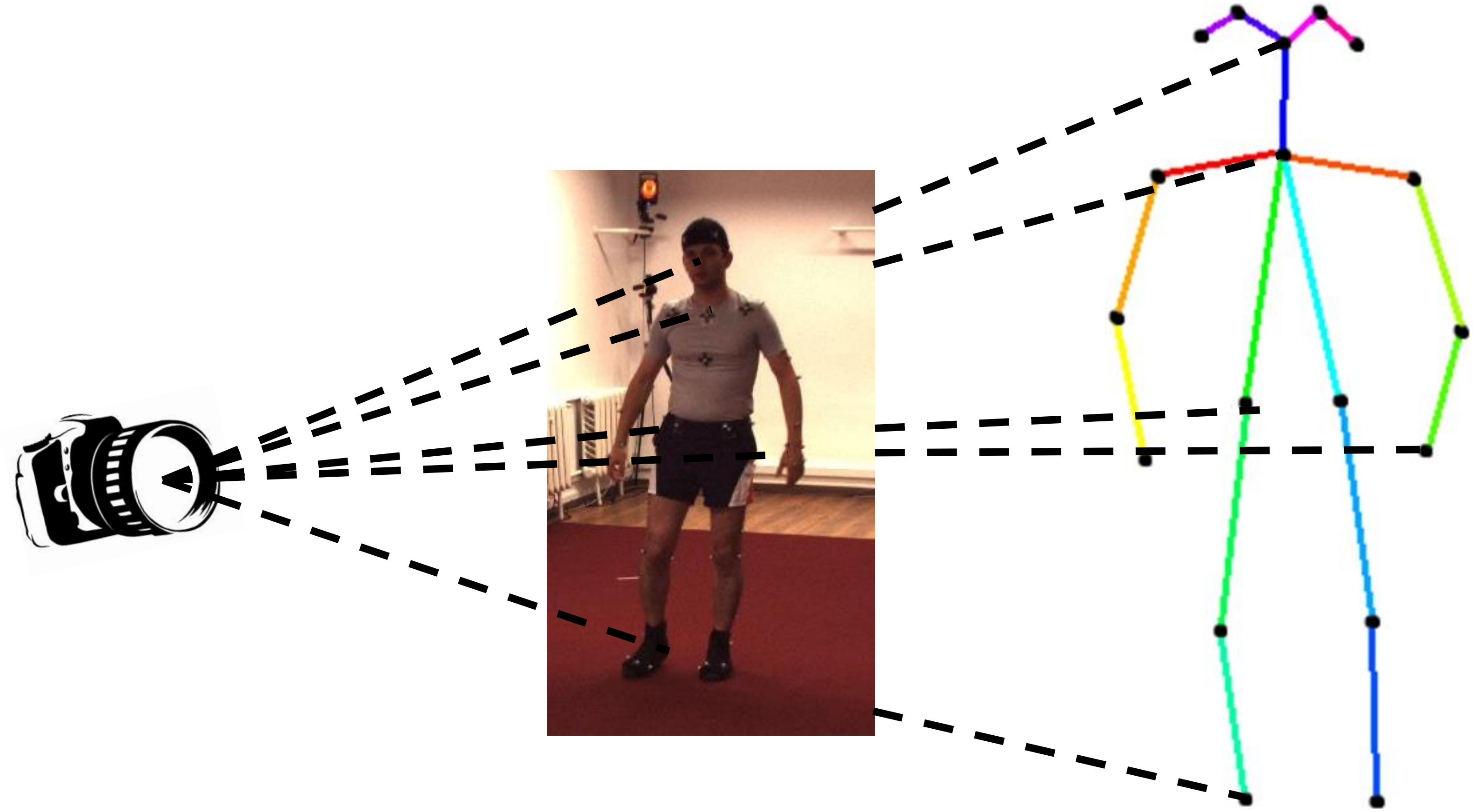


- **Sample** : 6 camera parameters for rotation and translation
- **PDF** : 6-variate normal density function
- **Objective function**: Reprojection error of 18 joint points

- **Sample** : Root-relative depths of 18 joint points
- **PDF** : 18-variate normal density function
- **Objective function**: Weighted sum of
  1. Spatial constancy loss
  2. Temporal constancy loss
  3. Body “flatness” loss
  4. Joint angle loss

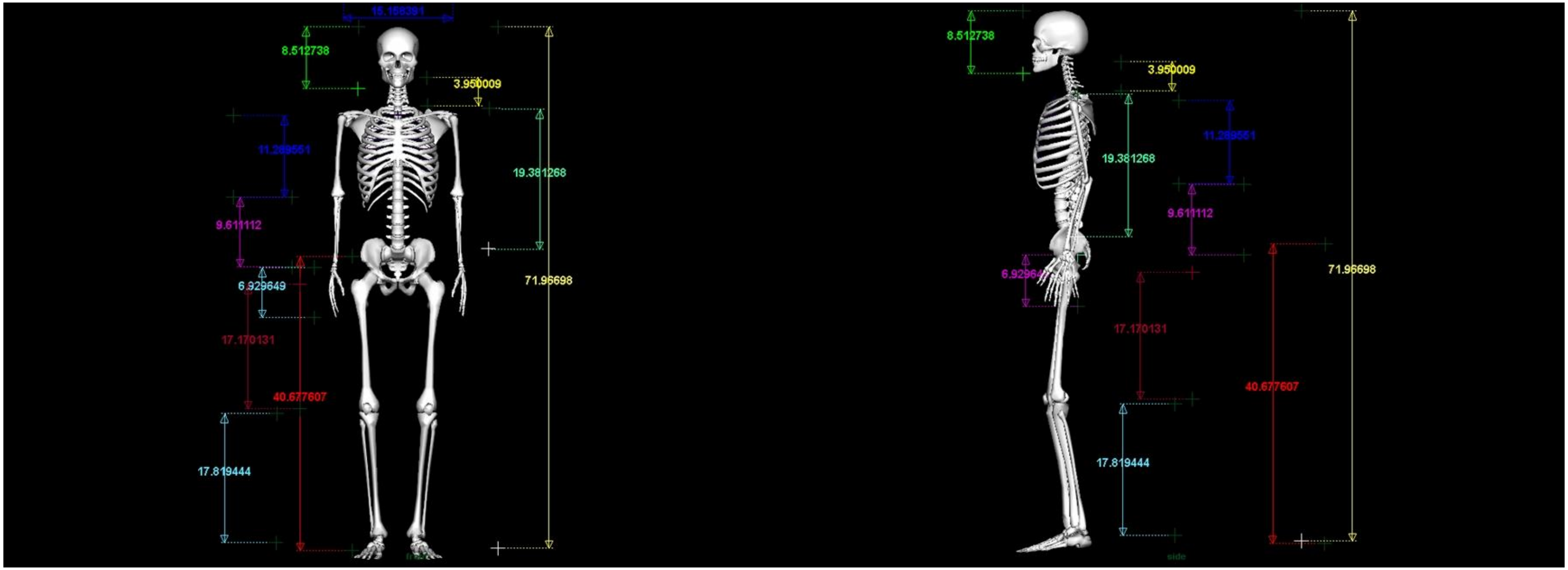
3D Pose Estimation

- Root-relative depths for human pose optimization



3D Pose Estimation

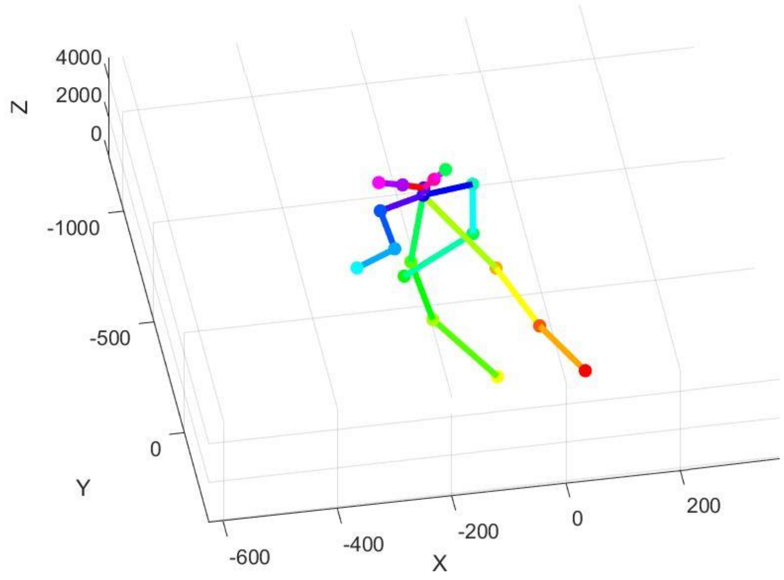
- Spatial constancy for human pose optimization



[ArchieMD Inc.]

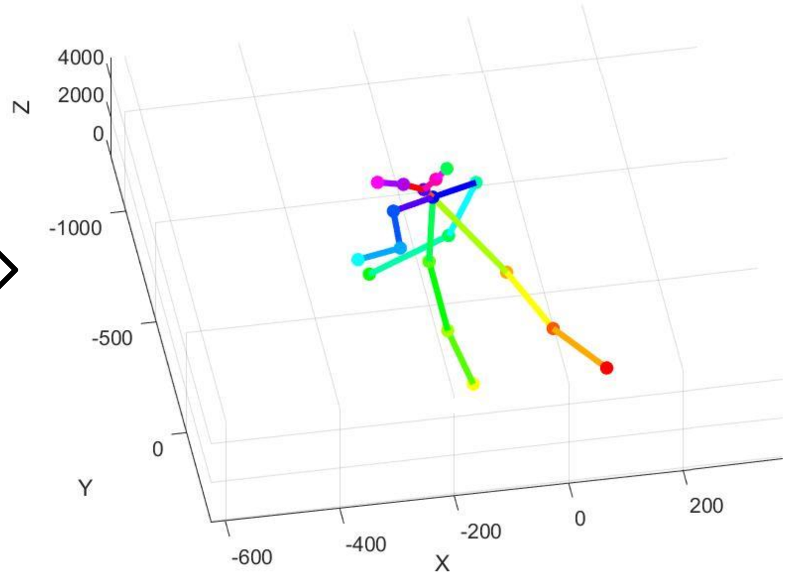
3D Pose Estimation

- Temporal constancy for human pose optimization



$t-\Delta$

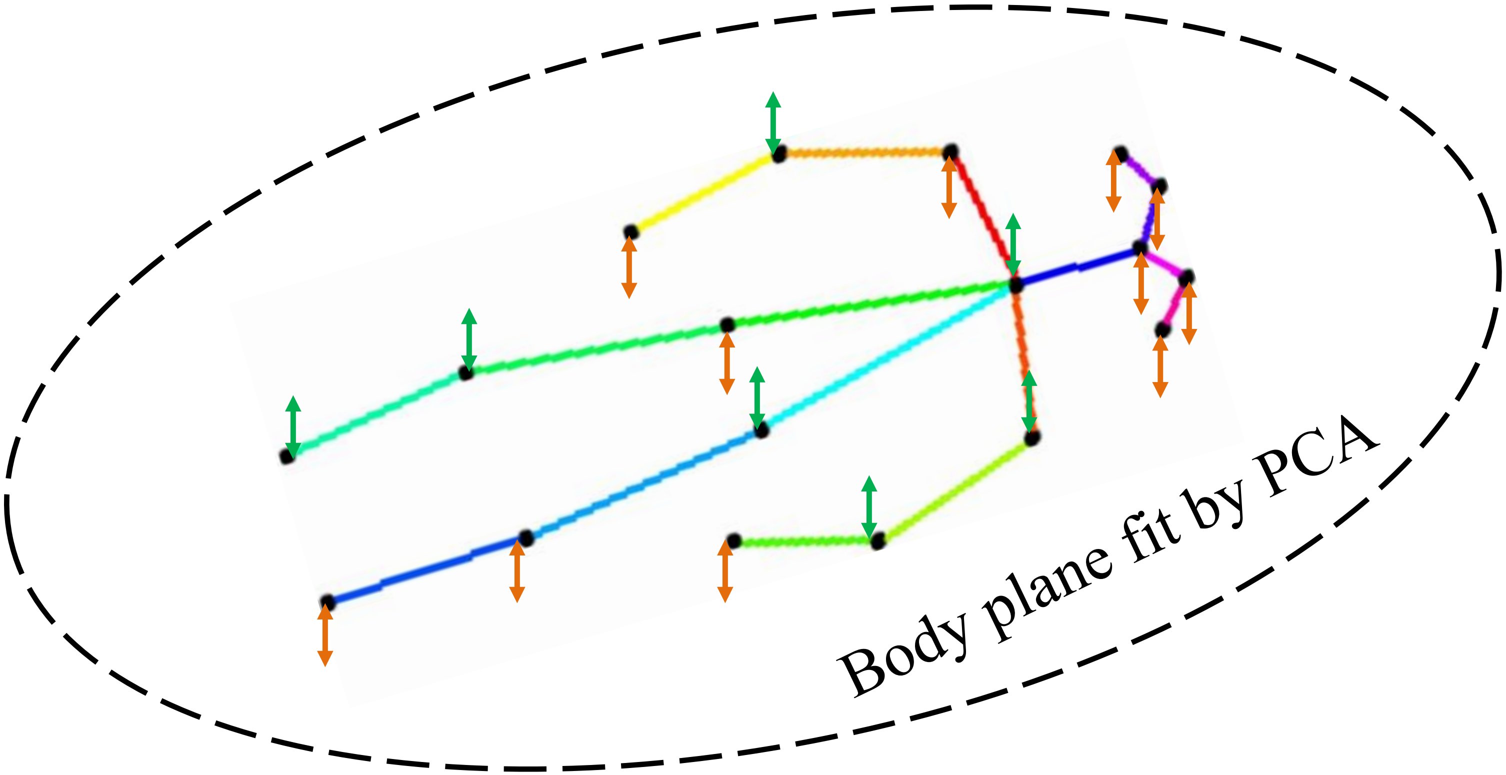
$\Leftrightarrow$



t

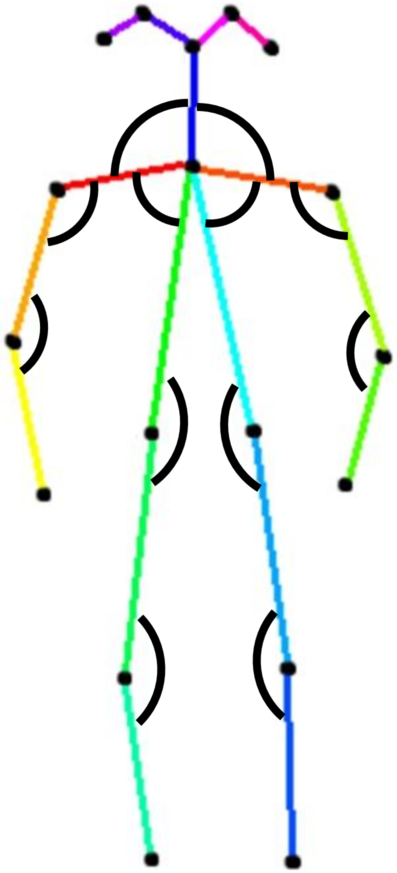
3D Pose Estimation

- Body flatness for human pose optimization



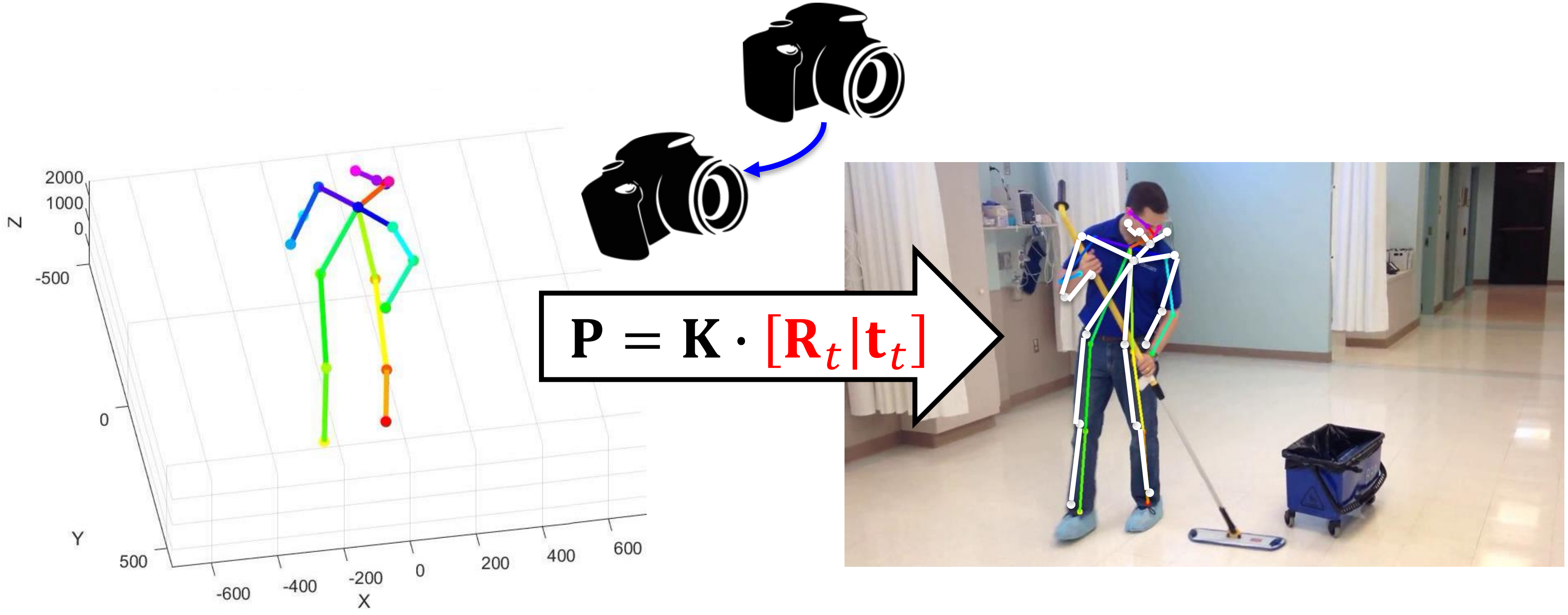
3D Pose Estimation

- Joint angle constraints for human pose optimization



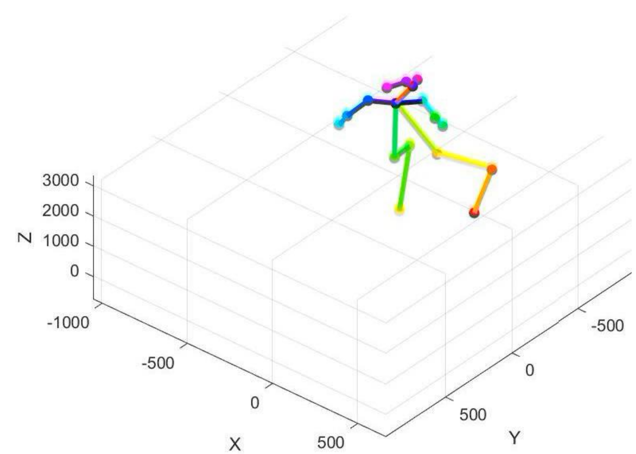
3D Pose Estimation

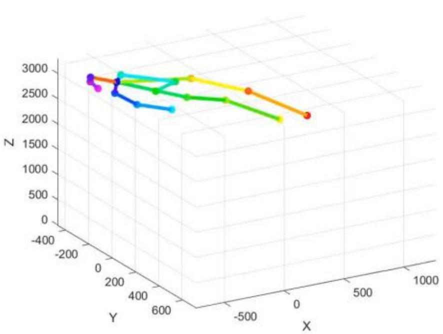
- Reprojection error for camera pose optimization

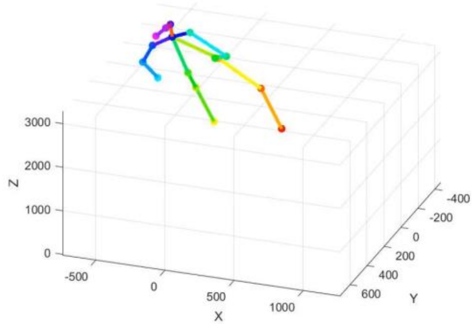


3D Pose Estimation

- Demo on people with small movement [ArchieMD Inc.]

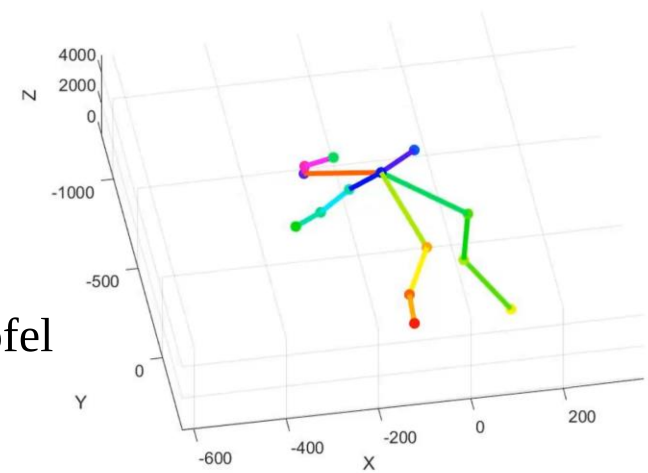


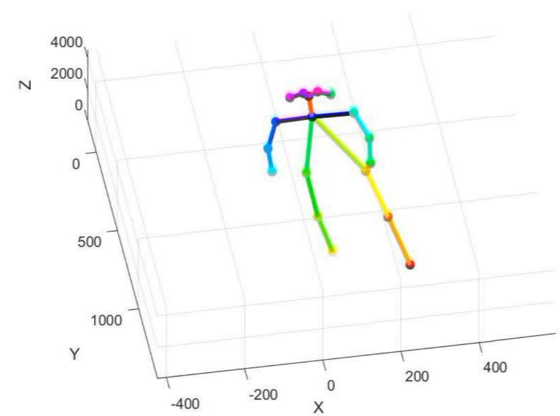




3D Pose Estimation

- Demo on people with large movement

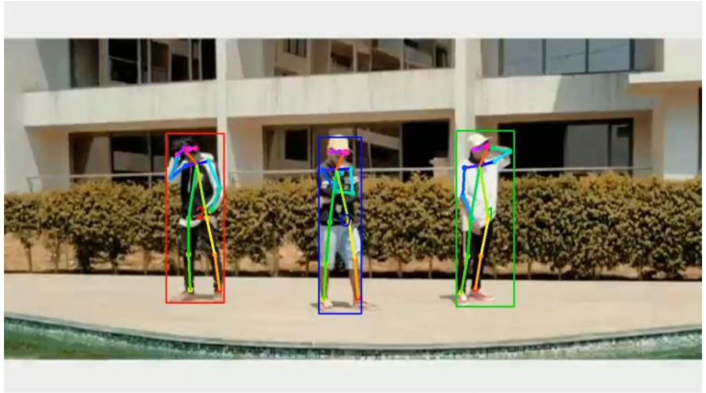


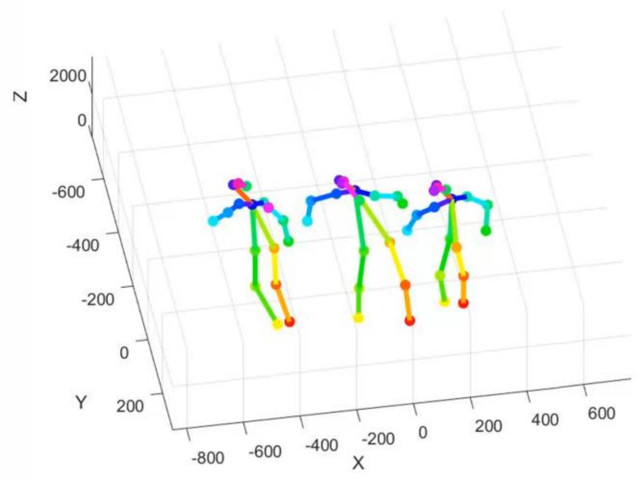


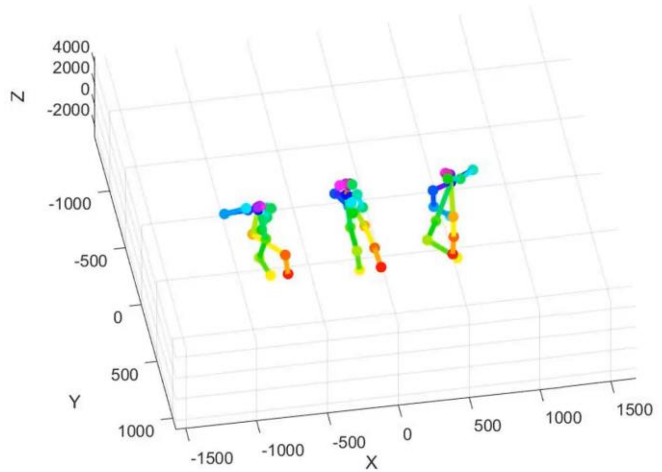
[Weinzaepfel *et al.*, arXiv'16]

3D Pose Estimation

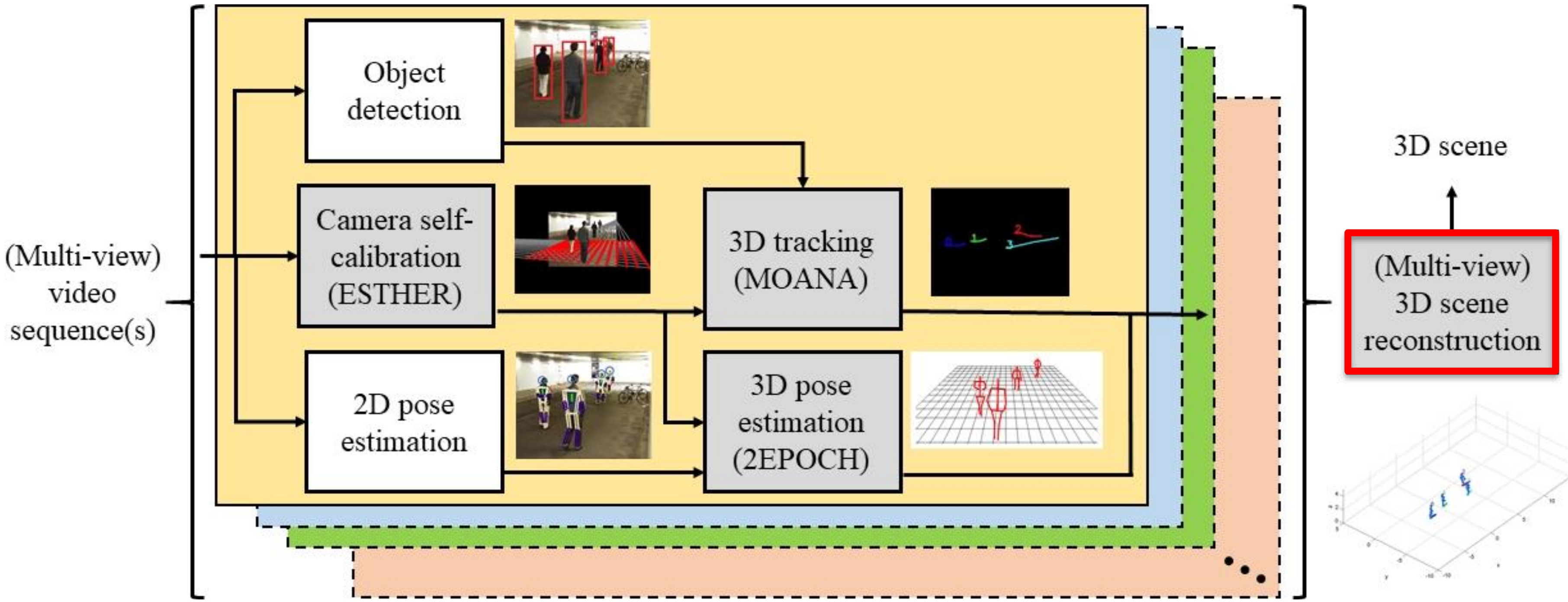
- Demo of multi-object 3D pose estimation [YouTube]







Outline

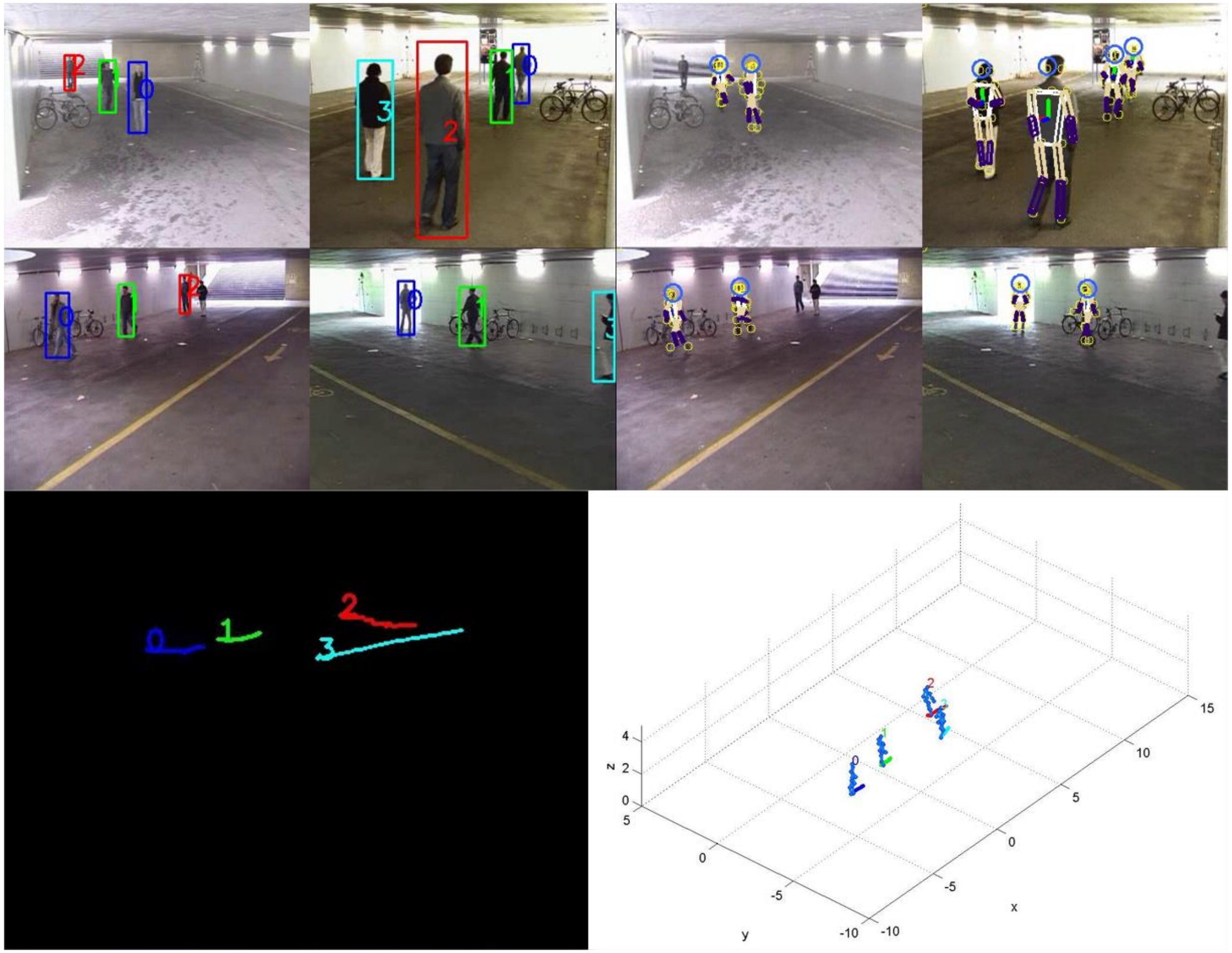


- **ESTHER**: Evolutionary Self-calibration from Tracking of Humans for Enhancing Robustness
- **MOANA**: Modeling of Object Appearance by Normalized Adaptation
- **2EPOCH**: Two-step Evolutionary Pose Optimization for Camera and Humans
- Extension to multi-view 3D scene reconstruction

3D Scene Reconstruction

Multi-view
2D tracking

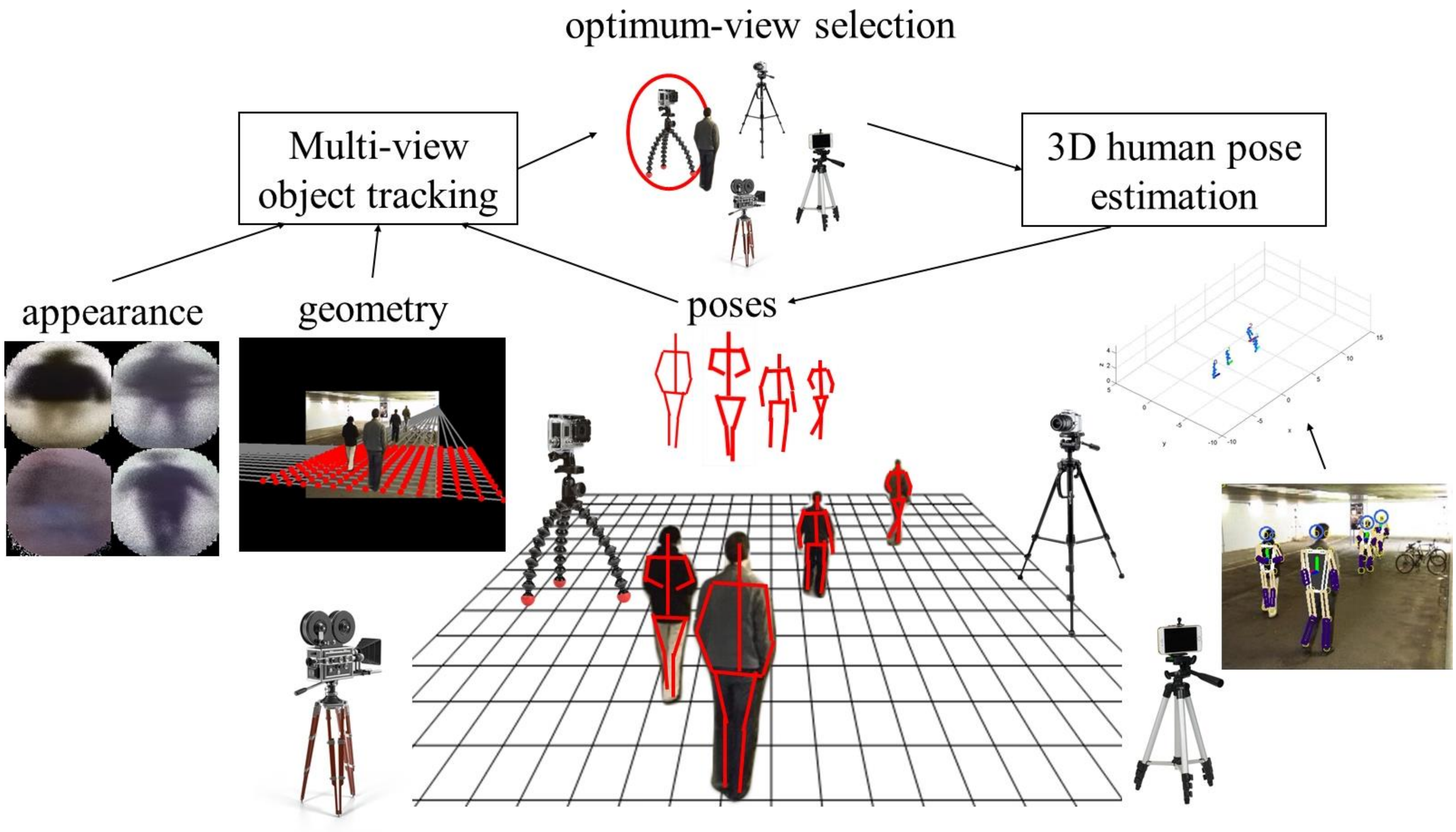
2D pose
estimation



3D
tracking
(top view)

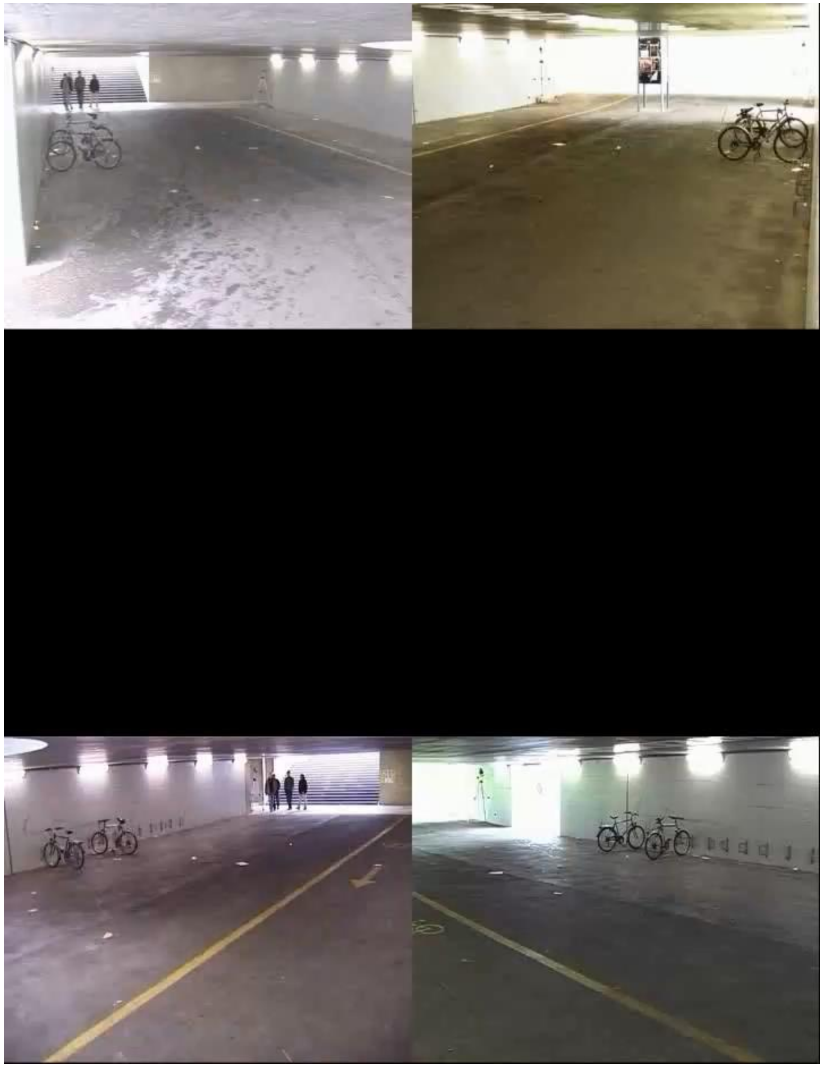
3D scene
reconstruction

3D Scene Reconstruction

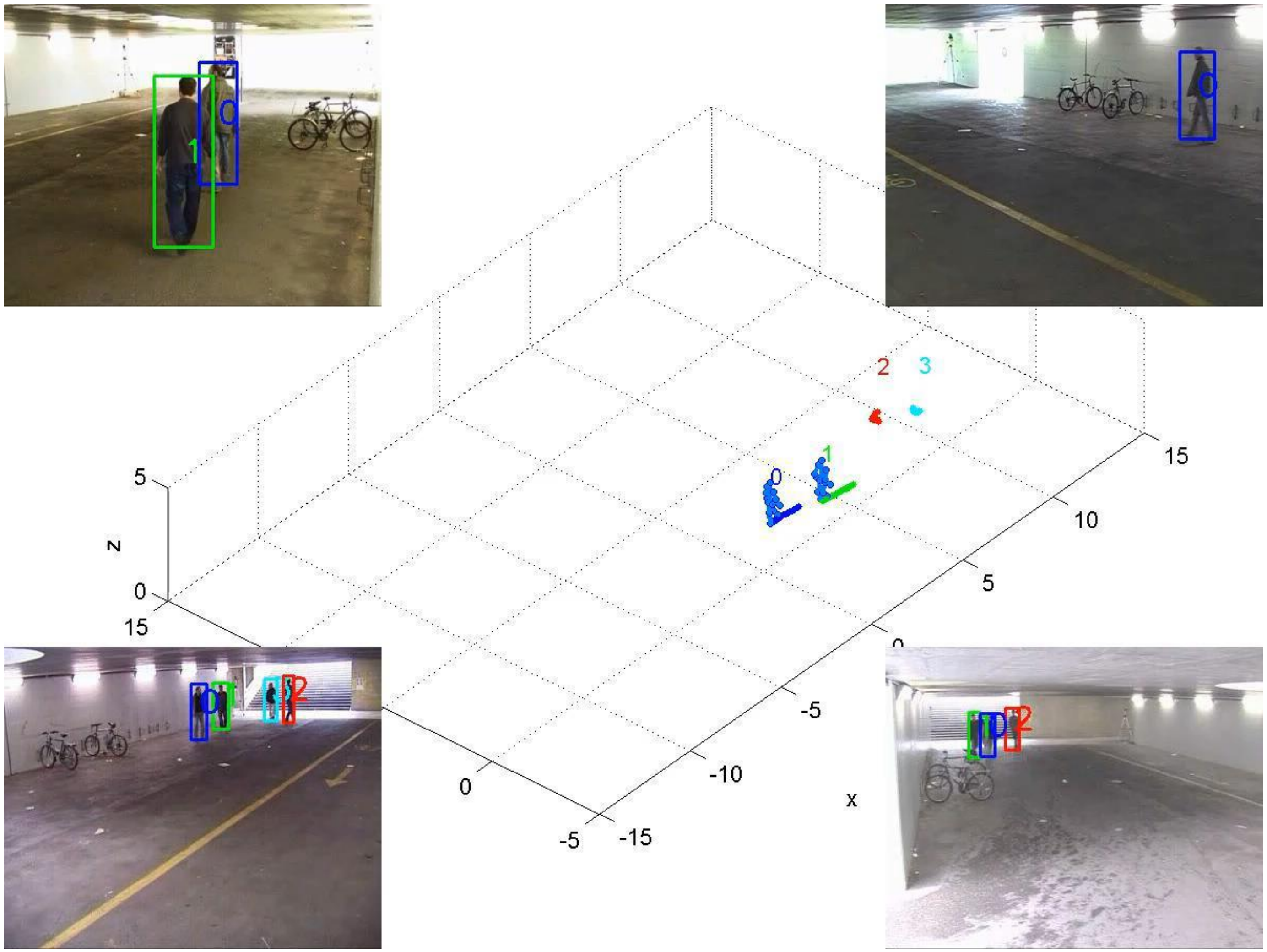


3D Scene Reconstruction

- 2D tracking in each camera view
- Multi-view 3D tracking based on data association with visual and semantic attributes



3D Scene Reconstruction



3D Scene Reconstruction

- Cross-view tracking results on EPFL [Fleuret *et al.*, TPAMI'08]

Method	MODA(%)	MODP(%)	MOTA(%)	MOTP(%)
Ours	**61.04**	**73.13**	**60.26**	**72.26**
Xu *et al.*, CVPR'16	*43.75*	*67.11*	*43.75*	*67.11*
Berclaz *et al.*, TPAMI'11	40.46	58.88	40.46	57.24
Fleuret *et al.*, TPAMI'08	32.57	62.50	32.57	60.86

MODA (Multiple Object Detection Accuracy): The accuracy considering two error sources: false positives and false negatives/missed targets

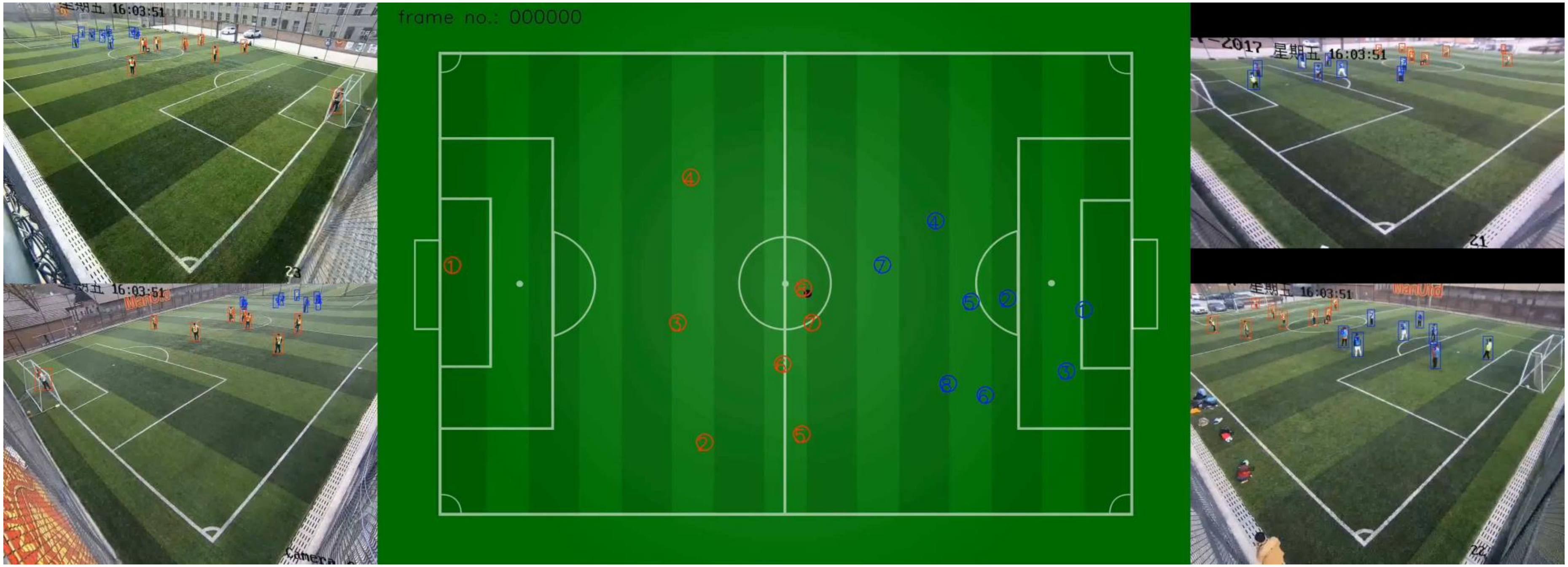
MODP (Multiple Object Detection Precision): The precision of alignment between the annotated and the predicted bounding boxes

MOTA (Multiple Object Tracking Accuracy): The accuracy considering three error sources: false positives, false negatives/missed targets and identity switches

MOTP (Multiple Object Tracking Precision): The precision of alignment between the annotated and the predicted bounding boxes

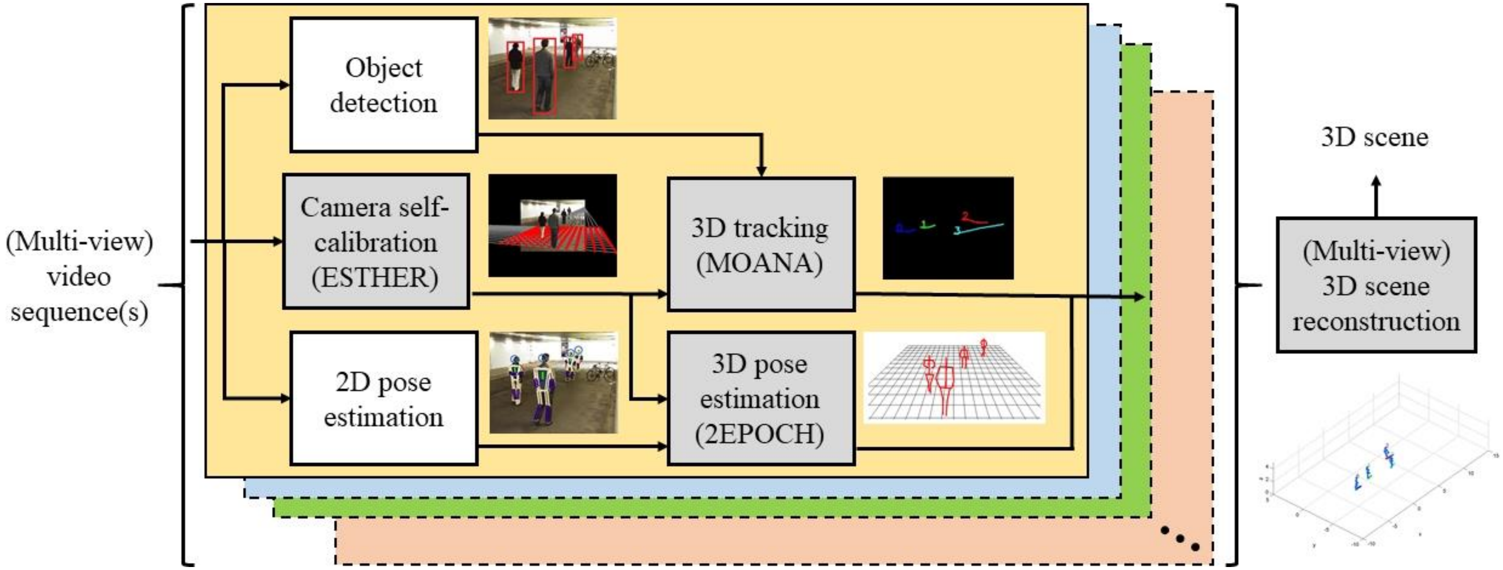
3D Scene Reconstruction

- Demo for soccer analytics



Conclusion

- 3D scene reconstruction
  - Camera self-calibration from walking humans
  - Adaptive appearance modeling for 3D tracking
  - Two-step evolutionary 3D pose estimation
  - Multi-view scene reconstruction



Publications

- Journals
  - **Z. Tang** and J.-N. Hwang, "MOANA: An online learned adaptive appearance model for robust multiple object tracking in 3D," *IEEE Access*, vol. 7, no. 1, pp. 31934-31945, 2019.
  - **Z. Tang**, Y.-S. Lin, K.-H. Lee, J.-N. Hwang and J.-H. Chuang, "ESTHER: Joint camera self-calibration and automatic radial distortion correction from tracking of walking humans," *IEEE Access*, vol. 7, no. 1, pp. 10754-10766, 2019.
  - Y.-G. Lee, **Z. Tang** and J.-N. Hwang, "Online-learning-based human tracking across non-overlapping cameras," *IEEE TCSVT*, vol. 28, no. 10, pp. 2870-2883, 2018.

Publications

- Conferences
  - **Z. Tang**, M. Naphade, M.-Y. Liu, X. Yang, S. Birchfield, S. Wang, R. Kumar, D. Anastasiu and J.-N. Hwang, "CityFlow: A city-scale benchmark for multi-target multi-camera vehicle tracking and re-identification," in *Proc. CVPR*, 2019.
  - **Z. Tang**, R. Gu and J.-N. Hwang, "Joint multi-view people tracking and pose estimation for 3D scene reconstruction," in *Proc. ICME*, 2018.
  - **Z. Tang**, G. Wang, H. Xiao, A. Zheng and J.-N. Hwang, "Single-camera and inter-camera vehicle tracking and 3D speed estimation based on fusion of visual and semantic features," in *Proc. CVPR Workshops*, pp. 108-115, 2018.
  - **Z. Tang**, G. Wang, T. Liu, Y.-G. Lee, A. Jahn, X. Liu, X. He and J.-N. Hwang, "Multiple-kernel based vehicle tracking using 3D deformable model and camera self-calibration," *arXiv:1708.06831*, 2017.

Publications

- Conferences
  - **Z. Tang**, Y.-S. Lin, K.-H. Lee, J.-N. Hwang, J.-H. Chuang and Z. Fang, "Camera self-calibration from tracking of moving persons," in *Proc. ICPR*, pp. 260-265, 2016.
  - **Z. Tang**, J.-N. Hwang, Y.-S. Lin and J.-H. Chuang, "Multiple-kernel adaptive segmentation and tracking (MAST) for robust object tracking," in *Proc. ICASSP*, pp. 1115-1119, 2016.
  - Y.-G. Lee, **Z. Tang**, J.-N. Hwang and Z. Fang, "Inter-camera tracking based on fully unsupervised online learning," in *Proc. ICIP*, pp. 2607-2611, 2017.
  - T. Liu, Y. Liu, **Z. Tang** and J.-N. Hwang, "Adaptive ground plane estimation for moving camera-based 3D object tracking," in *Proc. MMSP*, 2017.
  - N. Wang, H. Du, Y. Liu, **Z. Tang** and J.-N. Hwang, "Self-calibration of traffic surveillance cameras based on moving vehicle appearance and 3-D vehicle modeling," in *Proc. ICIP*, pp. 3064-3068, 2018.

Honors

"Image Analysis and Machine Learning for Scene Understanding"

23rd. INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION

FINALIST BEST IBM TRACK 3 STUDENT PAPER AWARD

is hereby granted to:

Tang, Zheng; Lin, Yen-Shuo; Lee, Kuan-Hui; Hwang, Jenq-Neng; Chuang, Jen-Hui; Fang, Zhijun

for their participation in:

Camera Self-Calibration from Tracking of Moving Persons

4-8 DEC 2016 CANCUN ICC CANCUN MEXICO

Prof. Eduardo Bayro-Corrochano
General Chair ICPR2016
December 10, 2016

"Image Analysis and Machine Learning for Scene Understanding"

23rd. INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION

FINALIST BEST INTEL TRACK 3 STUDENT PAPER AWARD

is hereby granted to:

Tang, Zheng; Lin, Yen-Shuo; Lee, Kuan-Hui; Hwang, Jenq-Neng; Chuang, Jen-Hui; Fang, Zhijun

for their participation in:

Camera Self-Calibration from Tracking of Moving Persons

4-8 DEC 2016 CANCUN ICC CANCUN MEXICO

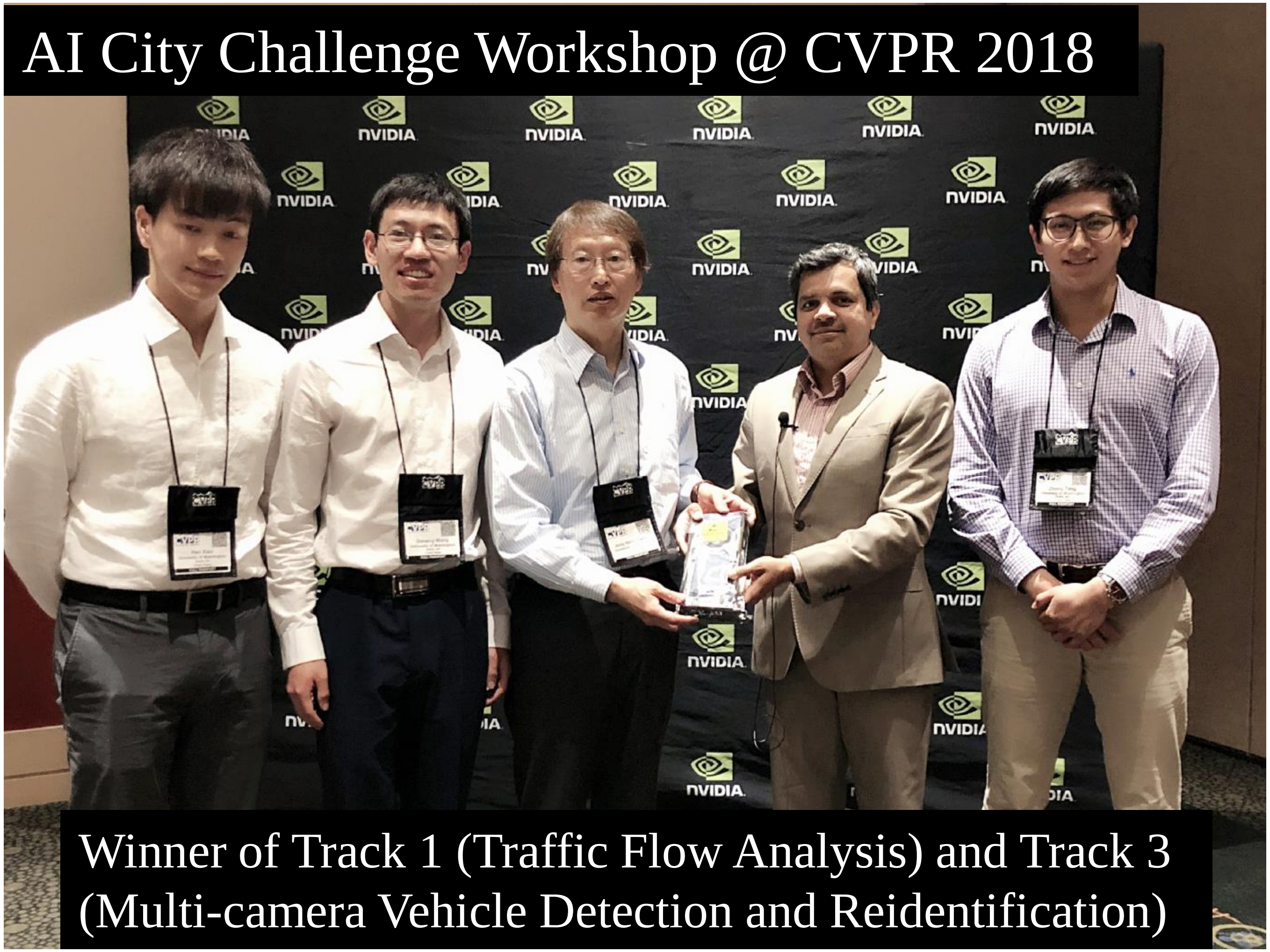
Prof. Eduardo Bayro-Corrochano
General Chair ICPR2016
December 10, 2016

Honors

2017 NVIDIA AI City Challenge

Winner of Track 2 (AI City Applications)

Honors



AI City Challenge Workshop @ CVPR 2018

Winner of Track 1 (Traffic Flow Analysis) and Track 3 (Multi-camera Vehicle Detection and Reidentification)

Acknowledgement

- Advisor: Prof. Jenq-Neng Hwang
- Committee members
  - Prof. Shapiro, Prof. Sun, Prof. Luo, and Prof. Bube
- IPL lab mates, colleagues and co-authors
- Companies for internship and research funding agencies
  - NVIDIA Corp., ArchieMD Inc., Mr. Wanhai Cui, Prism Skylabs, and Madrona Venture Labs
- Family, friends, and Jesus

Thank You!

Q & A