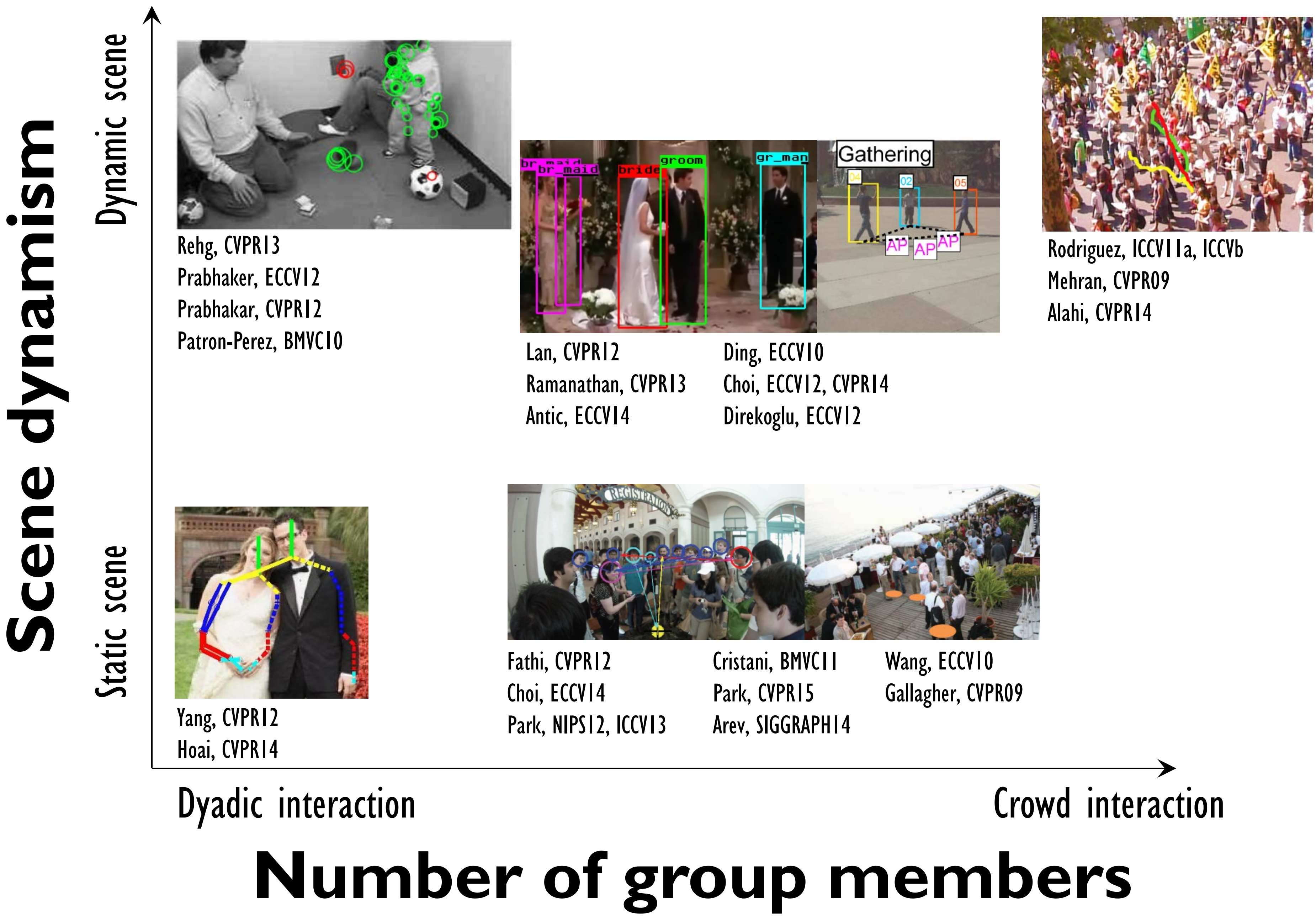
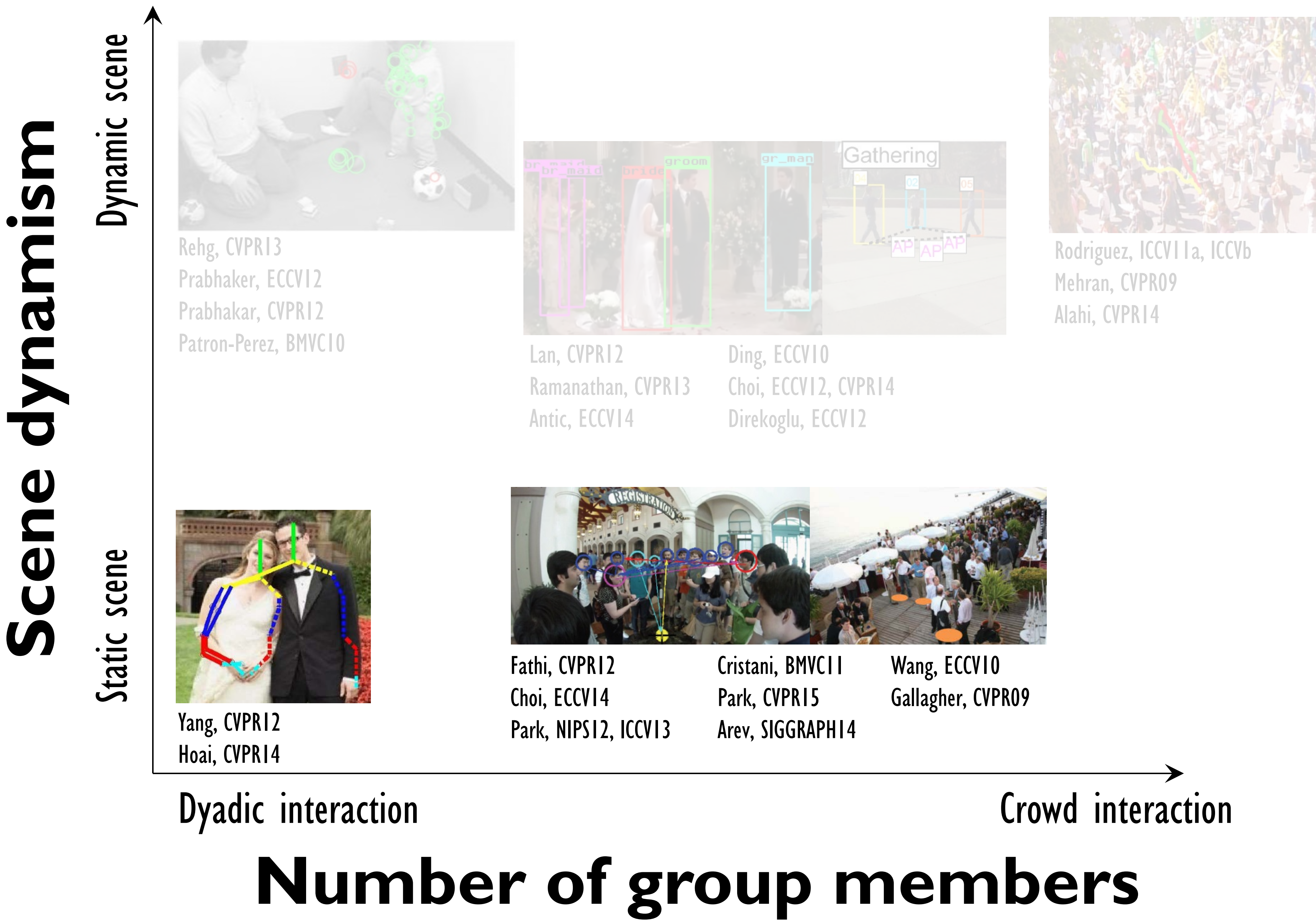
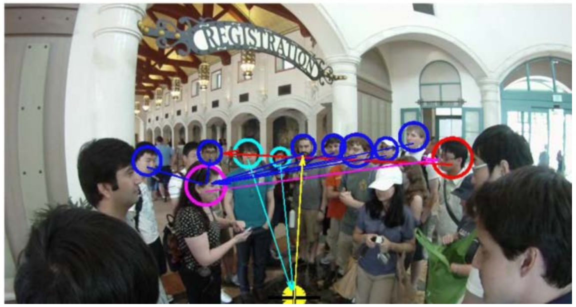
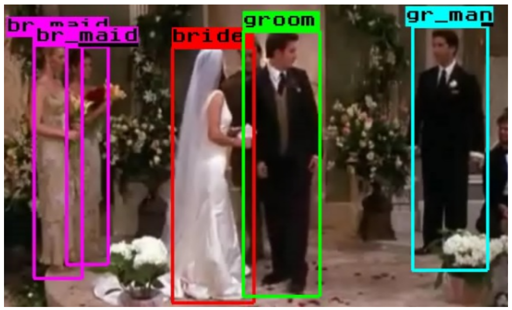
Social Statics

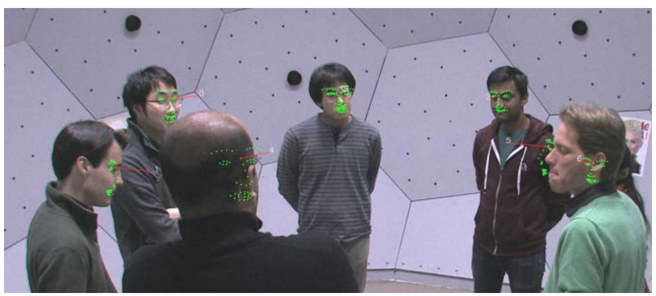

Scene dynamism
Dynamic scene
Static scene
Rehg, CVPR13
Prabhaker, ECCV12
Prabhakar, CVPR12
Patron-Perez, BMVC10
br_maid
bride
groom
gr_man
Gathering
AP
Lan, CVPR12
Ramanathan, CVPR13
Antic, ECCV14
Ding, ECCV10
Choi, ECCV12, CVPR14
Direkoglu, ECCV12
Rodriguez, ICCV11a, ICCVb
Mehran, CVPR09
Alahi, CVPR14
Yang, CVPR12
Hoai, CVPR14
Fathi, CVPR12
Choi, ECCV14
Park, NIPS12, ICCV13
Cristani, BMVC11
Park, CVPR15
Arev, SIGGRAPH14
Wang, ECCV10
Gallagher, CVPR09
Dyadic interaction
Crowd interaction
Number of group members

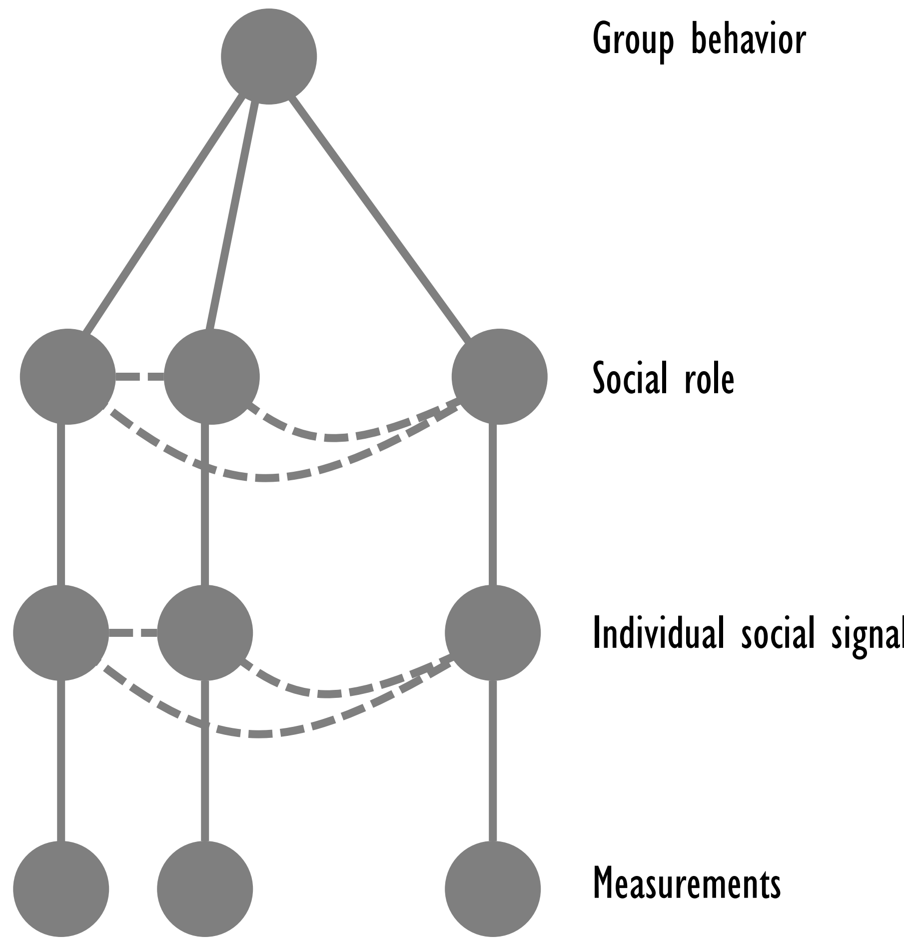


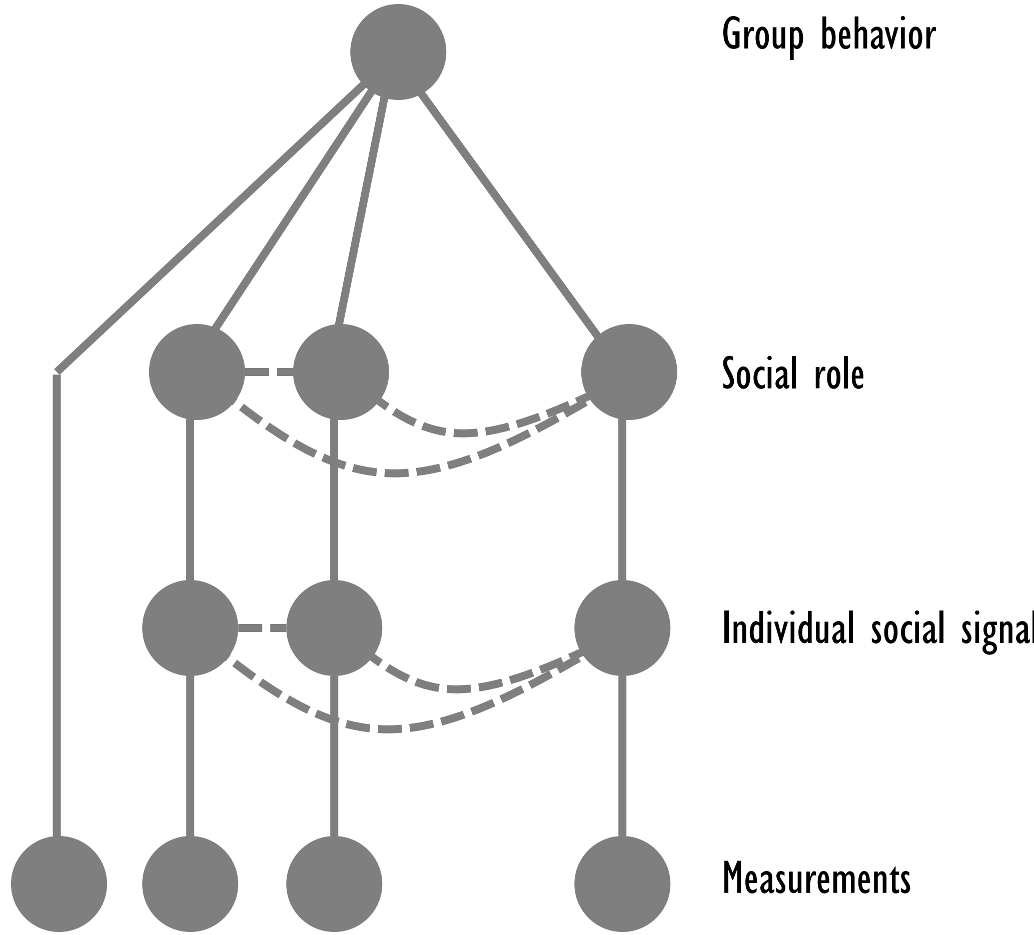
Scene dynamism
Dynamic scene
Static scene
Rehg, CVPR13
Prabhaker, ECCV12
Prabhakar, CVPR12
Patron-Perez, BMVC10
Gathering
Lan, CVPR12
Ramanathan, CVPR13
Antic, ECCV14
Ding, ECCV10
Choi, ECCV12, CVPR14
Direkoglu, ECCV12
Rodriguez, ICCV11a, ICCVb
Mehran, CVPR09
Alahi, CVPR14
Yang, CVPR12
Hoai, CVPR14
REGISTRATION
Fathi, CVPR12
Choi, ECCV14
Park, NIPS12, ICCV13
Cristani, BMVC11
Park, CVPR15
Arev, SIGGRAPH14
Wang, ECCV10
Gallagher, CVPR09
Dyadic interaction
Crowd interaction
Number of group members

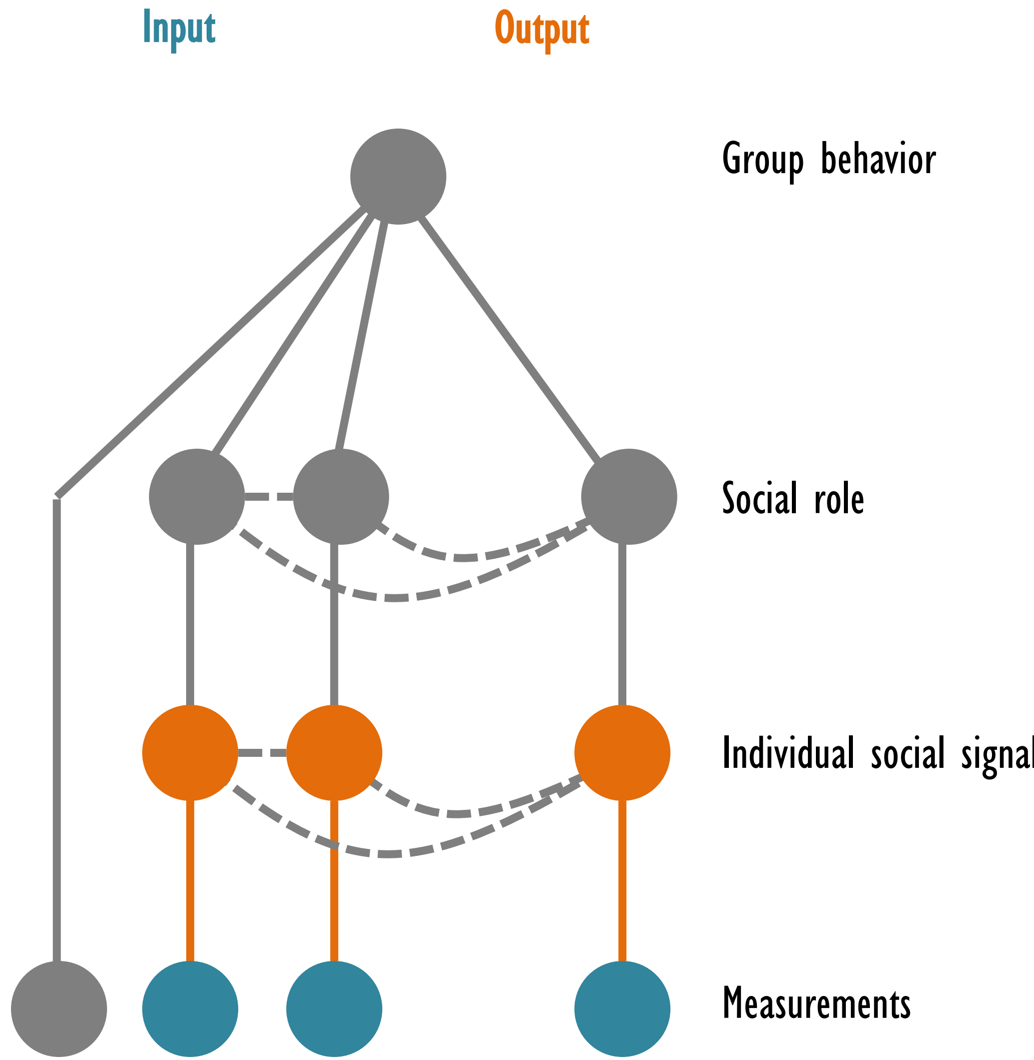

REGISTRATION

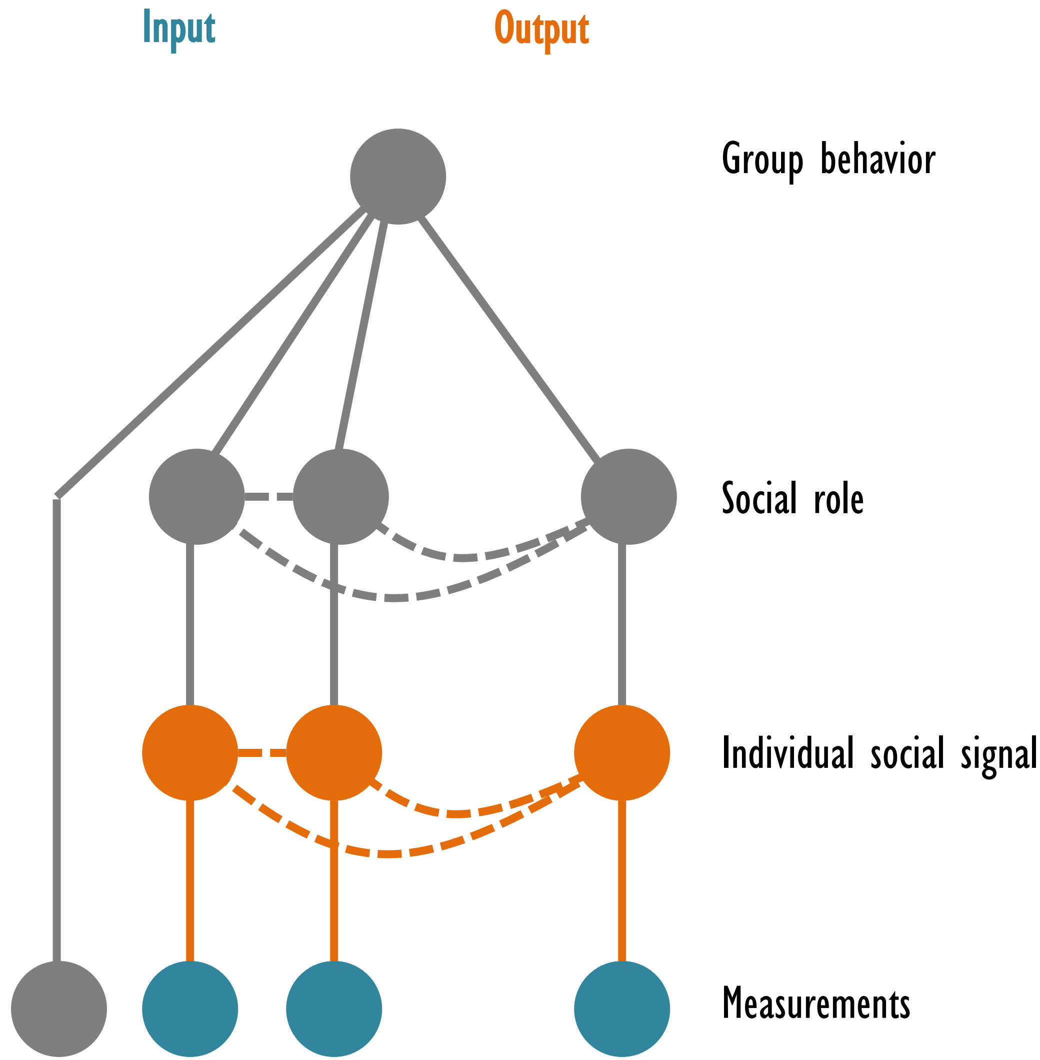

br_maid
bride
groom
gr_man

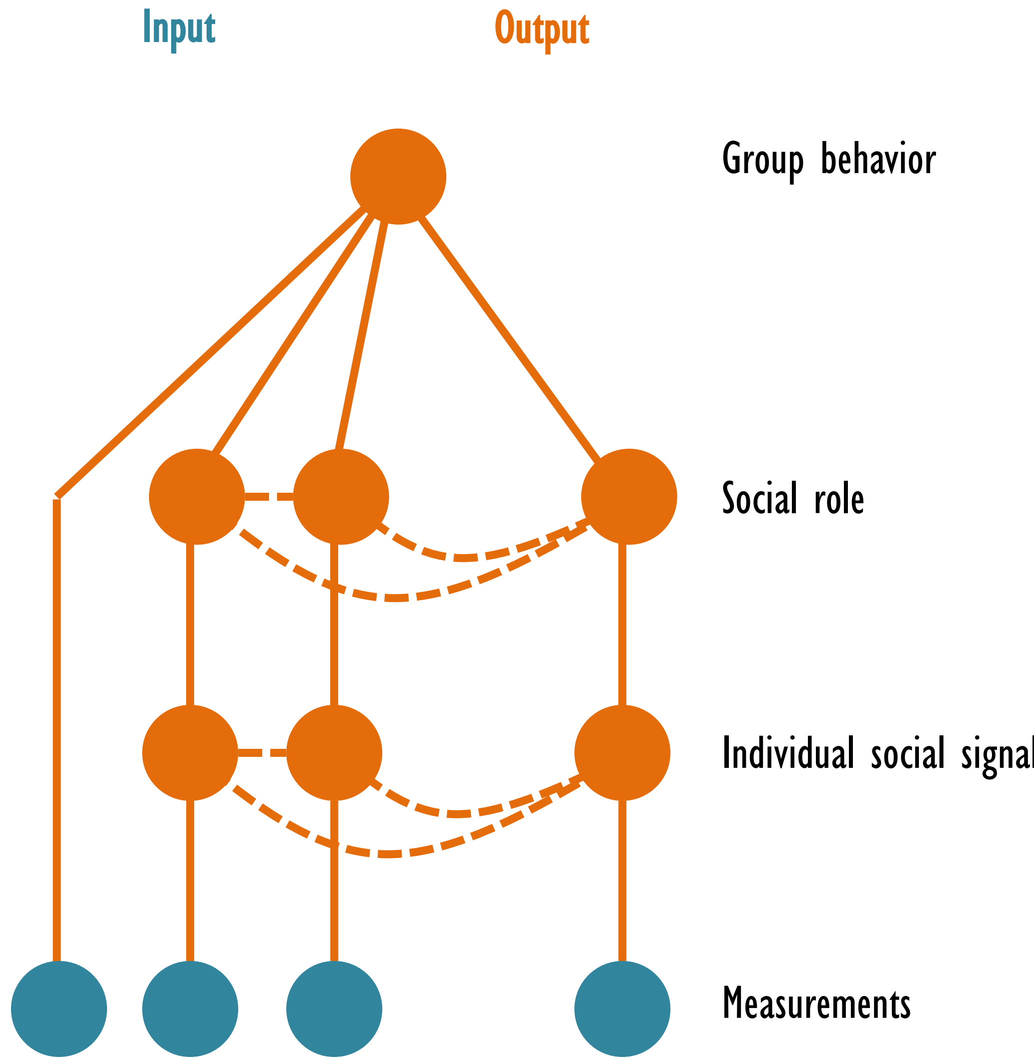


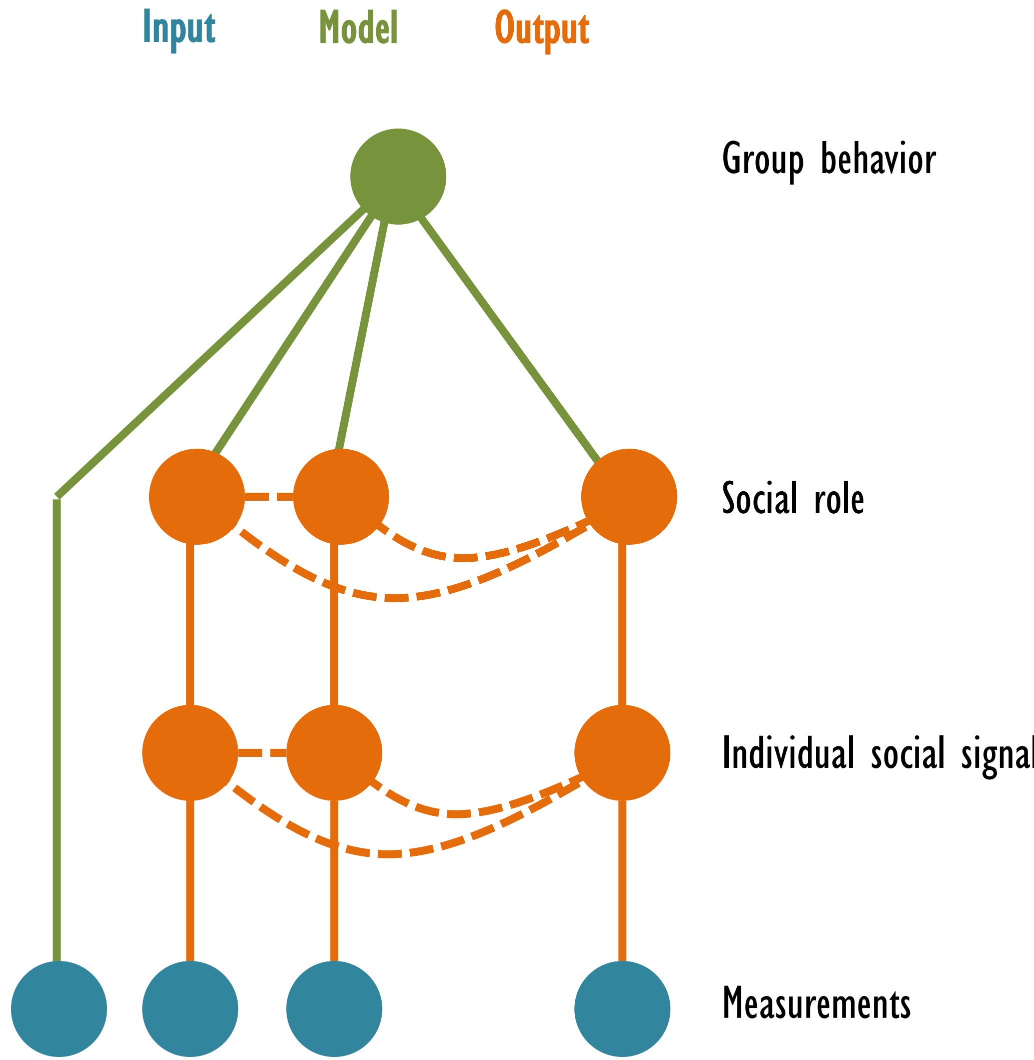

Group behavior
Social role
Individual social signal
Measurements

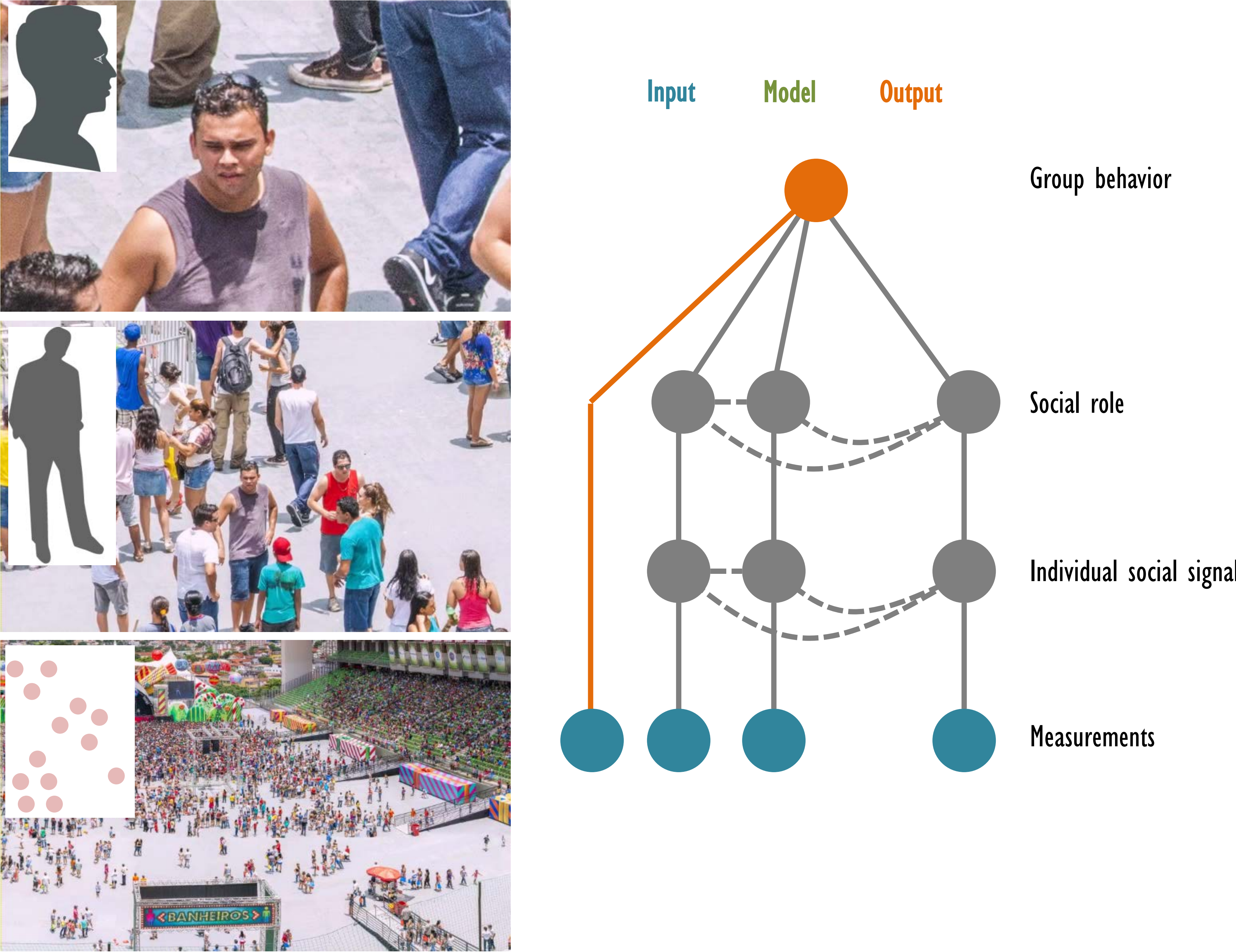

Group behavior
Social role
Individual social signal
Measurements


Input
Output
Group behavior
Social role
Individual social signal
Measurements

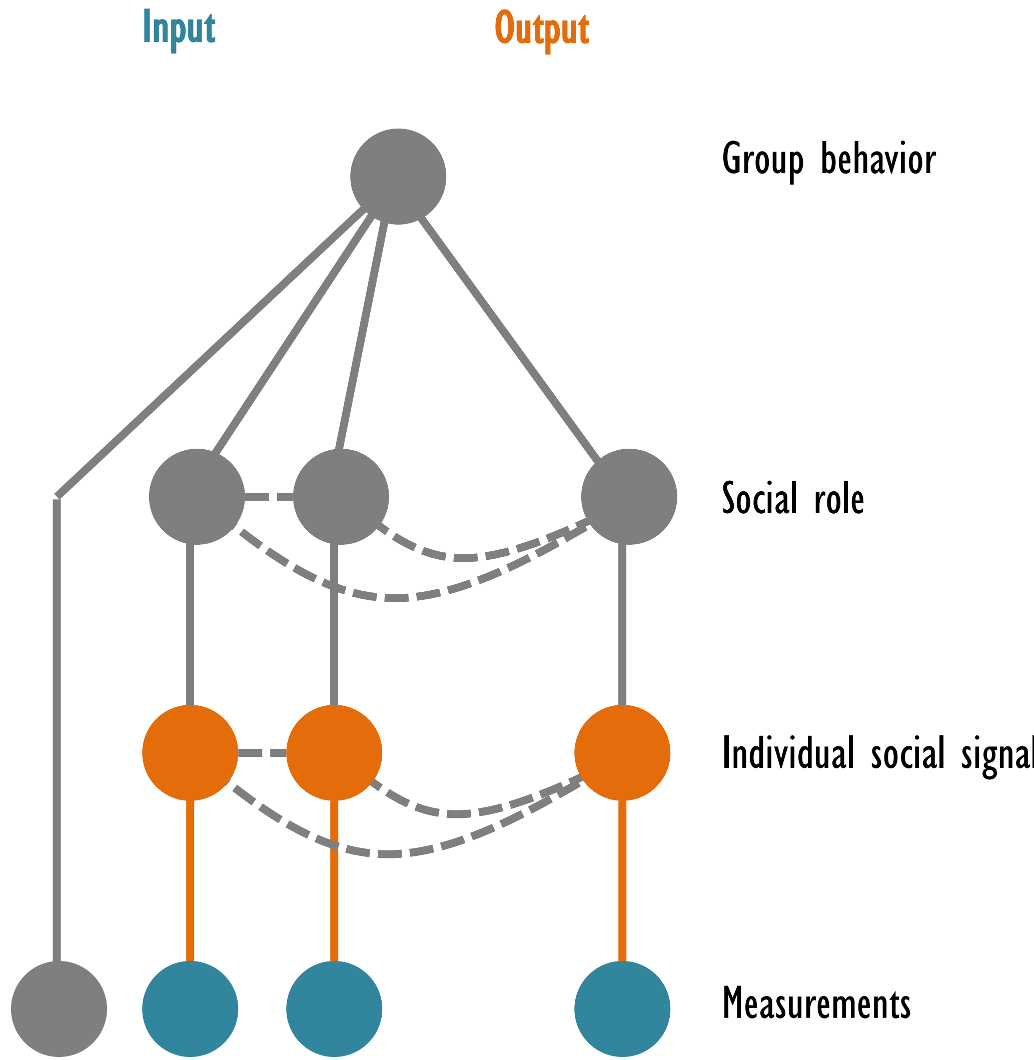

Input
Output
Group behavior
Social role
Individual social signal
Measurements

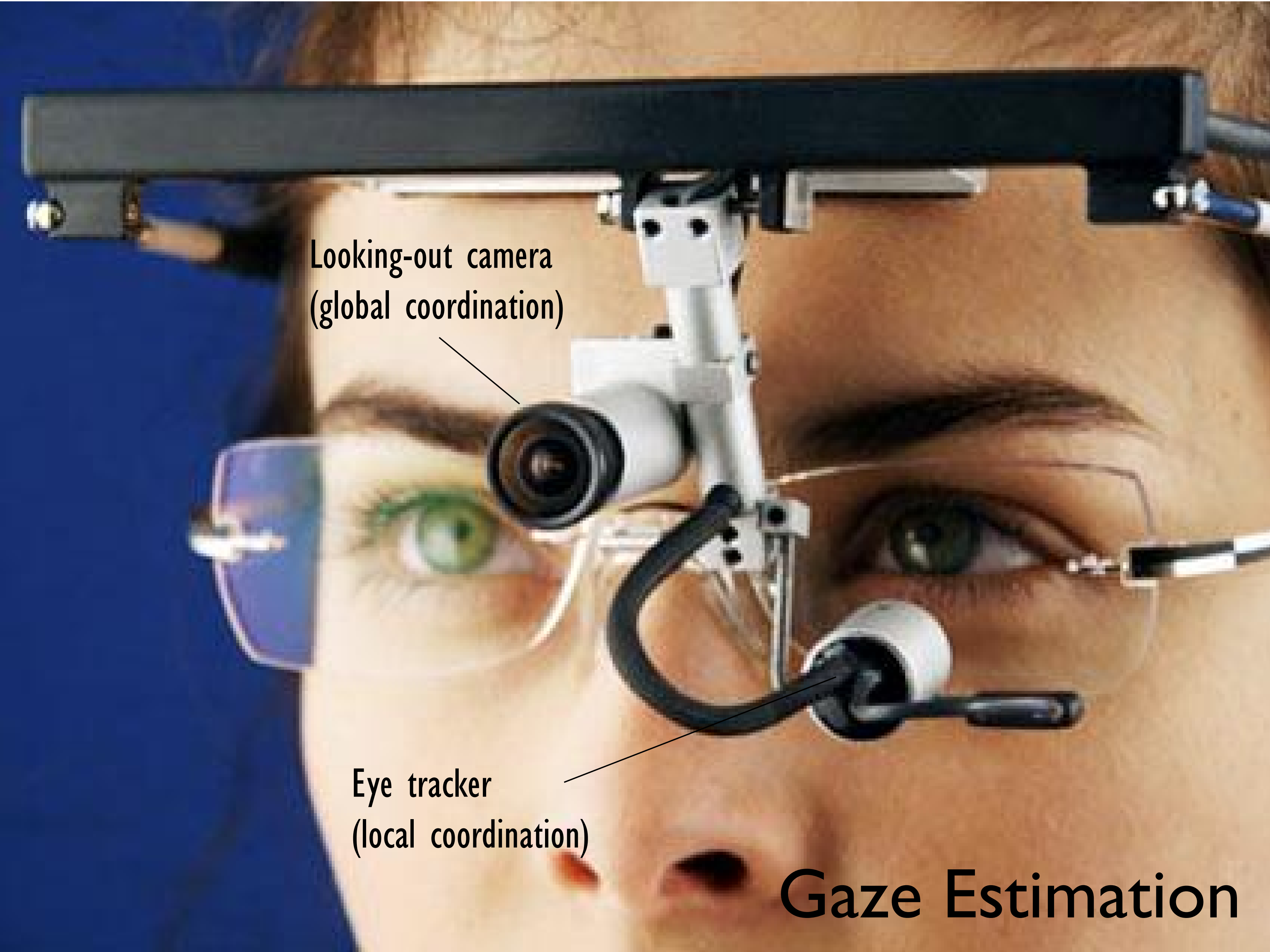

Input
Output
Group behavior
Social role
Individual social signal
Measurements


Input
Model
Output
Group behavior
Social role
Individual social signal
Measurements

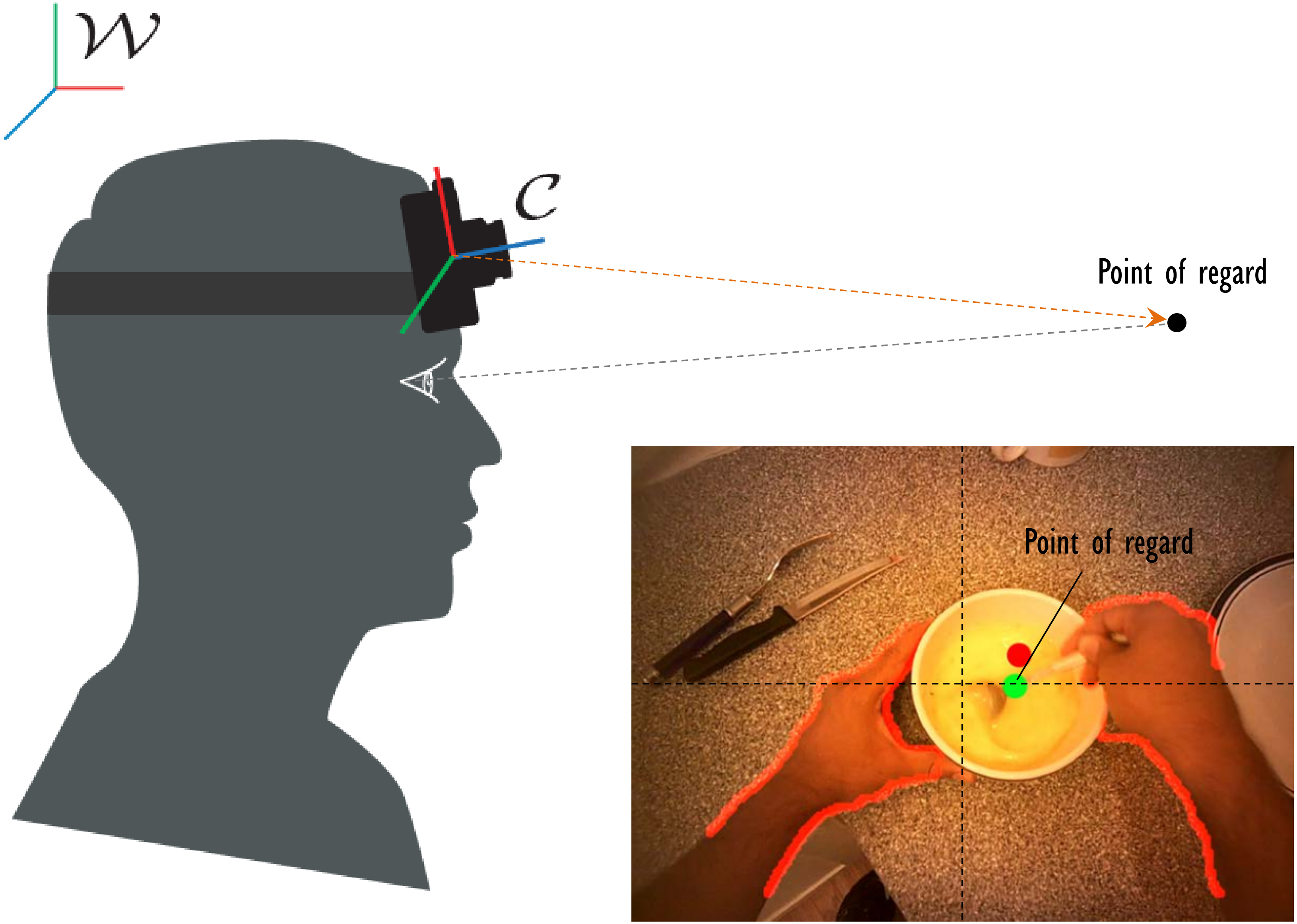


Input
Model
Output
Group behavior
Social role
Individual social signal
Measurements
BANHEIROS


Input
Output
Group behavior
Social role
Individual social signal
Measurements


Looking-out camera
(global coordination)
Eye tracker
(local coordination)
Gaze Estimation

Gaze Estimation w/o Eye Tracker

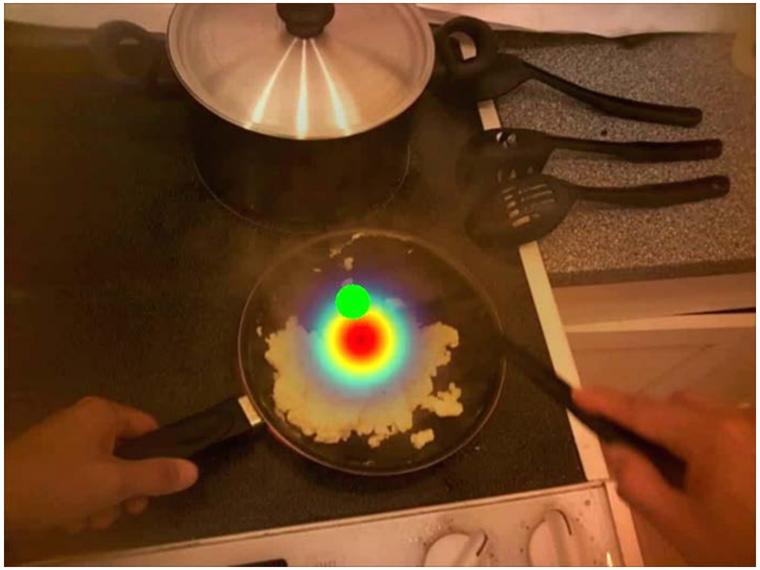


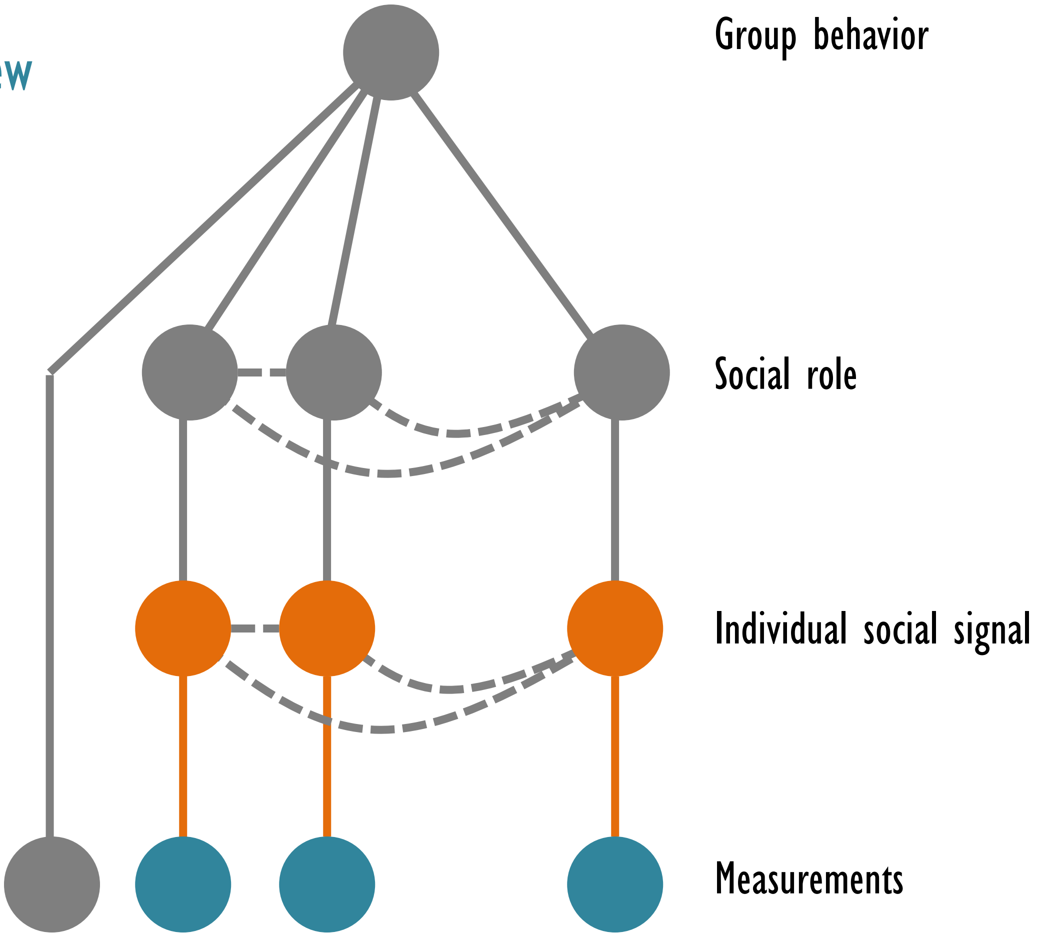
Social Signal Perception (Gaze)

[Li ICCV13]

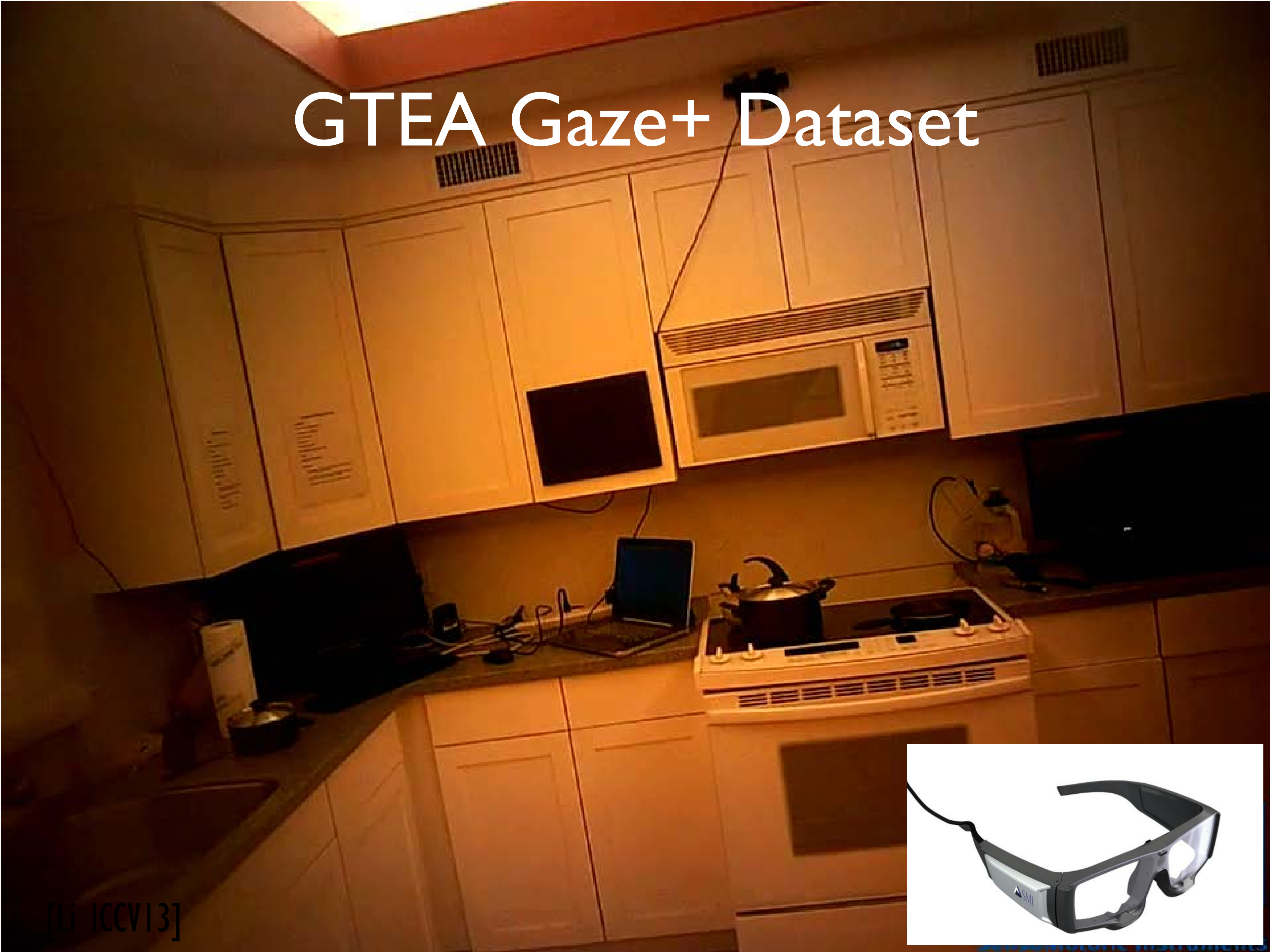
Input: image or video of first person view

Output: localization of fixation point.





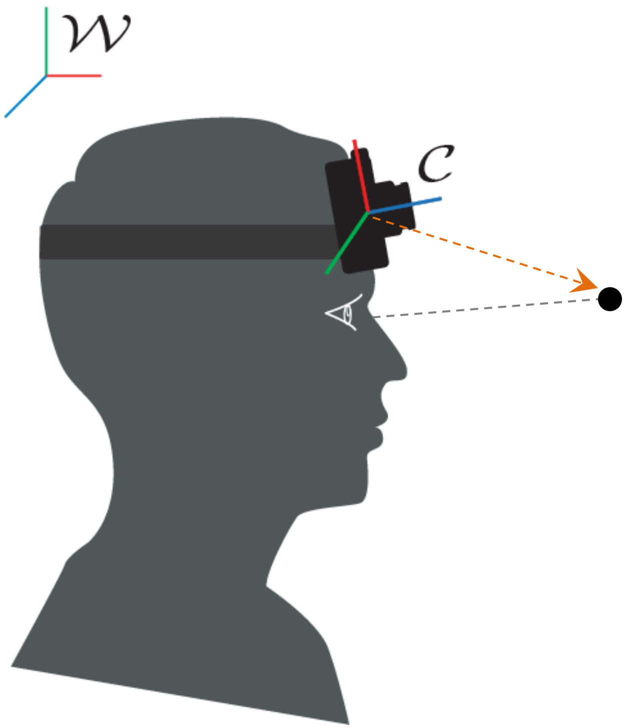
GTEA Gaze+ Dataset



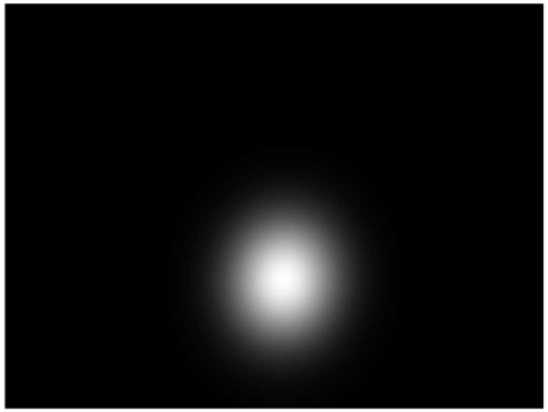
[Li ICCV13]

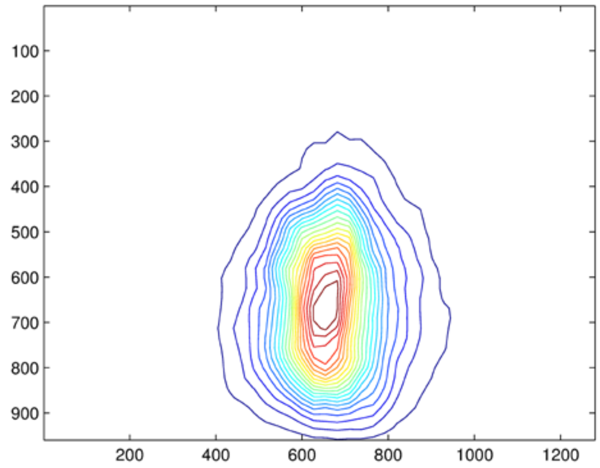
Egocentric Cue I

Eye-in-head Orientation (Center Prior)



GTEA Gaze+ Dataset





[Li ICCV13]

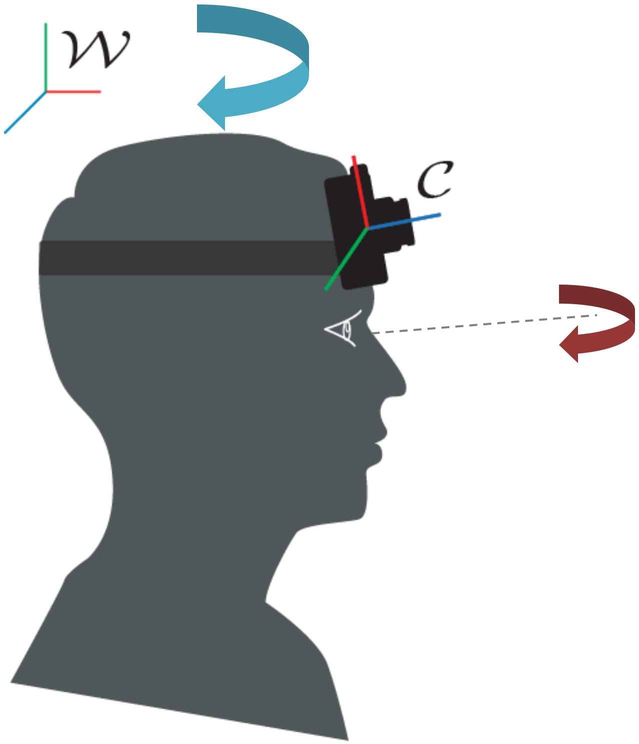
Egocentric Cue II

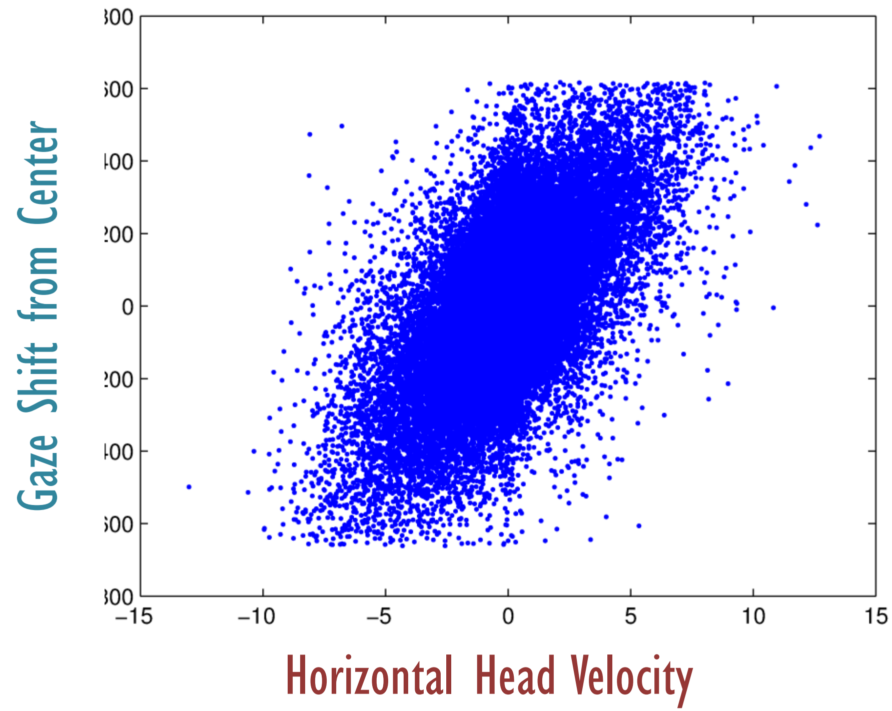
Head Motion

[Li ICCV13]

Egocentric Cue II

Head Motion

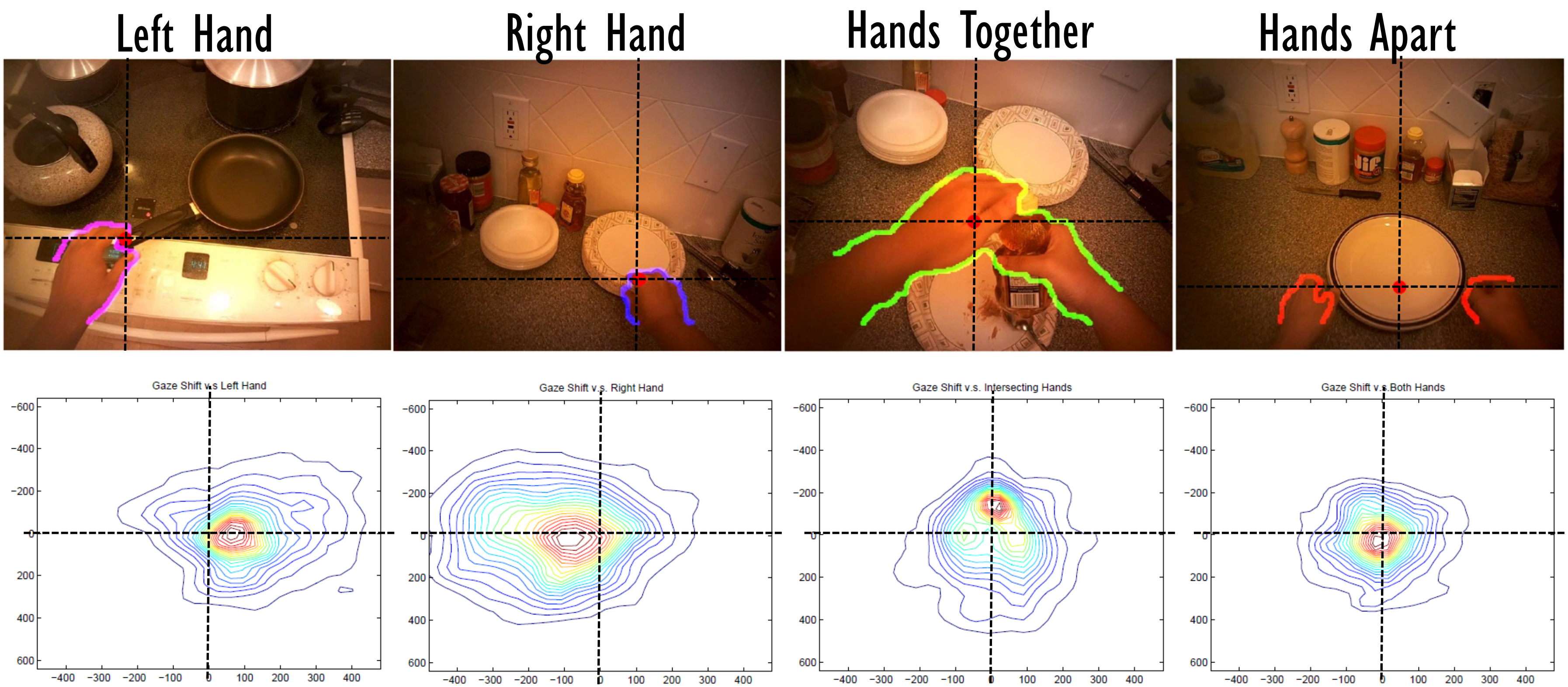




[Li ICCV13]

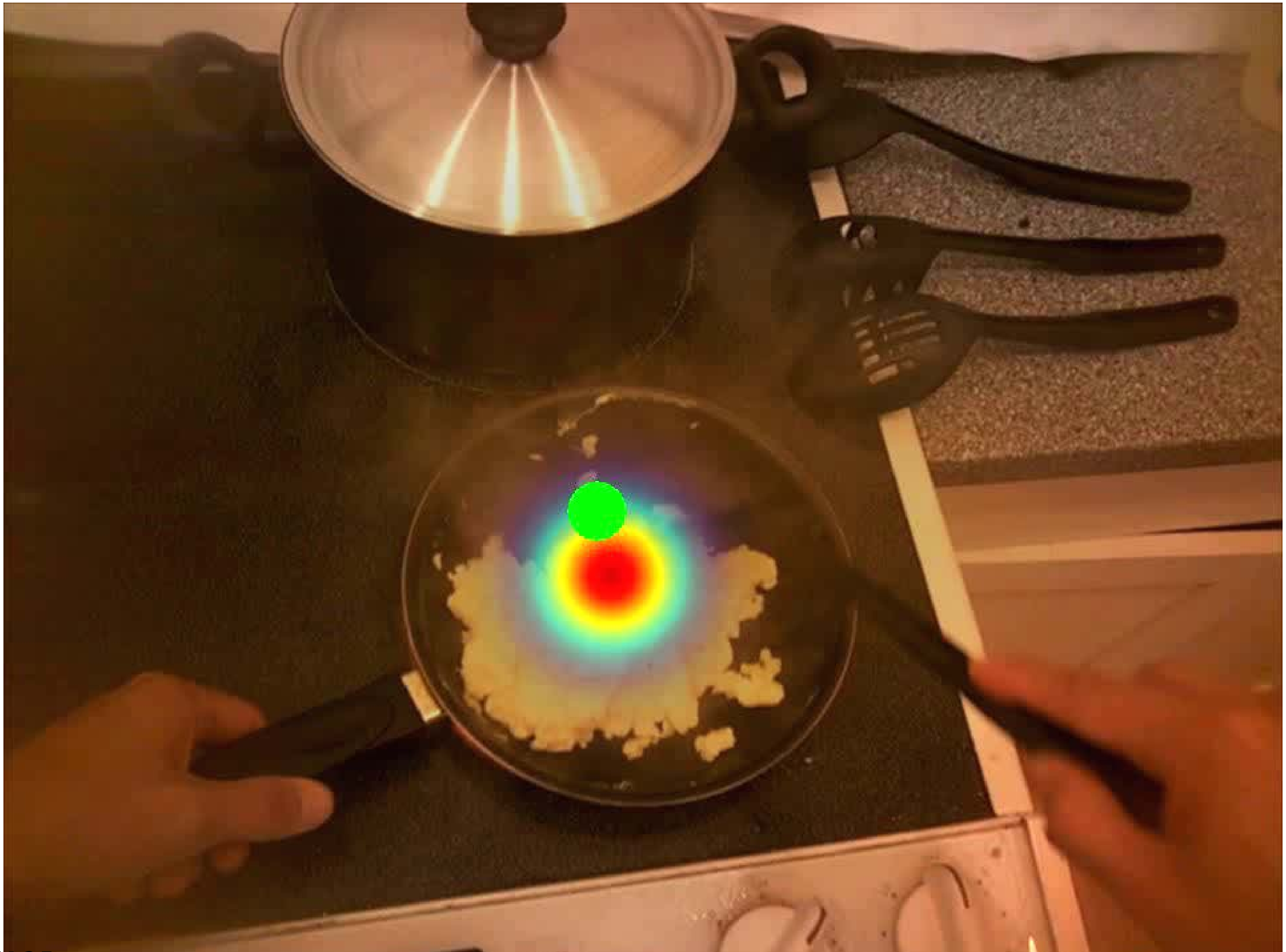
Egocentric Cue III

Hand Position



[Li ICCV13]

Gaze Prediction



[Li ICCV13]

Application to Foreground Segmentation

Foreground Object

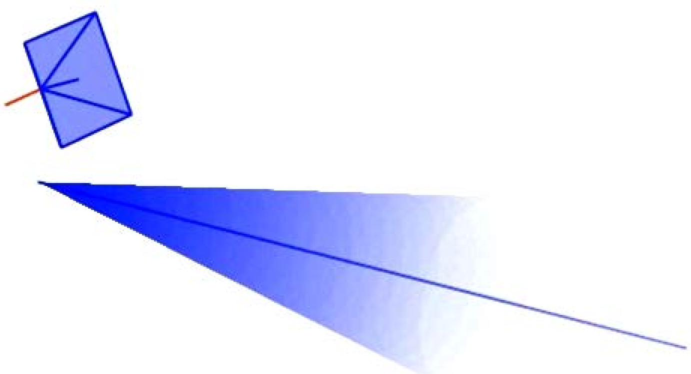
[Li ICCV13]

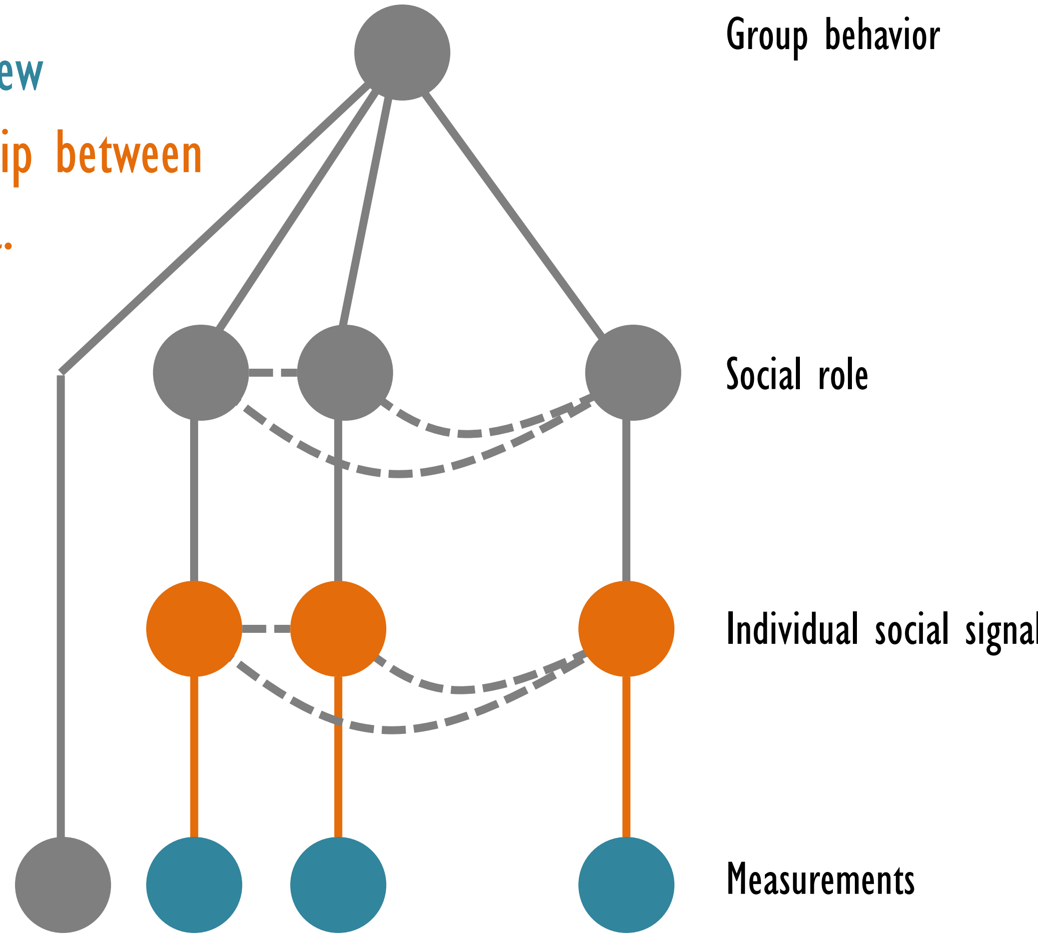
Social Signal Perception (Gaze)

[Park NIPS12]

Input: image or video of first person view

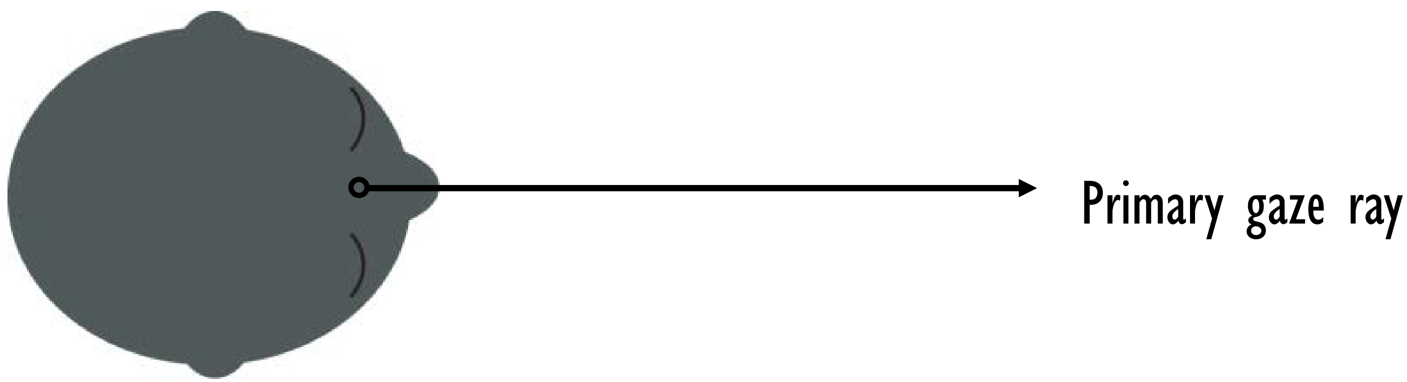
Output: to calibrate the fixed relationship between gaze and head-mounted camera.



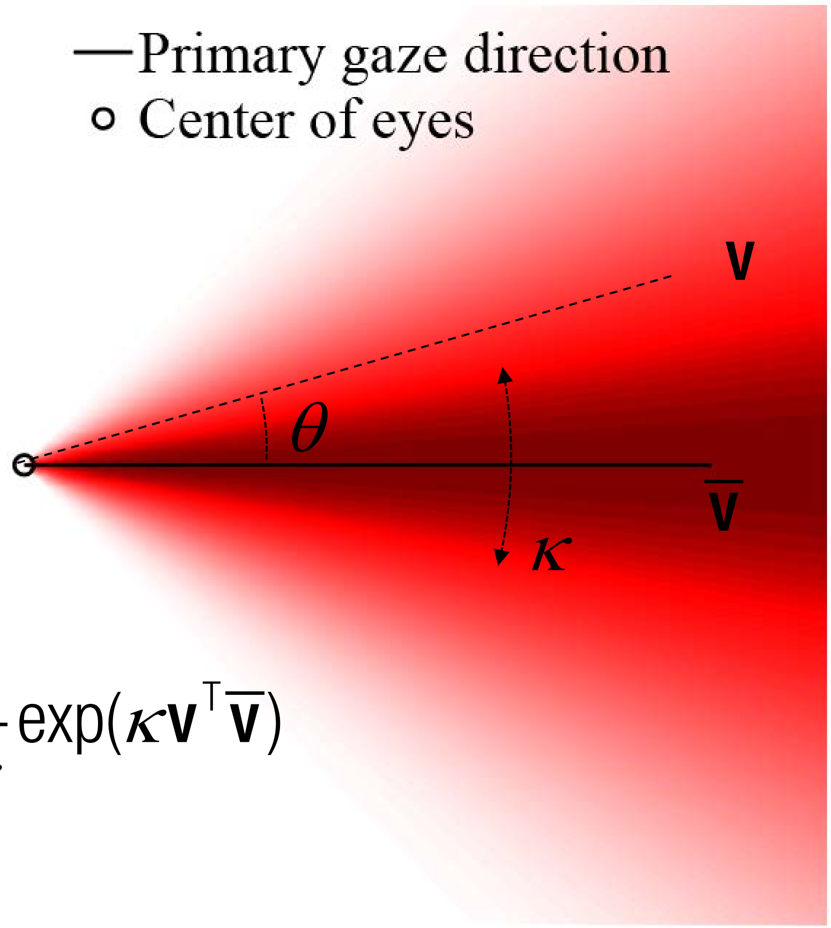


Eye-in-head Motion

[Park NIPS12]



Gaze Distribution



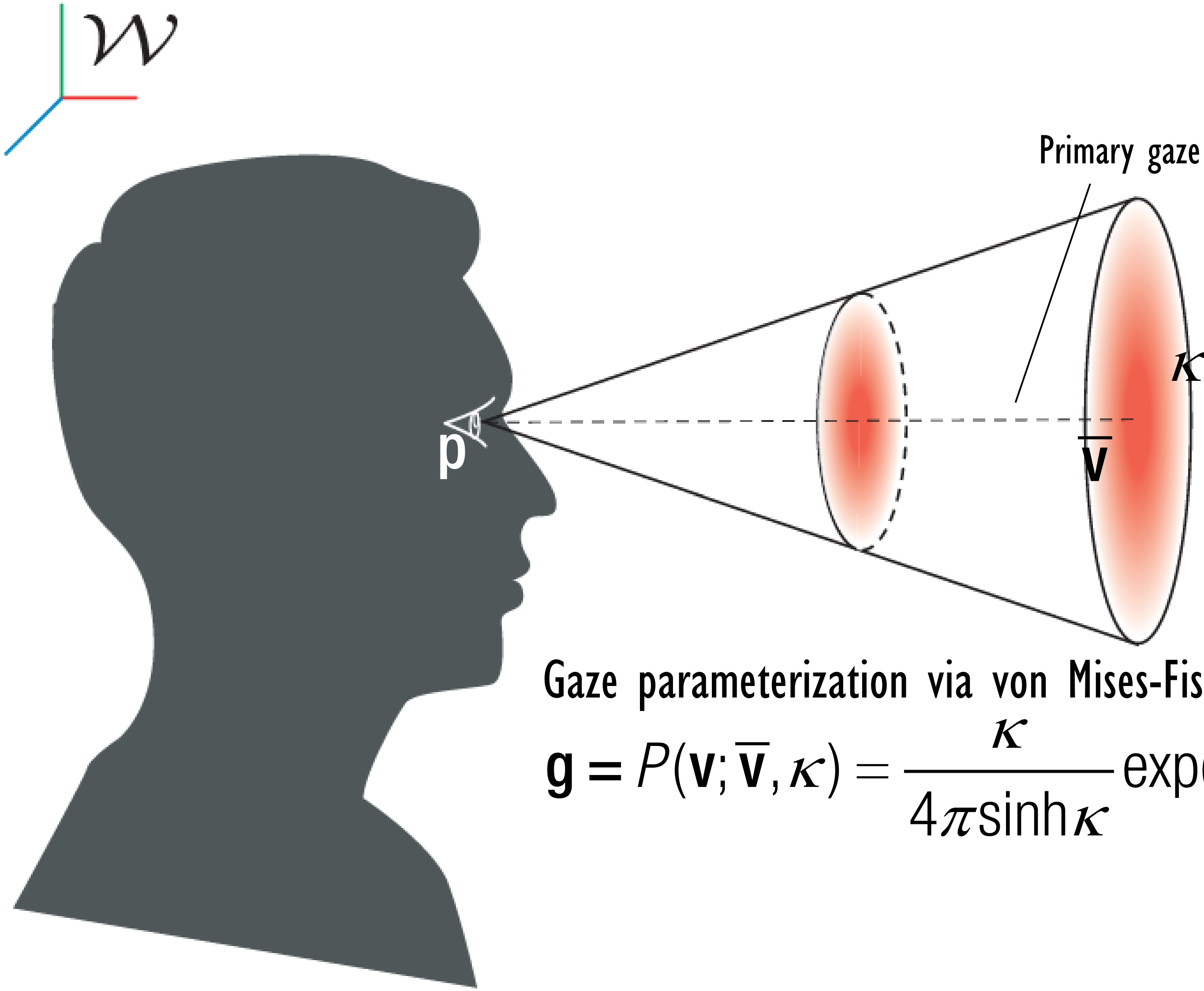
$$\cos(\theta) = \mathbf{v}^{\top}\bar{\mathbf{v}}$$

von Mises-Fisher distribution

$$P(\mathbf{v}; \bar{\mathbf{v}}, \kappa) = \frac{\kappa}{4\pi \sinh \kappa} \exp(\kappa \mathbf{v}^{\top}\bar{\mathbf{v}})$$

[Park NIPS12]

Cone-shaped Gaze Model

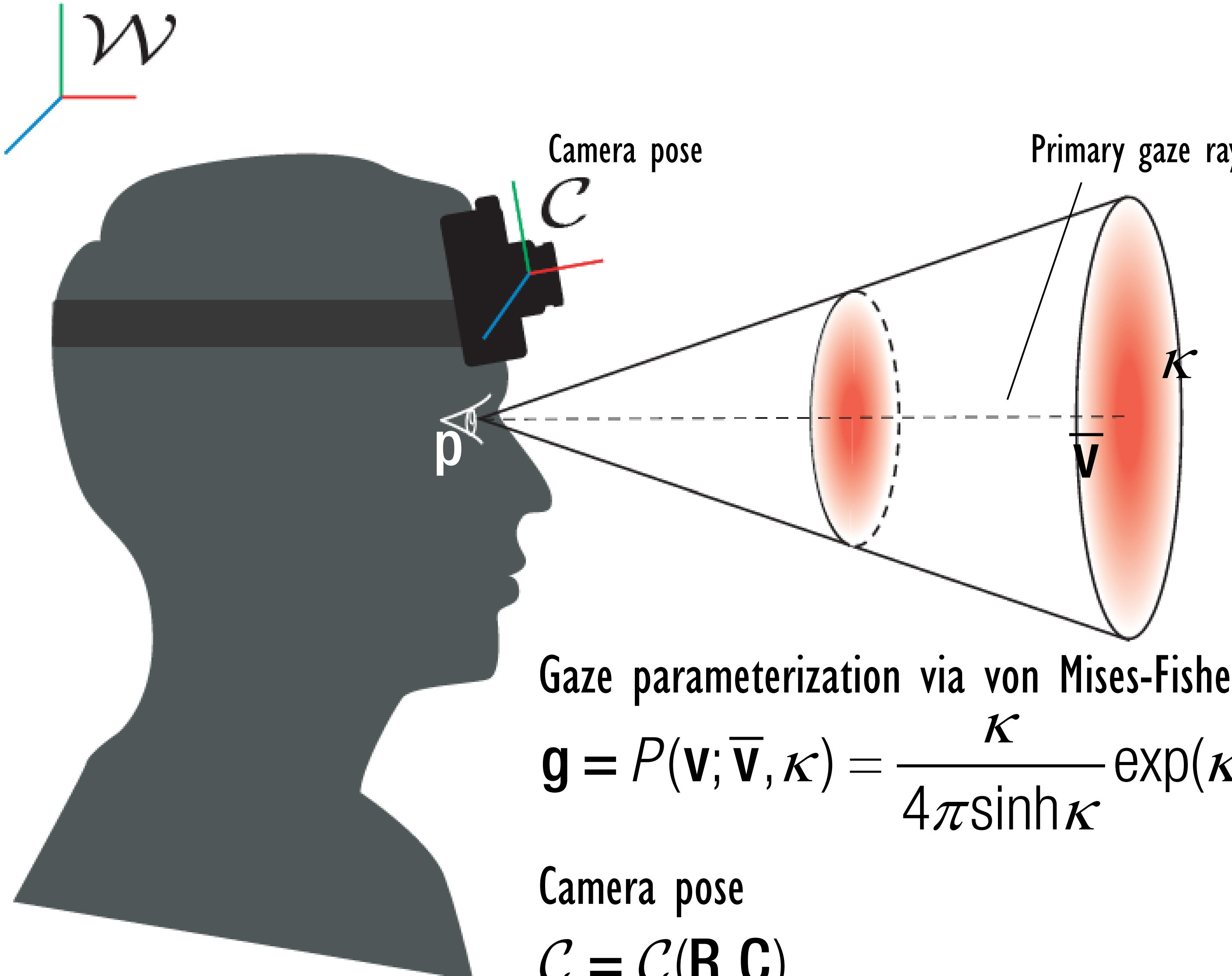


Gaze parameterization via von Mises-Fisher distribution

$$\mathbf{g} = P(\mathbf{v}; \overline{\mathbf{v}}, \kappa) = \frac{\kappa}{4\pi \sinh \kappa} \exp(\kappa \mathbf{v}^{\mathsf{T}} \overline{\mathbf{v}})$$

[Park NIPS12]

Gaze Calibration



Gaze parameterization via von Mises-Fisher distribution

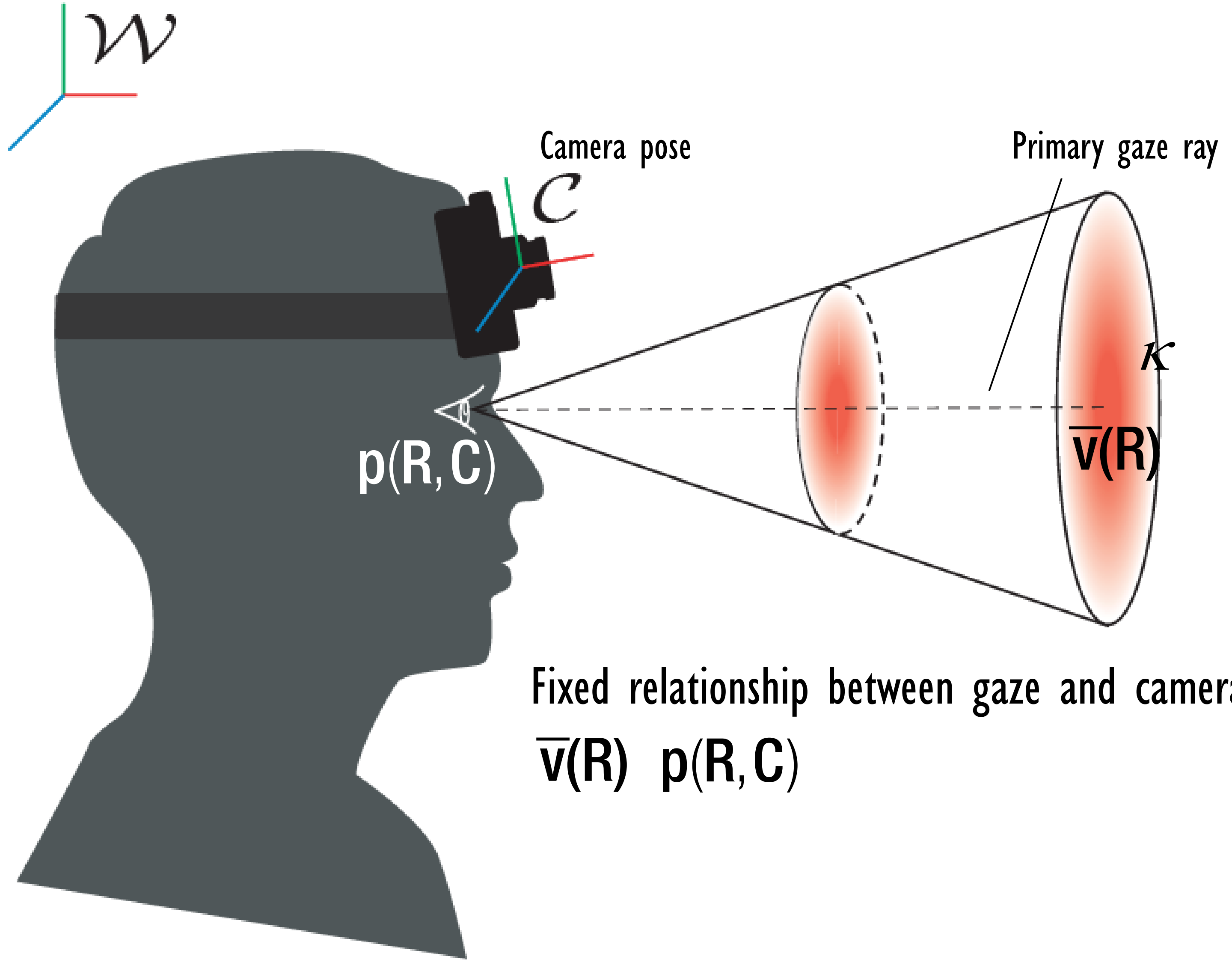
$$\mathbf{g} = P(\mathbf{v}; \overline{\mathbf{v}}, \kappa) = \frac{\kappa}{4\pi \sinh \kappa} \exp(\kappa \mathbf{v}^\top \overline{\mathbf{v}})$$

Camera pose

$$\mathcal{C} = \mathcal{C}(\mathbf{R}, \mathbf{C})$$

[Park NIPS12]

Gaze Calibration

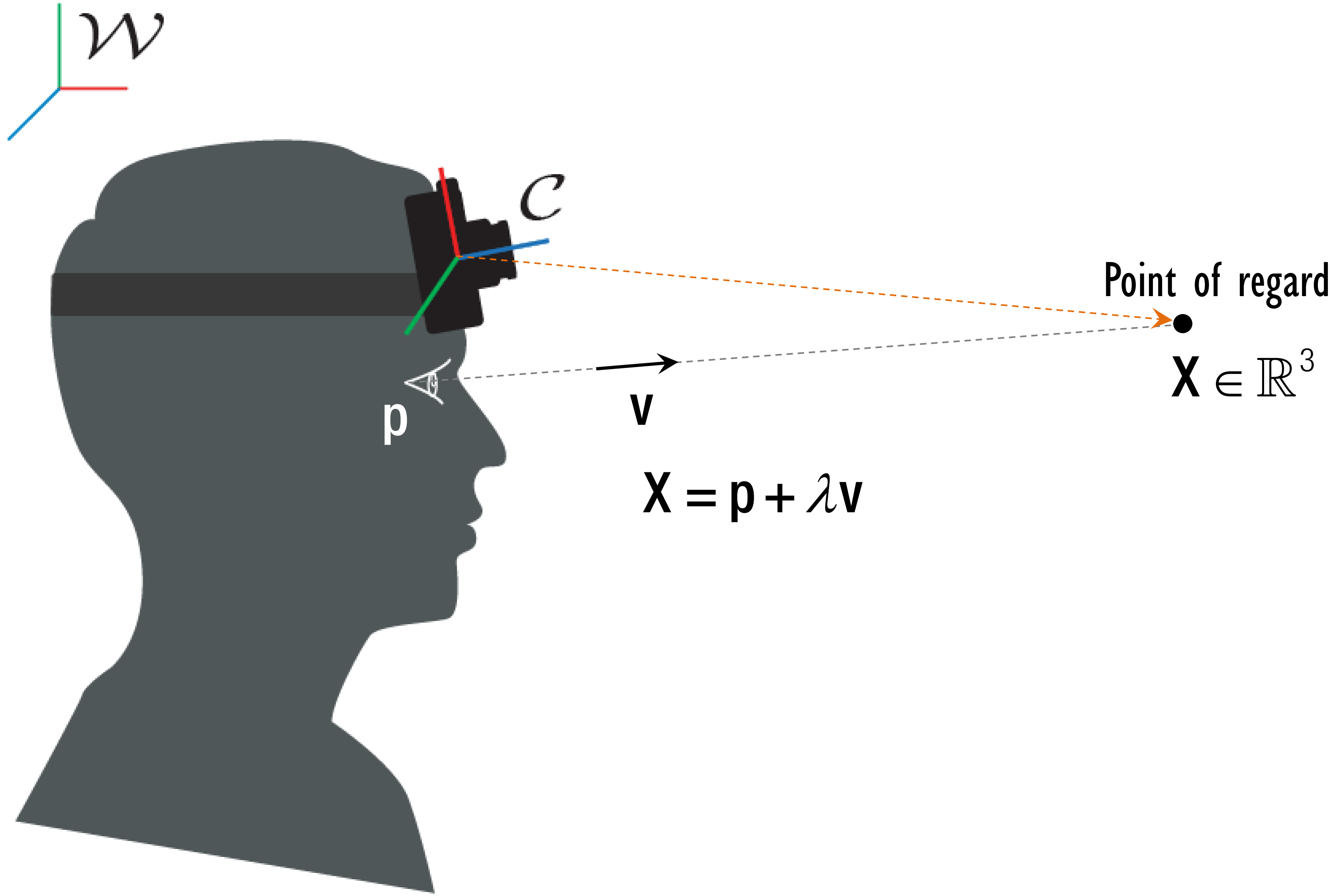


Fixed relationship between gaze and camera

$\overline{\mathbf{v}}(\mathbf{R}) \quad \mathbf{p}(\mathbf{R}, \mathbf{C})$

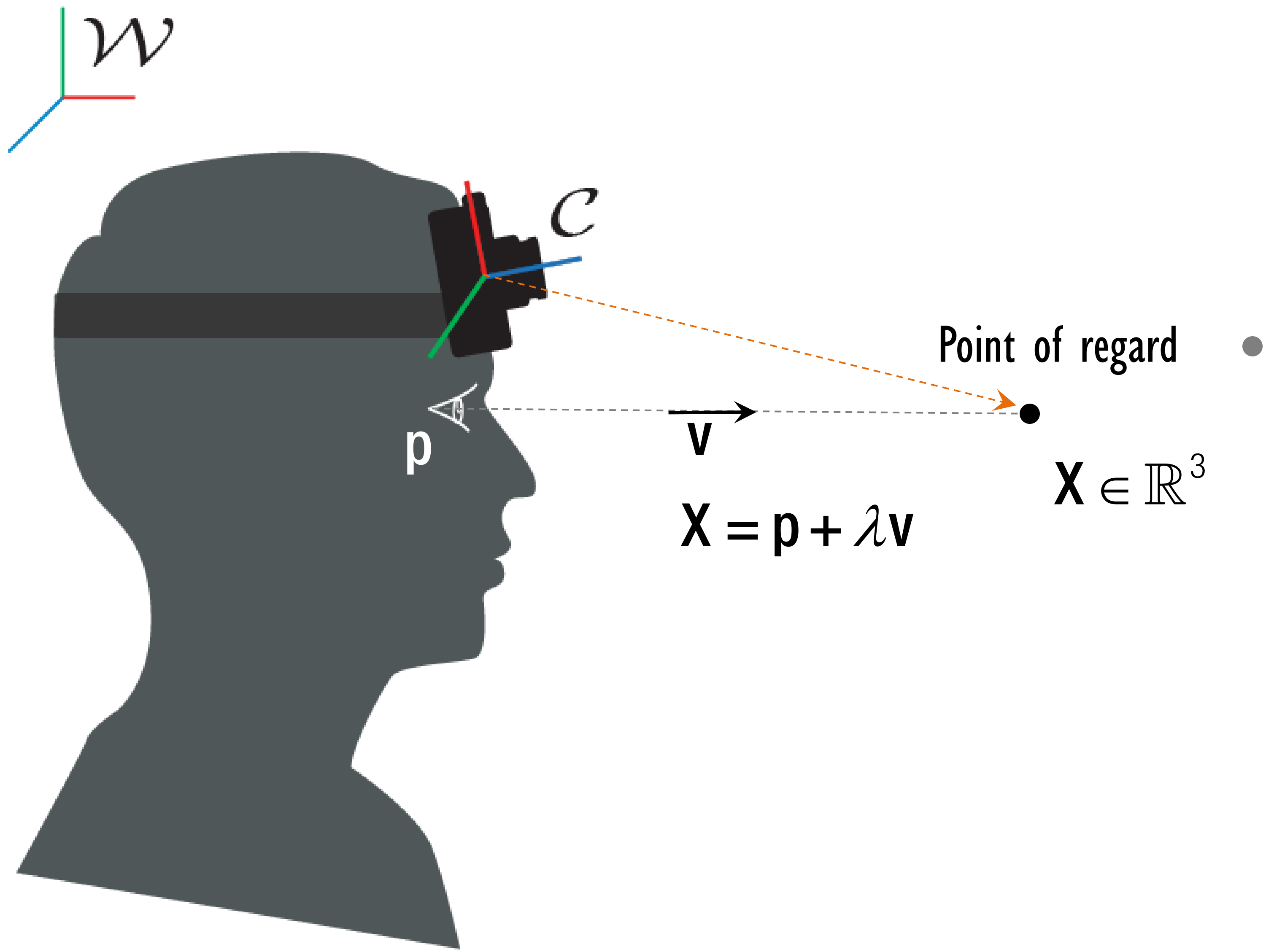
[Park NIPS12]

Gaze Calibration



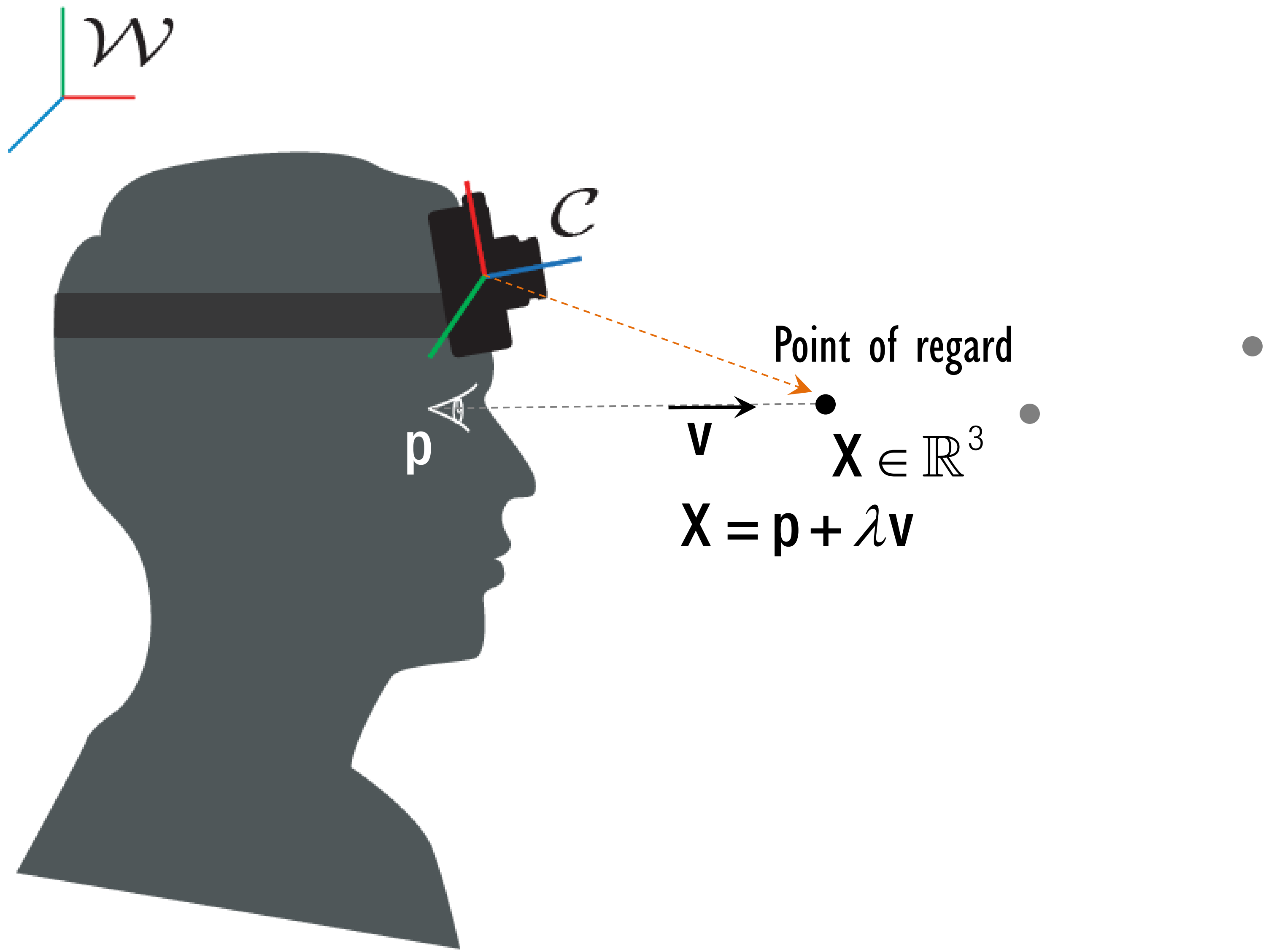
[Park NIPS12]

Gaze Calibration



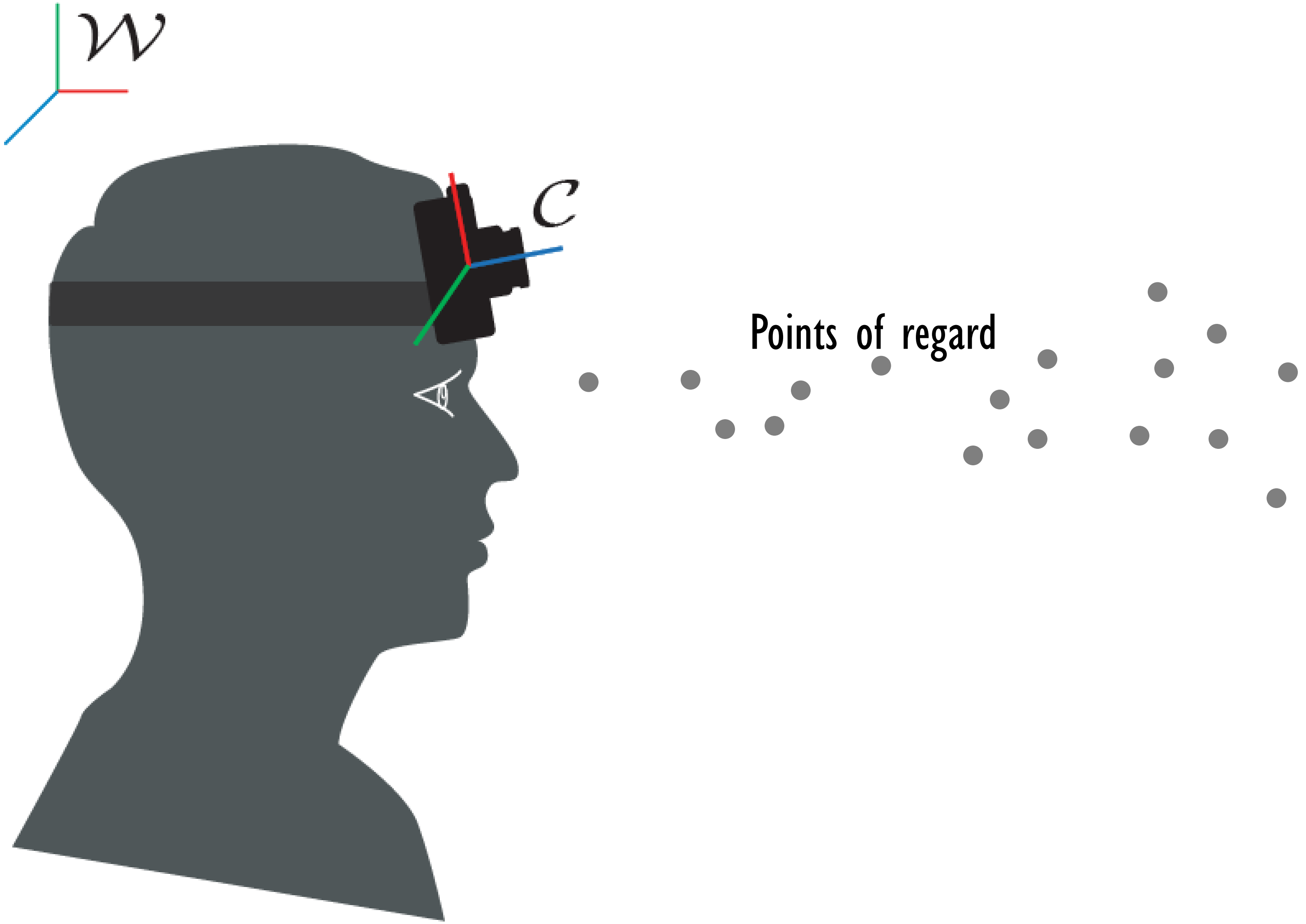
[Park NIPS12]

Gaze Calibration



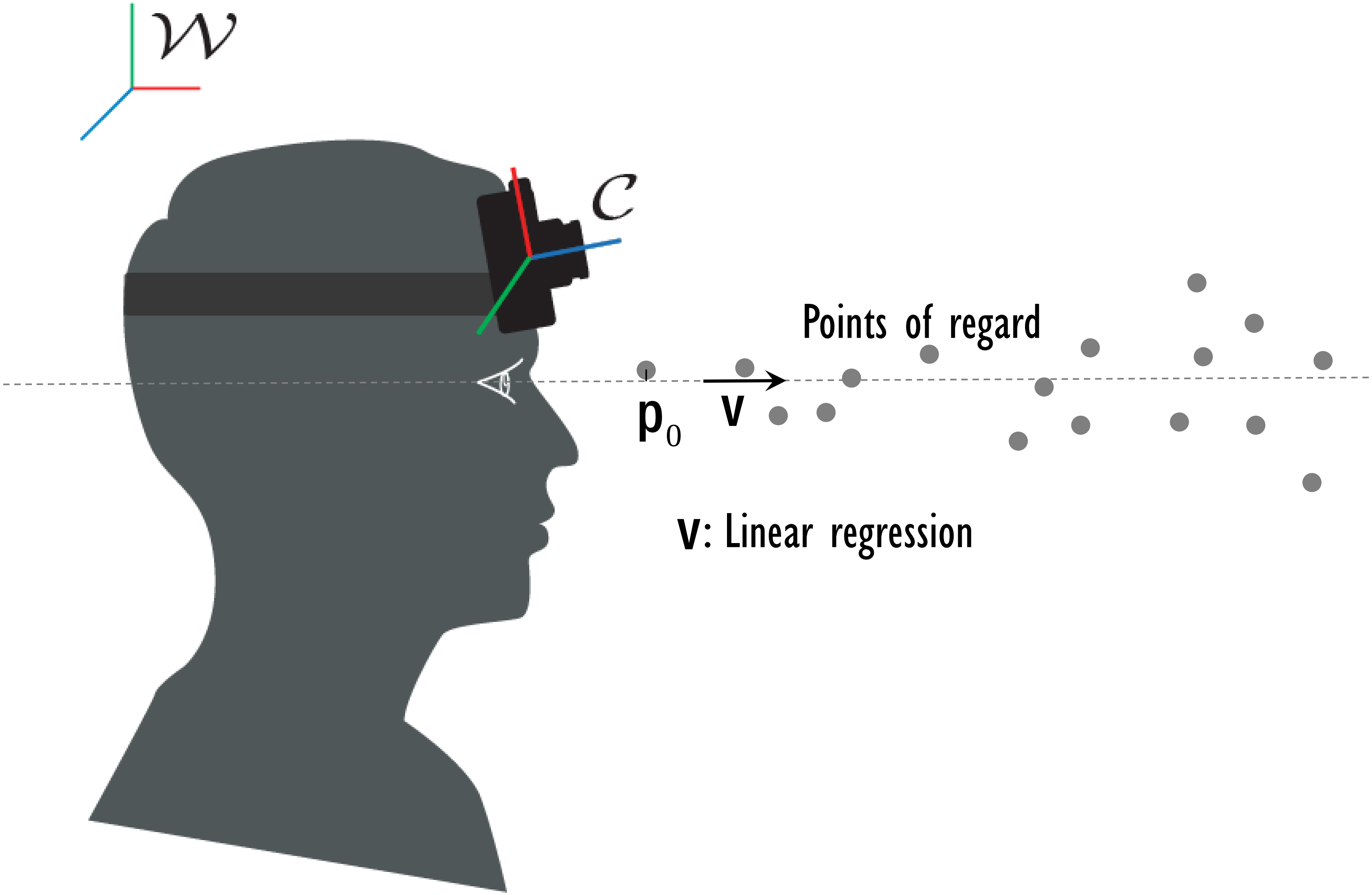
[Park NIPS12]

Gaze Calibration



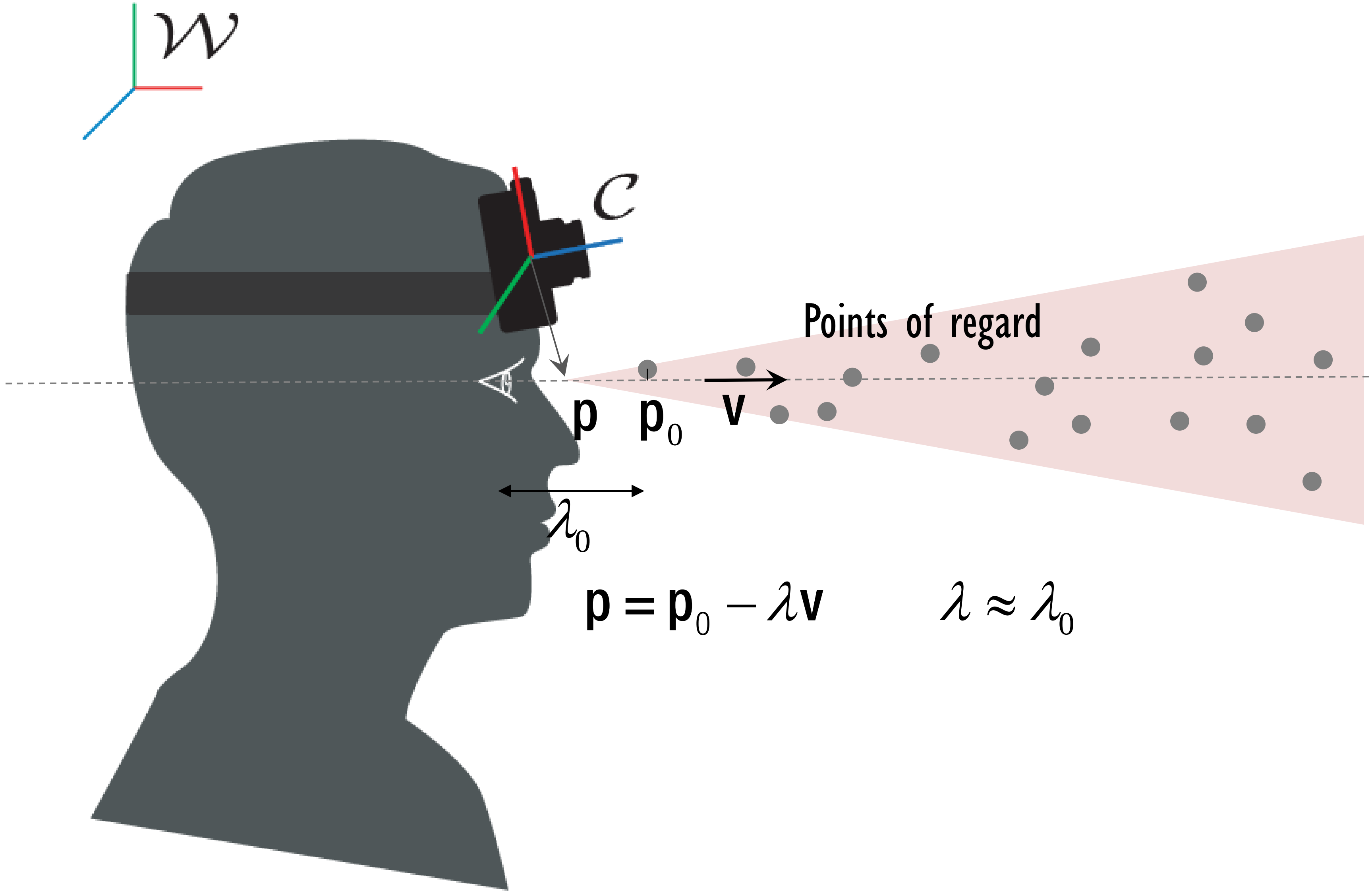
[Park NIPS12]

Gaze Calibration



[Park NIPS12]

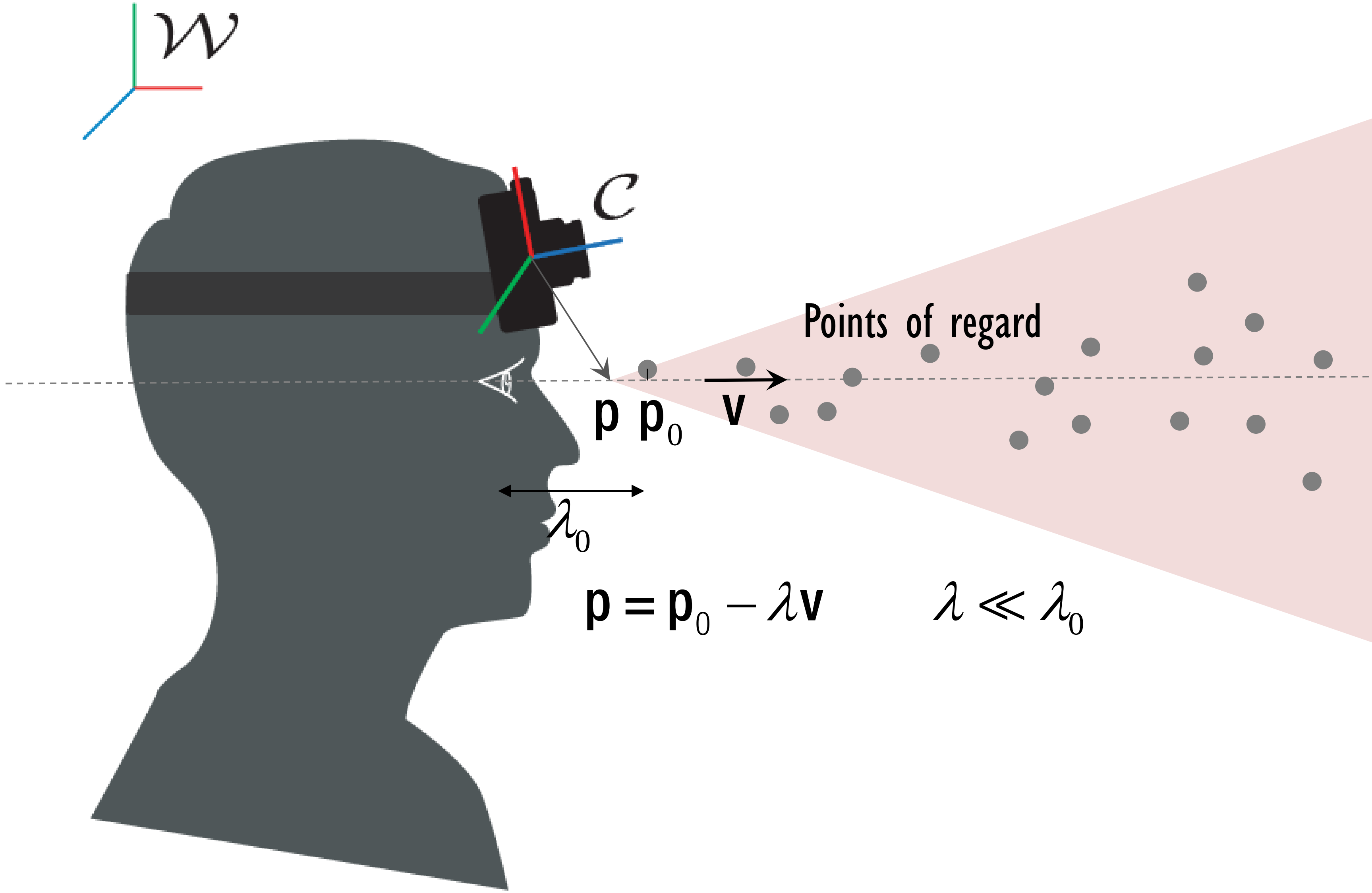
Gaze Calibration



$$\mathbf{p} = \mathbf{p}_0 - \lambda \mathbf{v} \qquad \lambda \approx \lambda_0$$

[Park NIPS12]

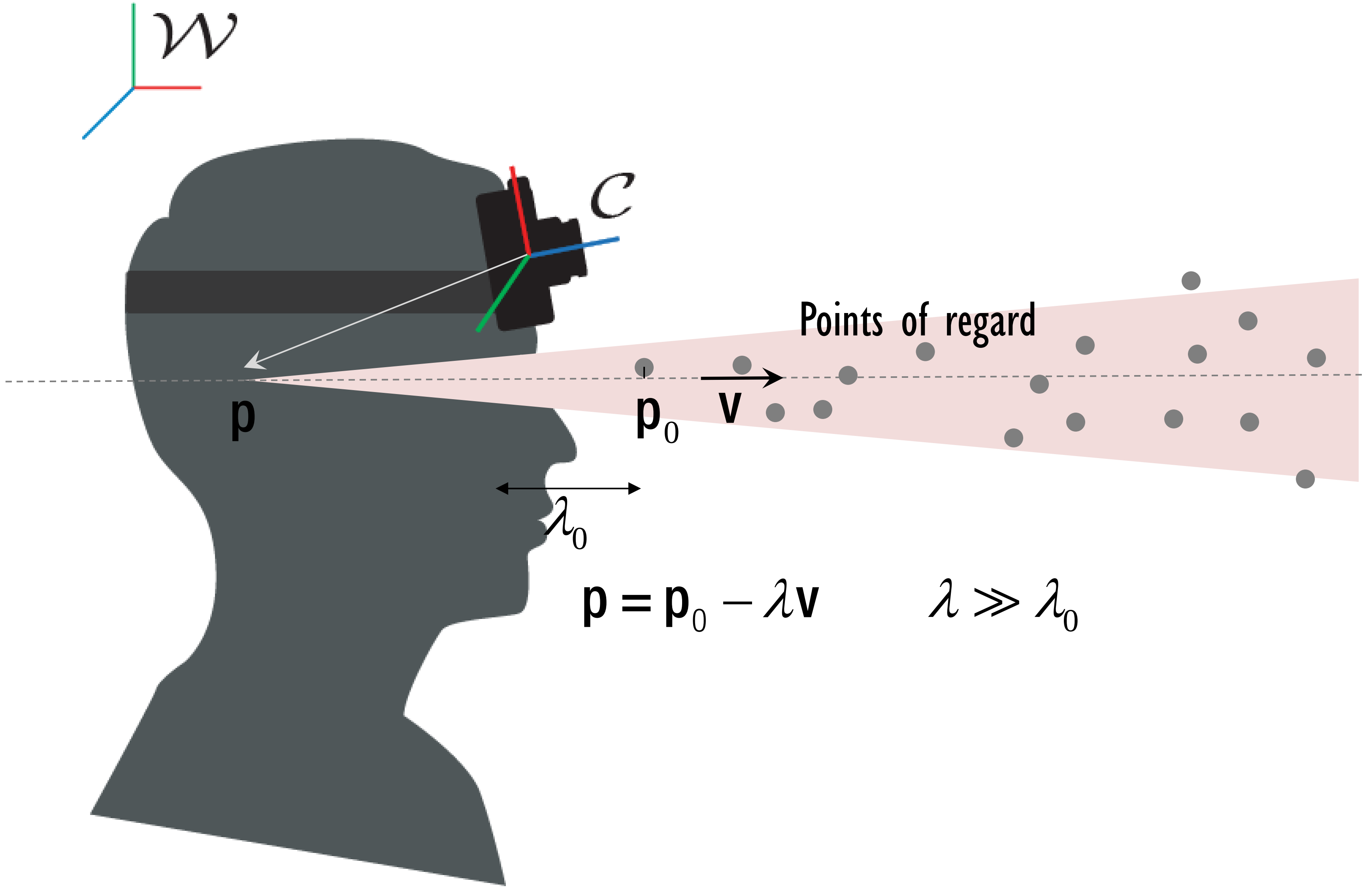
Gaze Calibration



$$\mathbf{p} = \mathbf{p}_0 - \lambda \mathbf{v} \qquad \lambda \ll \lambda_0$$

[Park NIPS12]

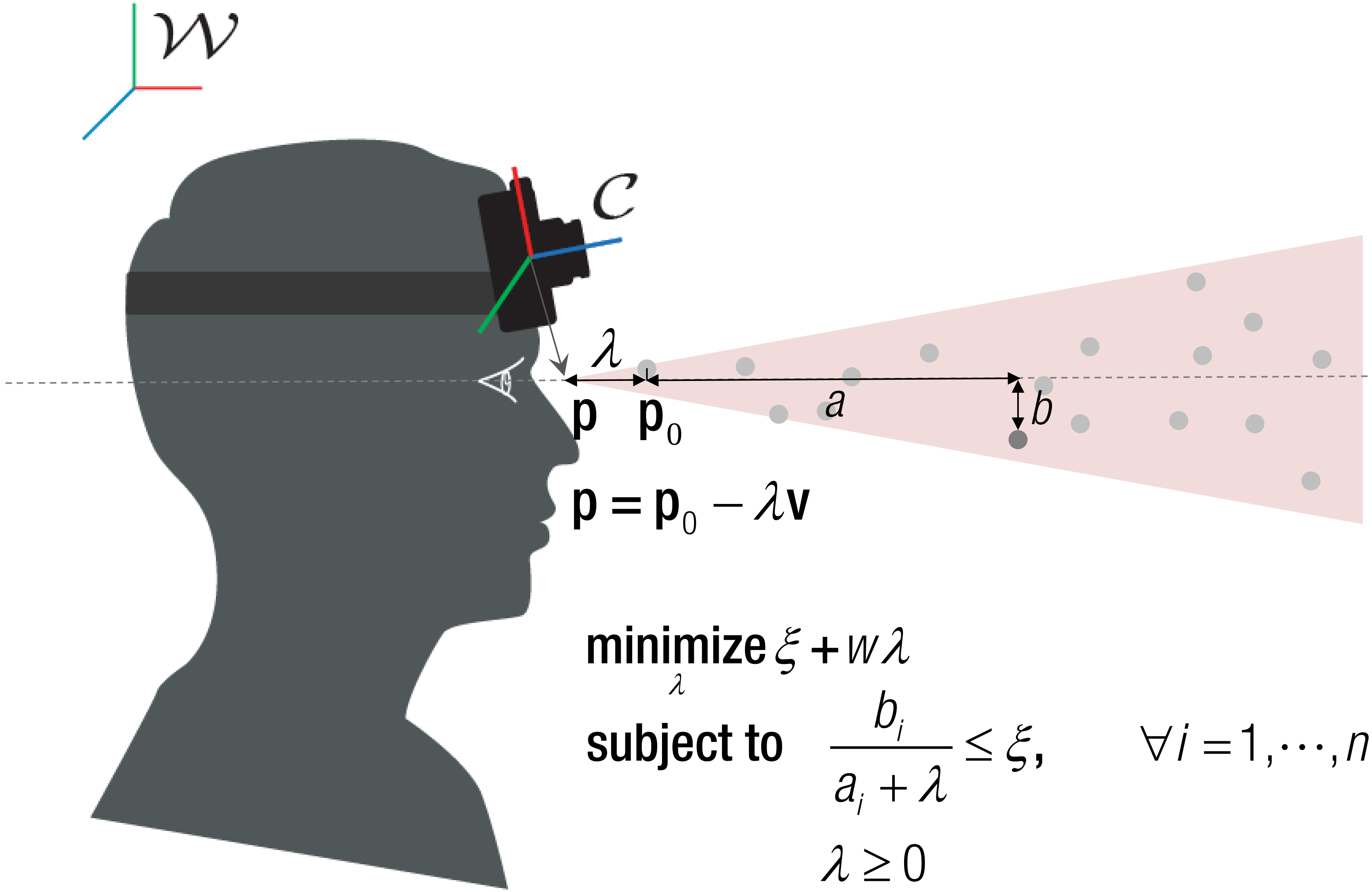
Gaze Calibration



$$\mathbf{p} = \mathbf{p}_0 - \lambda \mathbf{v} \qquad \lambda \gg \lambda_0$$

[Park NIPS12]

Gaze Calibration

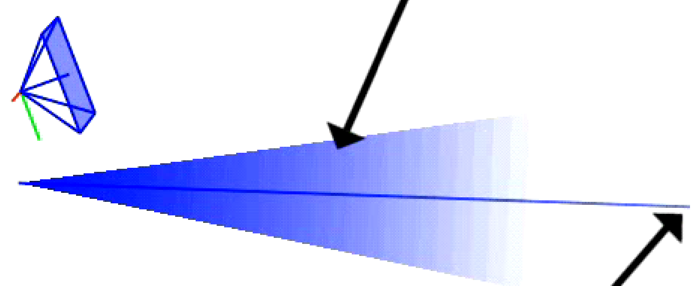


$$\mathbf{p} = \mathbf{p}_0 - \lambda \mathbf{v}$$

$$\begin{aligned} &\underset{\lambda}{\text{minimize}}\ \xi + w\lambda \\ &\text{subject to} \quad \frac{b_i}{a_i + \lambda} \le \xi, \qquad \forall i = 1, \cdots, n \\ &\lambda \ge 0 \end{aligned}$$

[Park NIPS12]

Gaze Ray Calibration



Primary gaze ray with respect to the camera pose

[Park NIPS12]

Gaze Ray Calibration

Back and forth motion　　Side to side motion

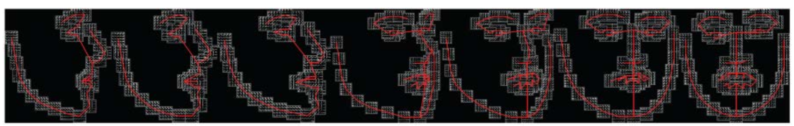
[Park NIPS12]

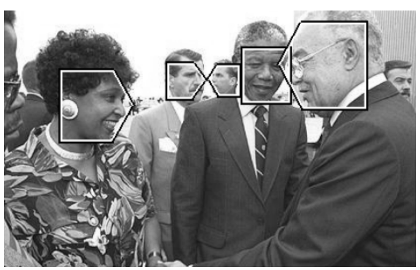
Head Detection/Alignment

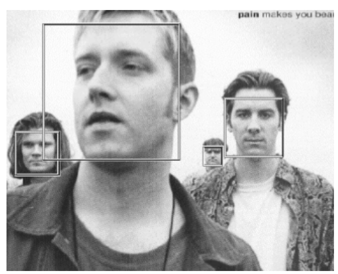
[Cootes, PAMI01, Schneiderman, CVPR00, Viola, IJCV01, Matthews, IJCV04, Saragih, ICCV09, Xiong, CVPR13, Zhu, CVPR12, Marin-Jimenez, IJCV14, …]

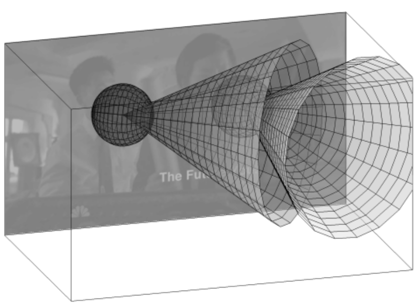
Input: image or video of third person view

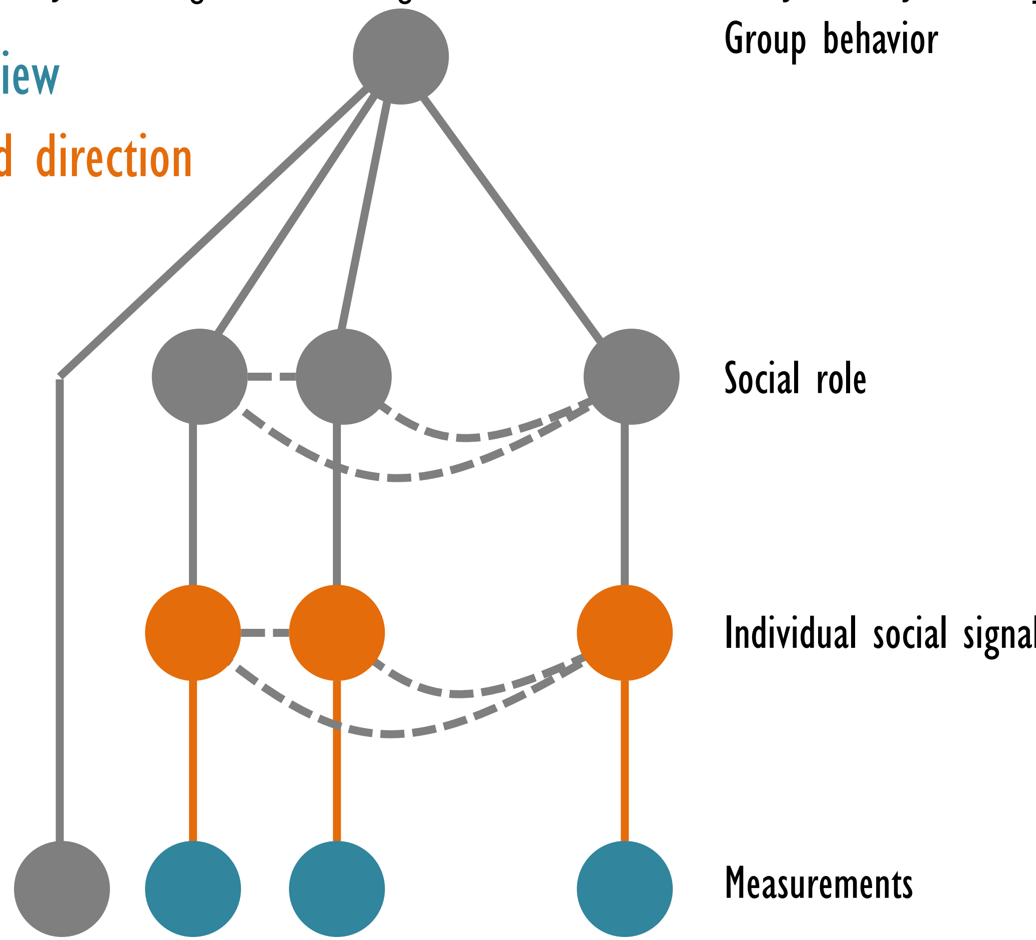
Output: to find head / to estimate head direction











[Schneiderman CVPR00, Xiong CVPR13]


[Schneiderman CVPR00, Xiong CVPR13]

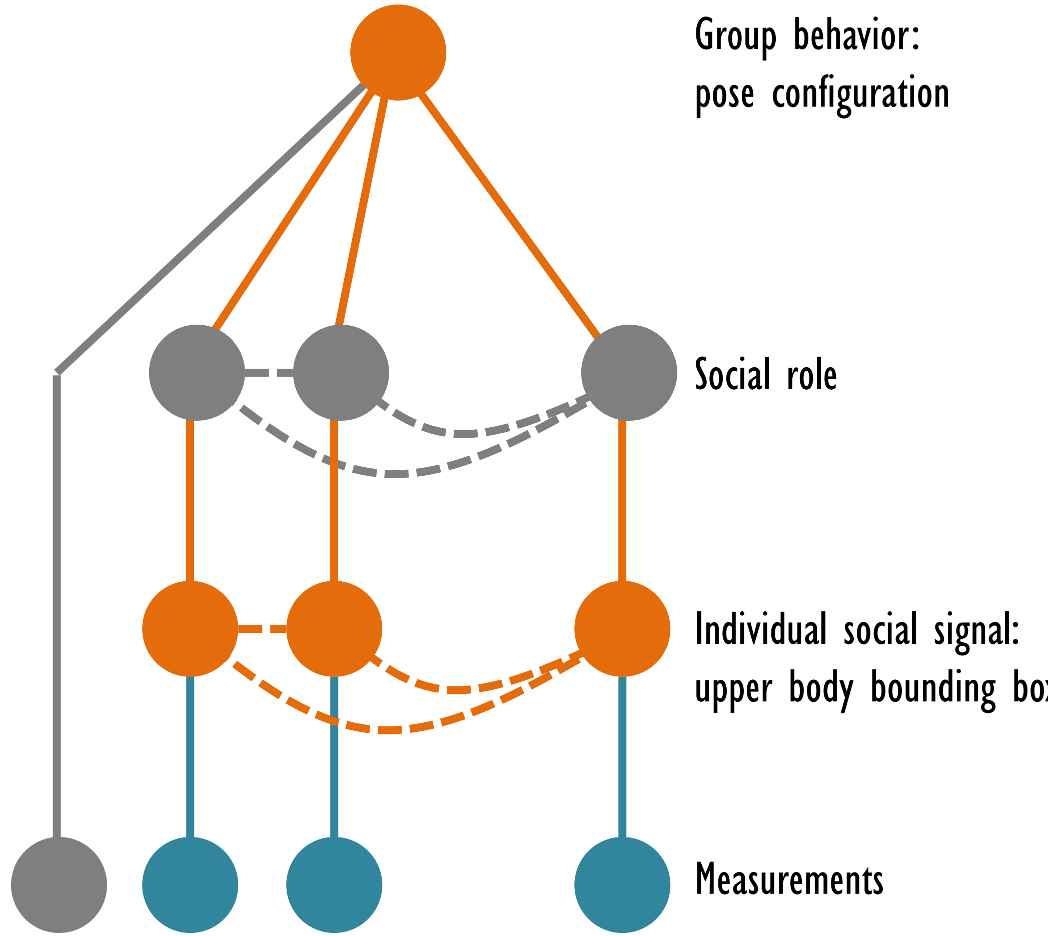
[Schneiderman CVPR00, Xiong CVPR13]

Body Configuration

[Hoai CVPR14]

Input: images of TV shows

Output: to detect social interactions



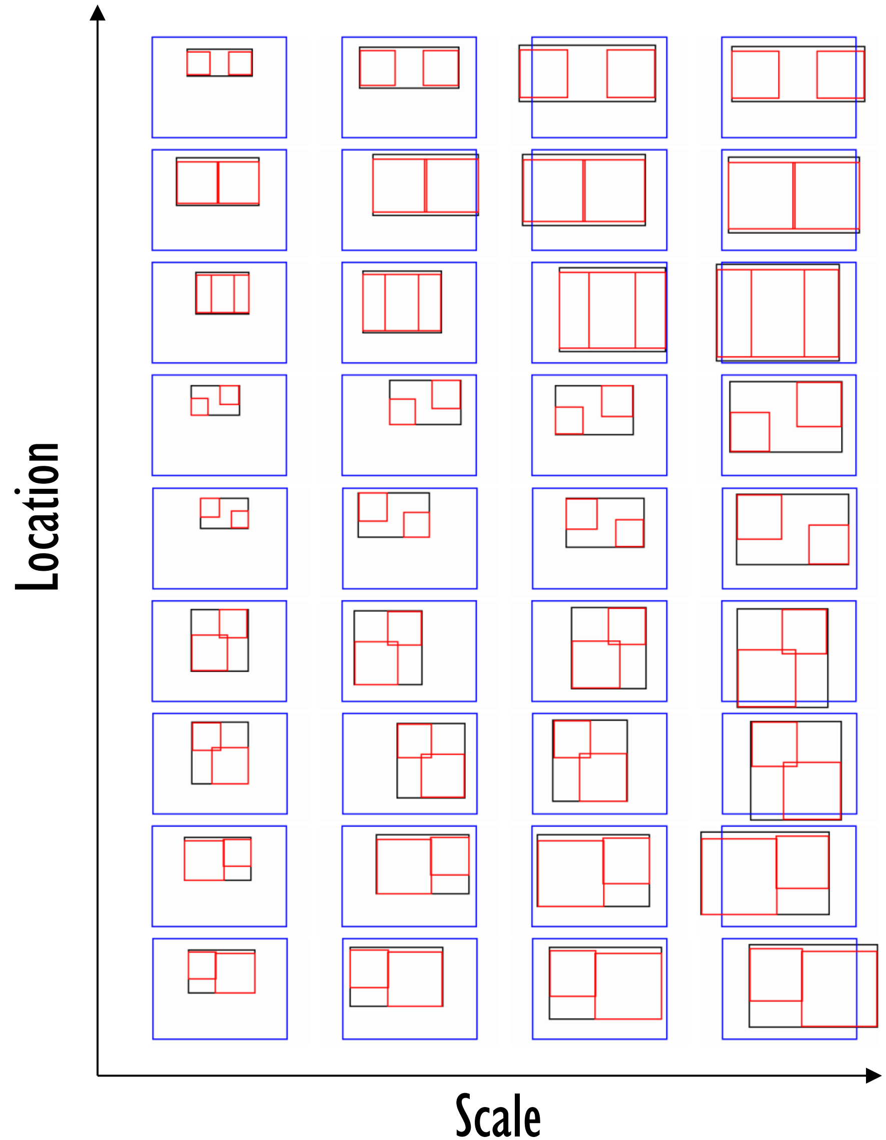
Characteristics of TV show interactions:

- Important contents mostly stay within frame.
- Particular camera angle is preferred.

[Hoai CVPR14]

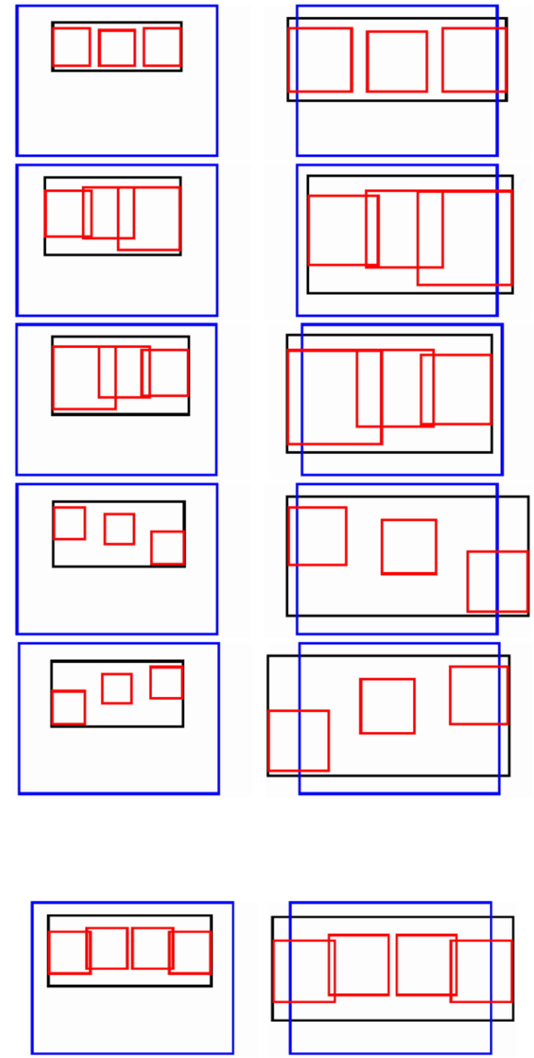
Common Configuration of Upper Body

Two People

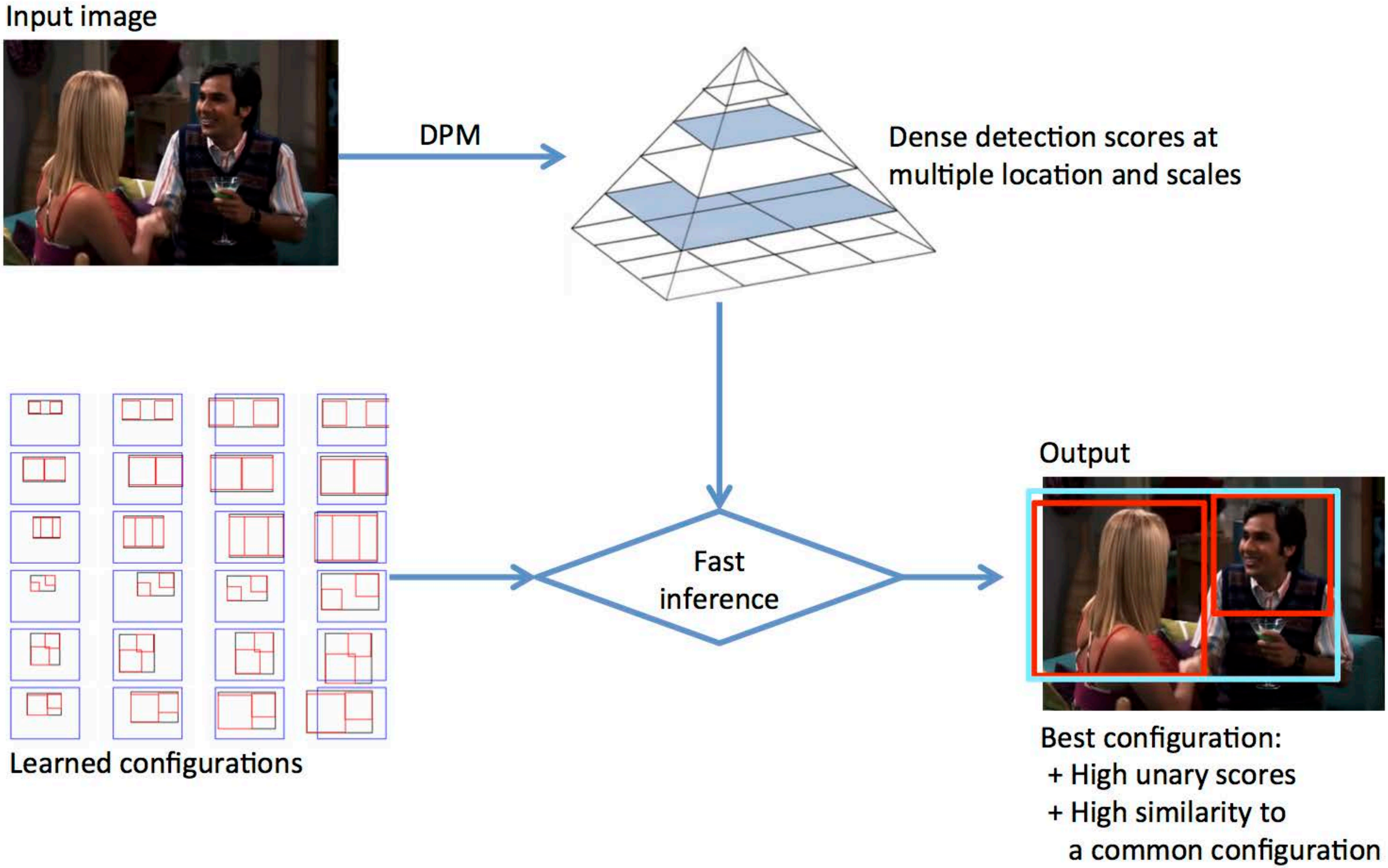


Common Configuration of Upper Body

More than Three People

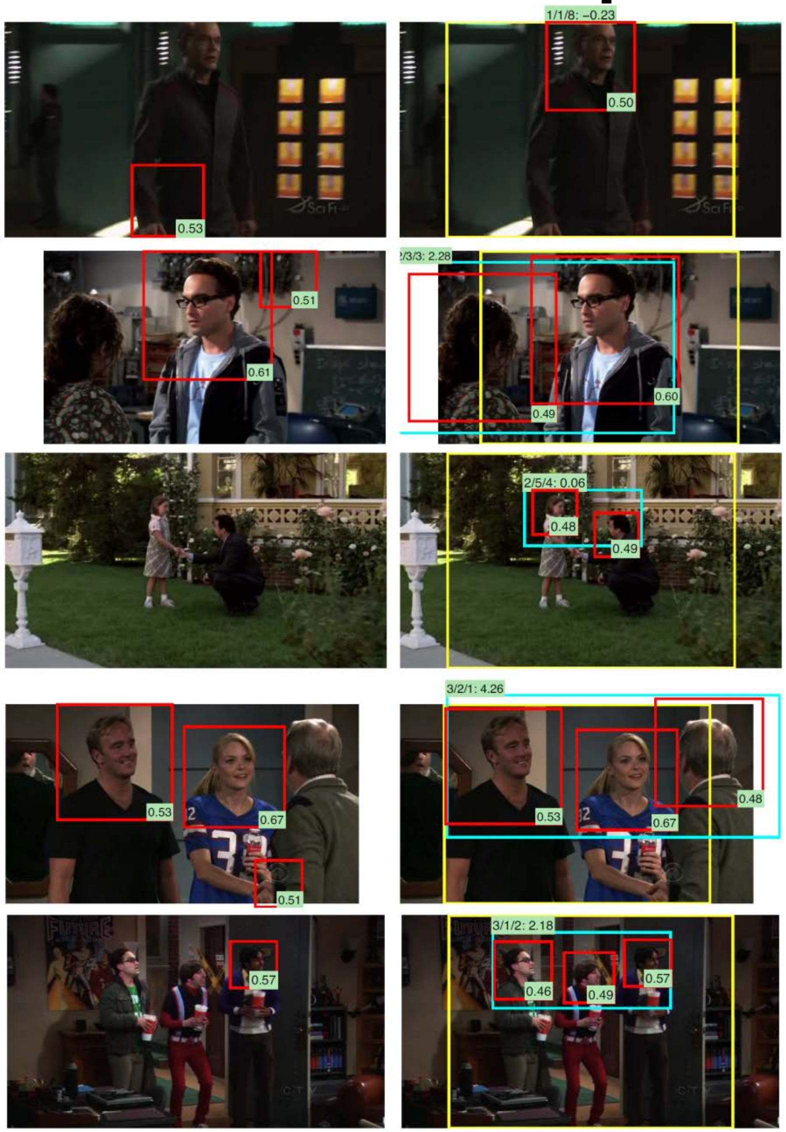


[Hoai CVPR14]

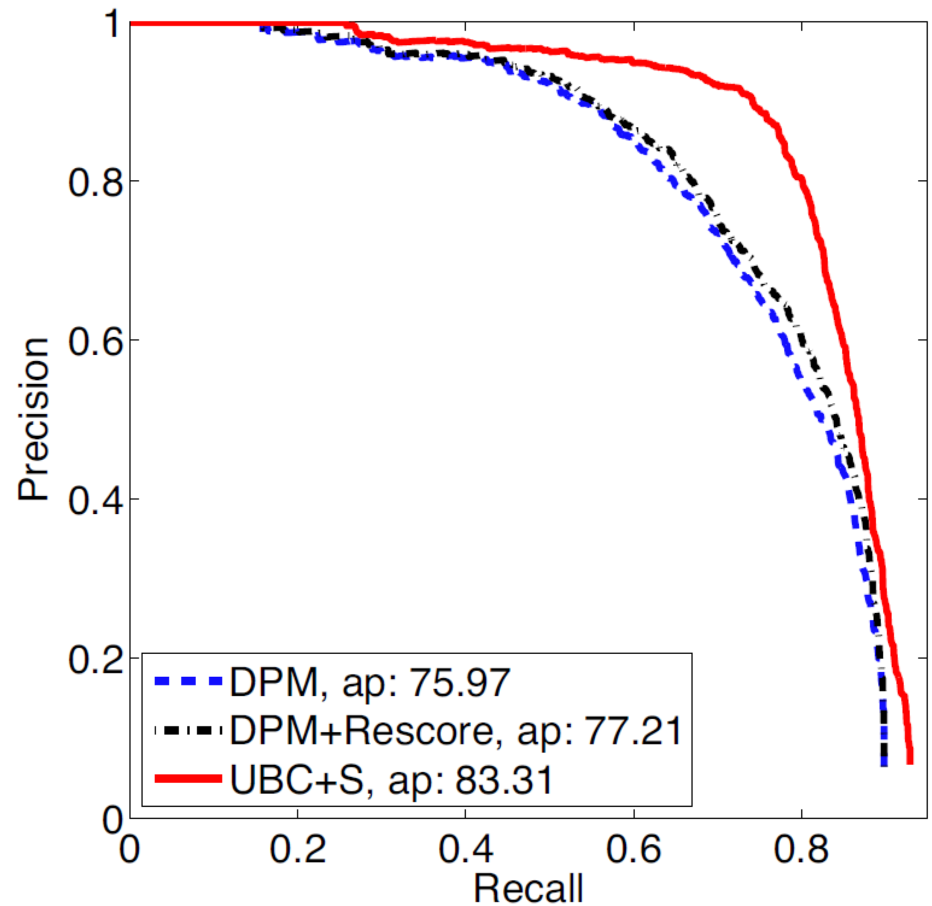

Input image
DPM
Dense detection scores at
multiple location and scales
Fast
inference
Output
Learned configurations
Best configuration:
+ High unary scores
+ High similarity to
a common configuration

[Hoai CVPR14]

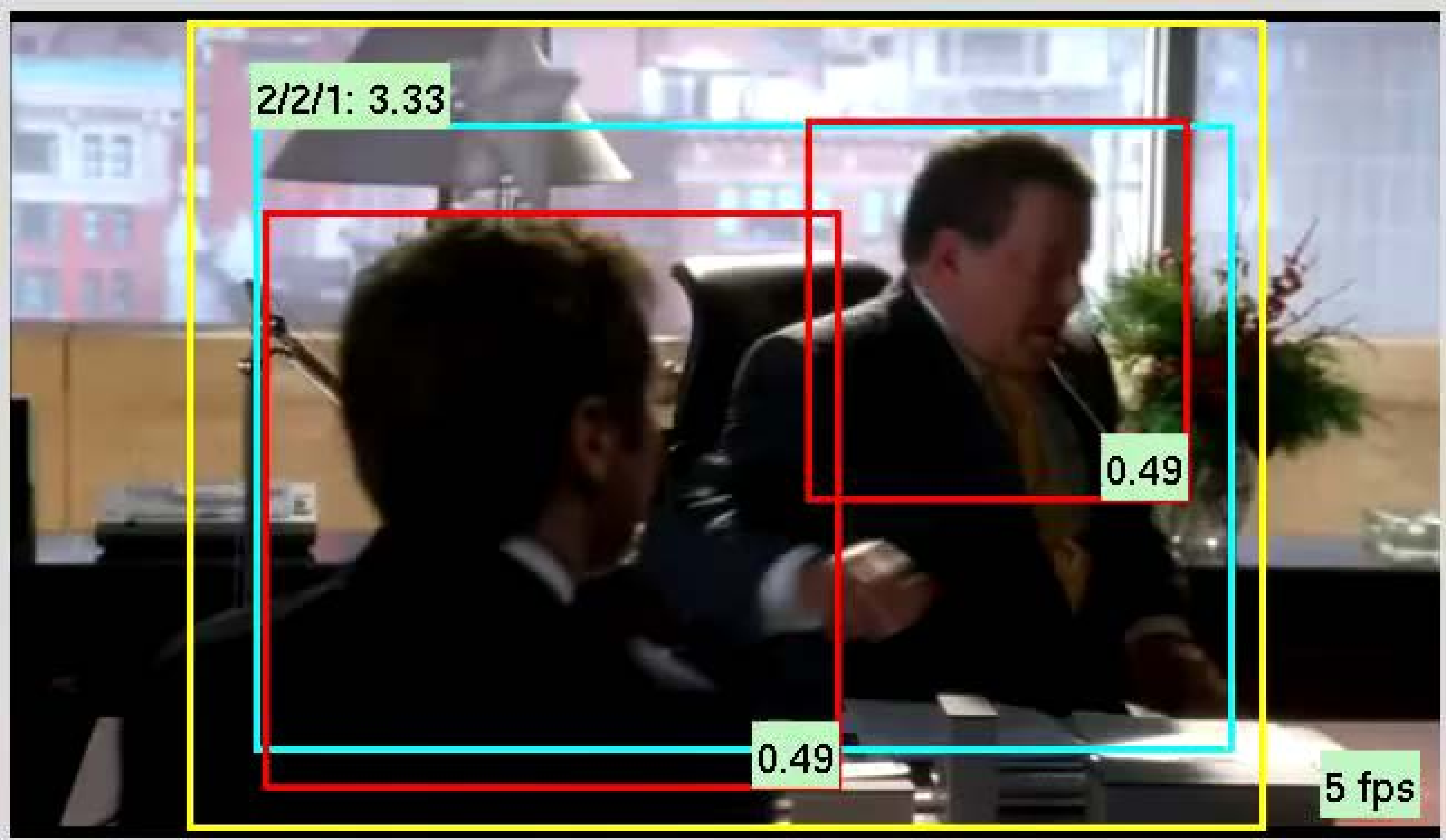
Comparison w/ DPM



DPM Hoai, CVPR14



[Hoai CVPR14]



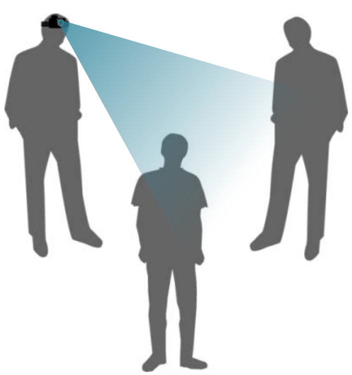
[Hoai CVPR14]

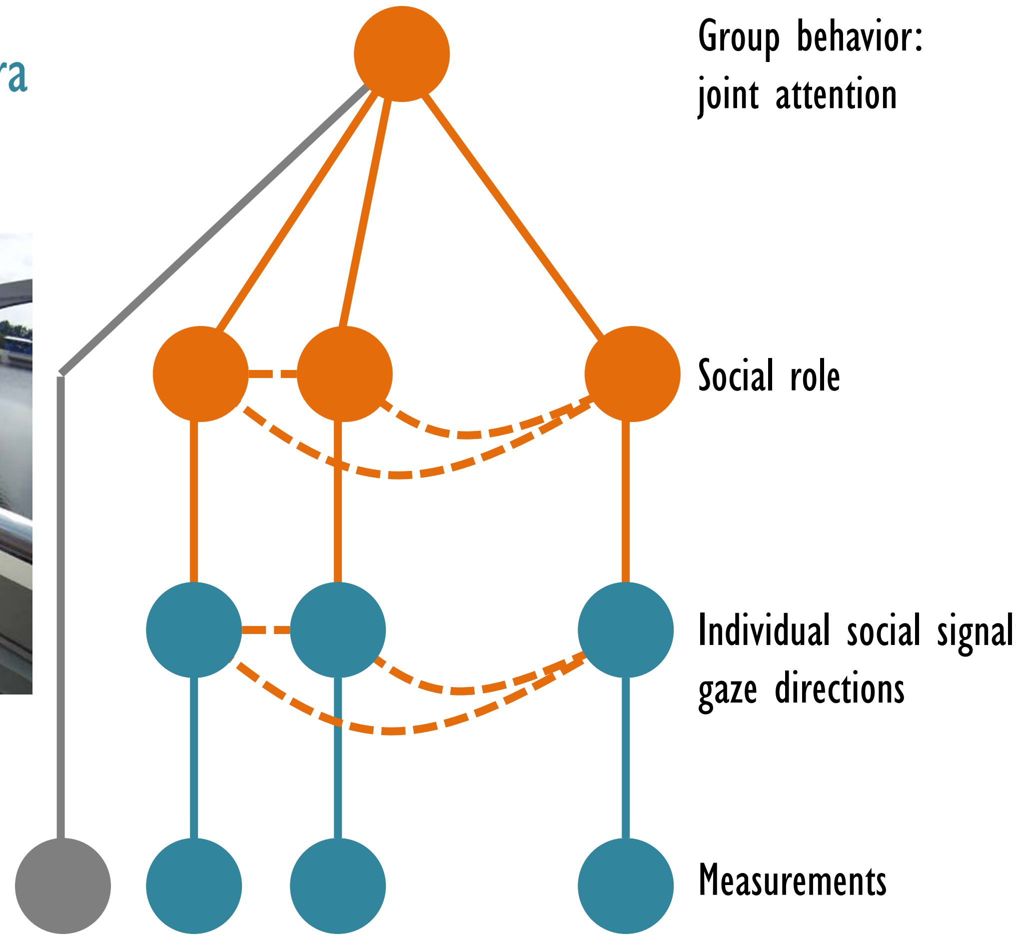
Joint Attention

[Fathi CVPR12]

Input: images from a first person camera

Output: to detect social interactions



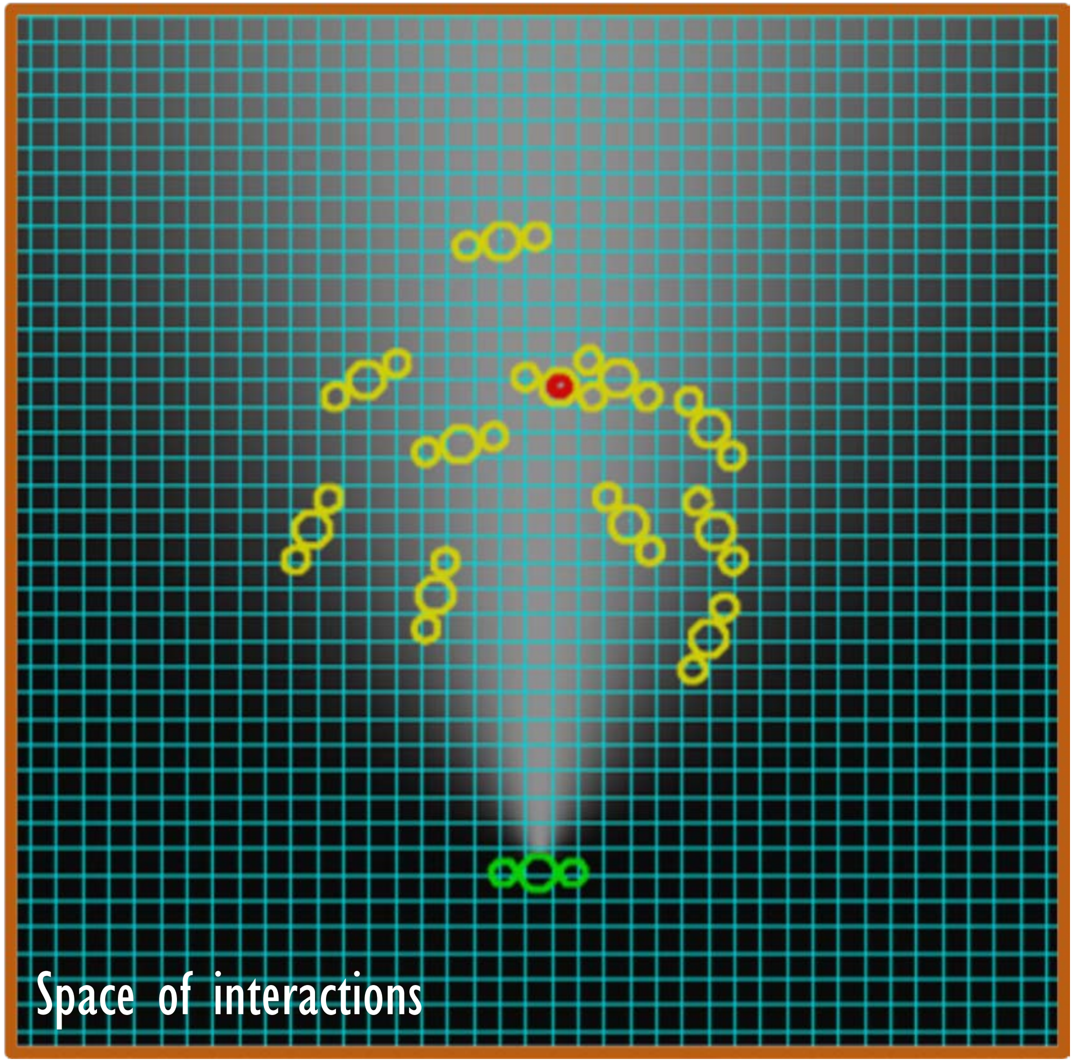


Where Do They Look?

[Fathi CVPR12]

MRF Modeling

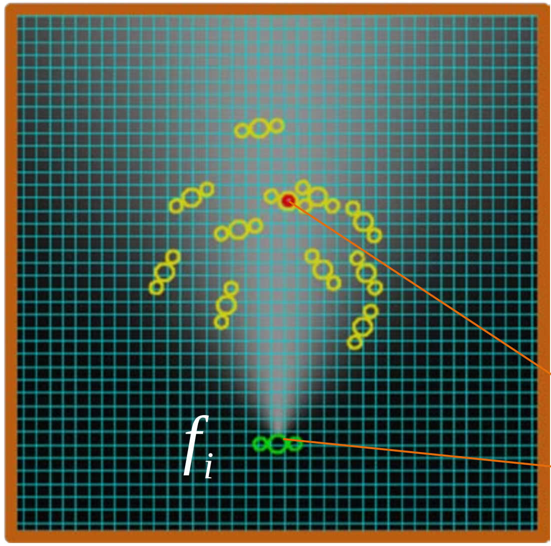
[Fathi CVPR12]



Top view

MRF Modeling

[Fathi CVPR12]



MRF modeling: unary potential

$$\phi_U(L_{f_i}, P_{f_1}, P_{f_2}, \ldots, P_{f_N}) = \phi_1(L_{f_i}, P_{f_i}) \times \phi_2(L_{f_i}, P_{f_i}) \times \phi_3(L_{f_i}, P_{f_1}, \ldots, P_{f_N})$$

L_{f_i} : location at which f_i is looking (label space)

P_{f_i} : position and orientation of f_i

- Head direction is aligned with the point of regard.

$$\phi_1(L_{f_i} = \ell, P_{f_i}) = \frac{1}{\sigma_1\sqrt{2\pi}} \exp\left\{-\frac{\|V_{f_i} - \overline{(\ell - T_{f_i})}\|^2}{2\sigma_1^2}\right\}$$

- The point of regard cannot be the himself.

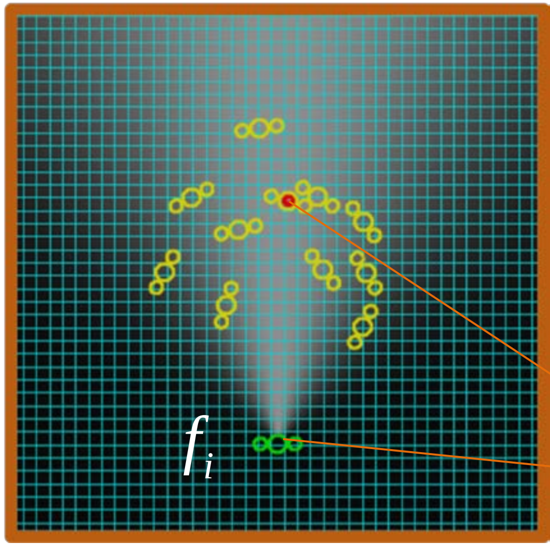
$$\phi_2(L_{f_i} = \ell, P_{f_i}) = \frac{1}{1 + \exp\{-(c_2.\|\ell - P_{f_i}\|)\}}$$

- The point of regard is likely to be a face.

$$\phi_3(L_{f_i} = \ell, P_{f_1}, \ldots, P_{f_N}) = \begin{cases} c_3 & \ell = P_{f_j} \forall j \neq i \\ 1 & otherwise \end{cases}$$

MRF Modeling

[Fathi CVPR12]



MRF modeling: unary potential

$$\phi_U(L_{f_i}, P_{f_1}, P_{f_2}, ..., P_{f_N}) = \phi_1(L_{f_i}, P_{f_i}) \times \phi_2(L_{f_i}, P_{f_i}) \times \phi_3(L_{f_i}, P_{f_1}, ..., P_{f_N})$$

L_{f_i} : location at which f_i is looking (label space)

P_{f_i} : position and orientation of f_i

MRF modeling: binary potential

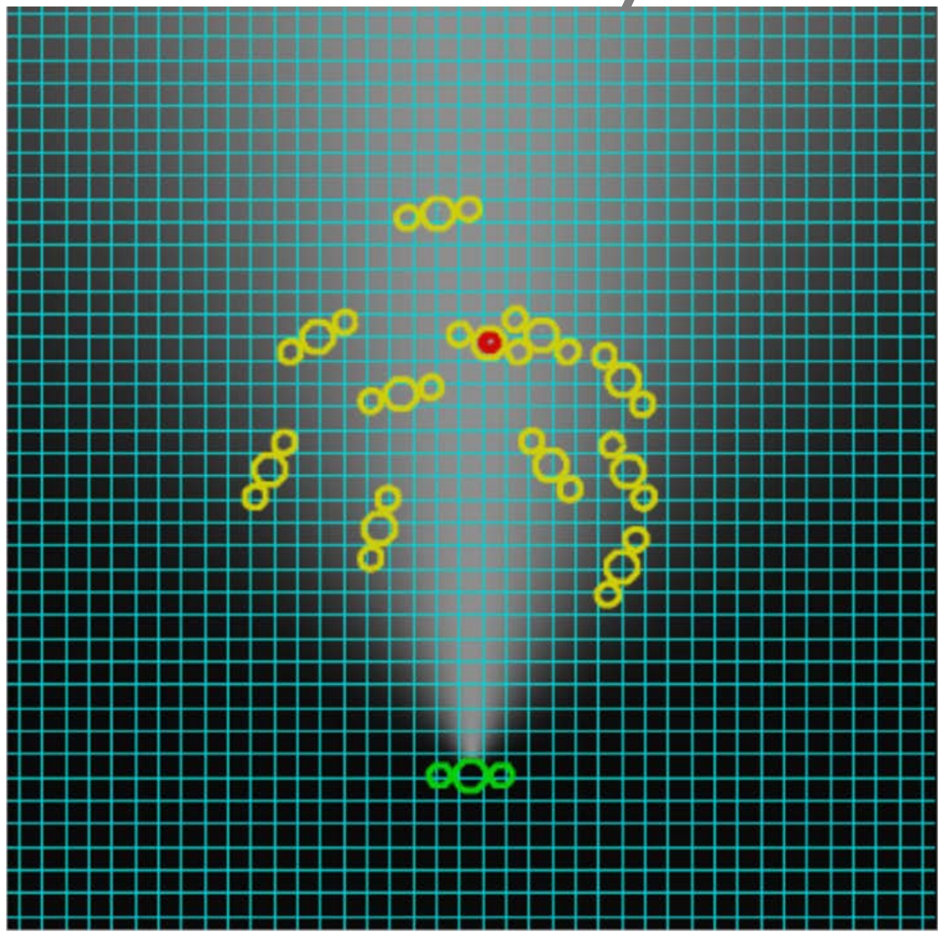
$$\phi_B(L_{f_i} = \ell_1, L_{f_j} = \ell_2) = \begin{cases} c_B & if(\ell_1 = \ell_2) \\ 1 - c_B & if(\ell_1 \neq \ell_2) \end{cases}$$

People engage joint attention.

MRF Inference

Where Do They Look?

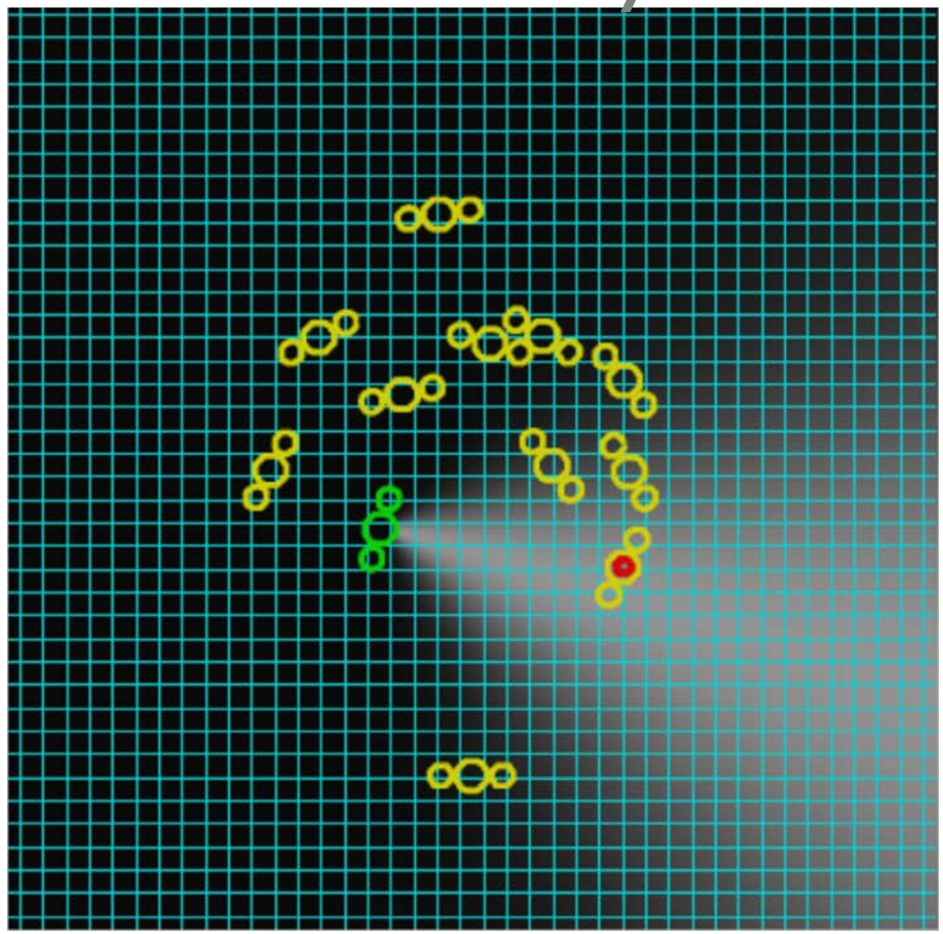
[Fathi CVPR12]



MRF Inference

Where Do They Look?

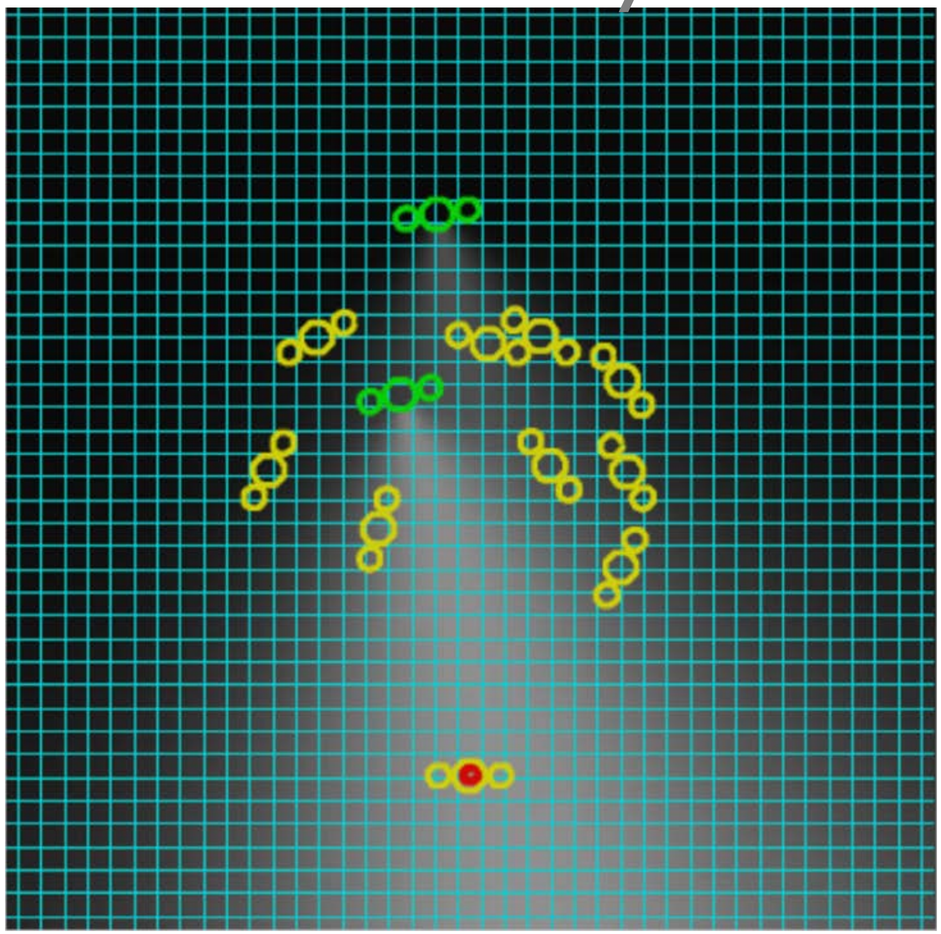
[Fathi CVPR12]



MRF Inference

Where Do They Look?

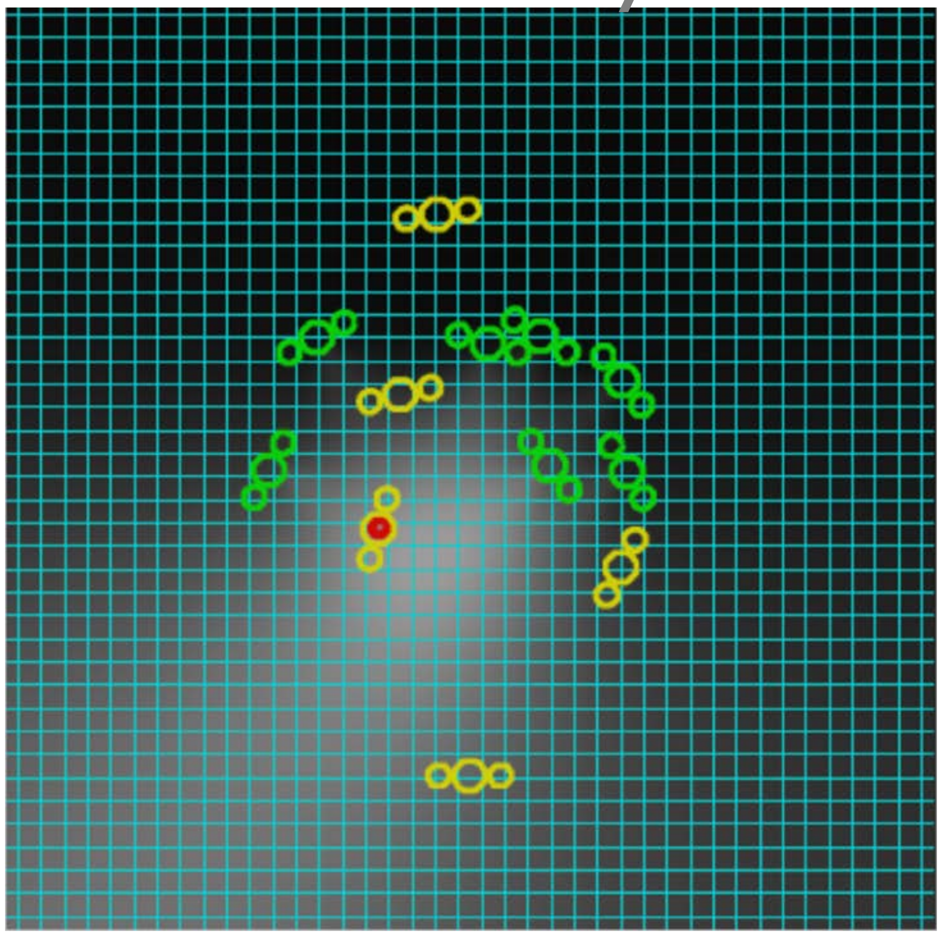
[Fathi CVPR12]



MRF Inference

Where Do They Look?

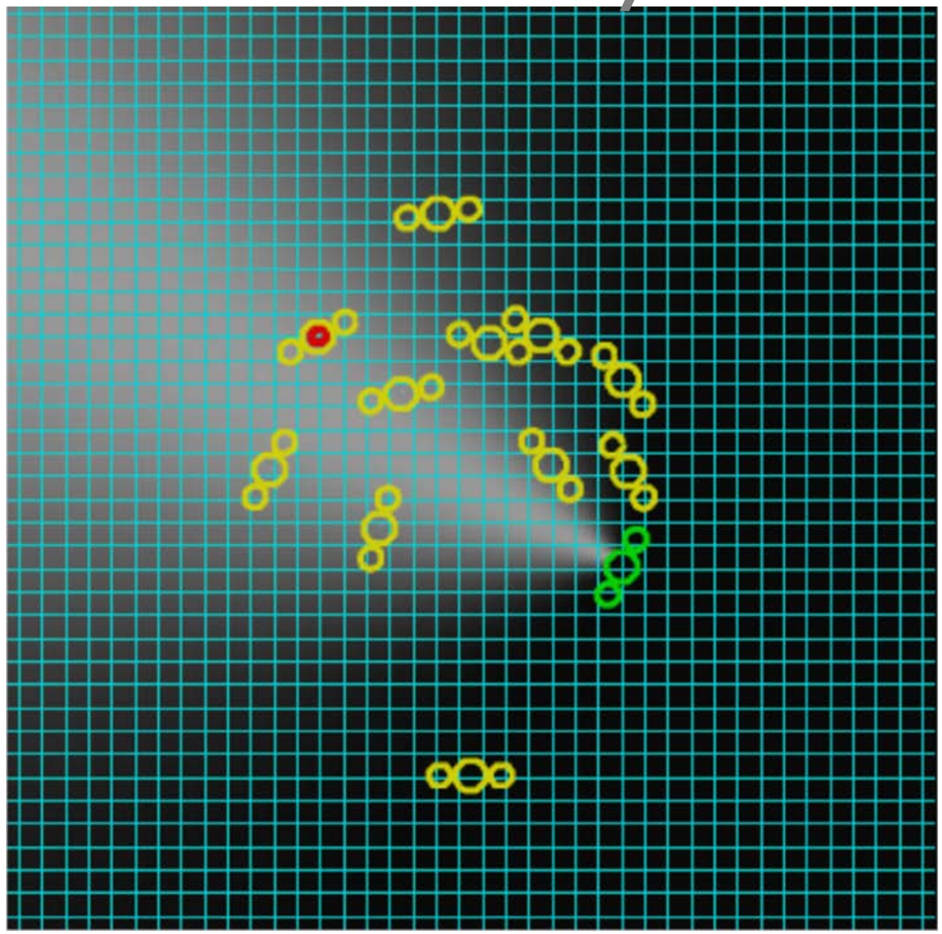
[Fathi CVPR12]



MRF Inference

Where Do They Look?

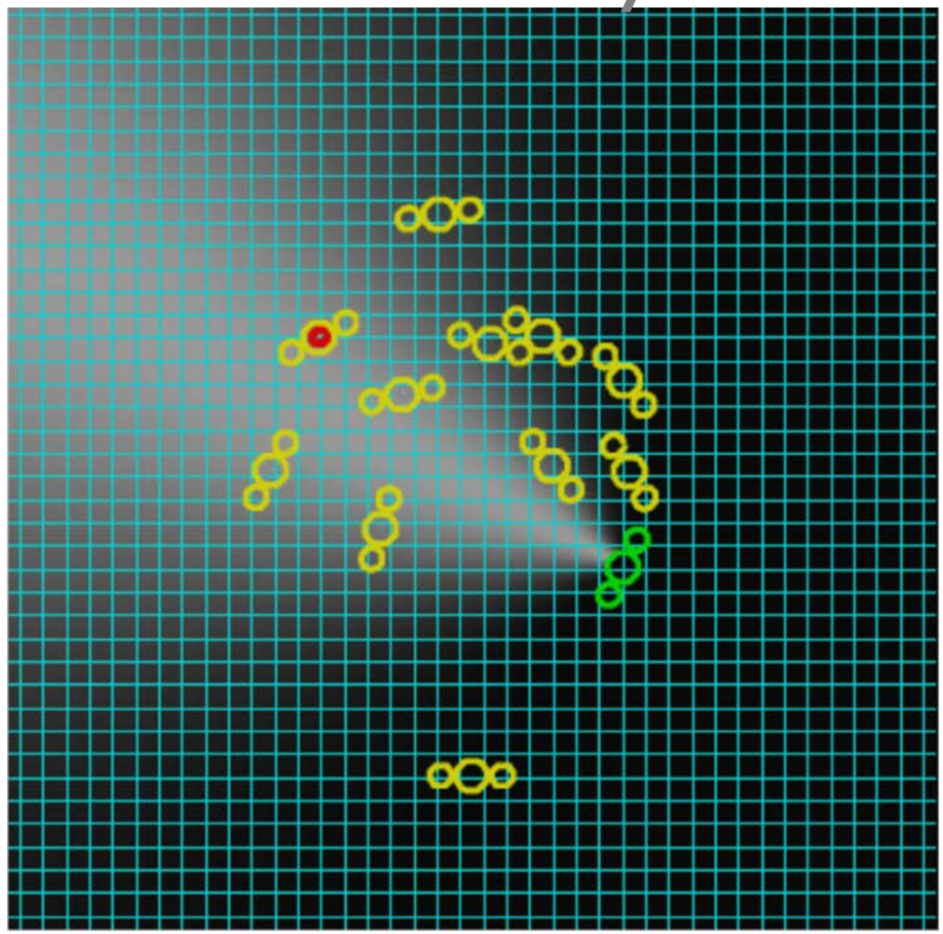
[Fathi CVPR12]



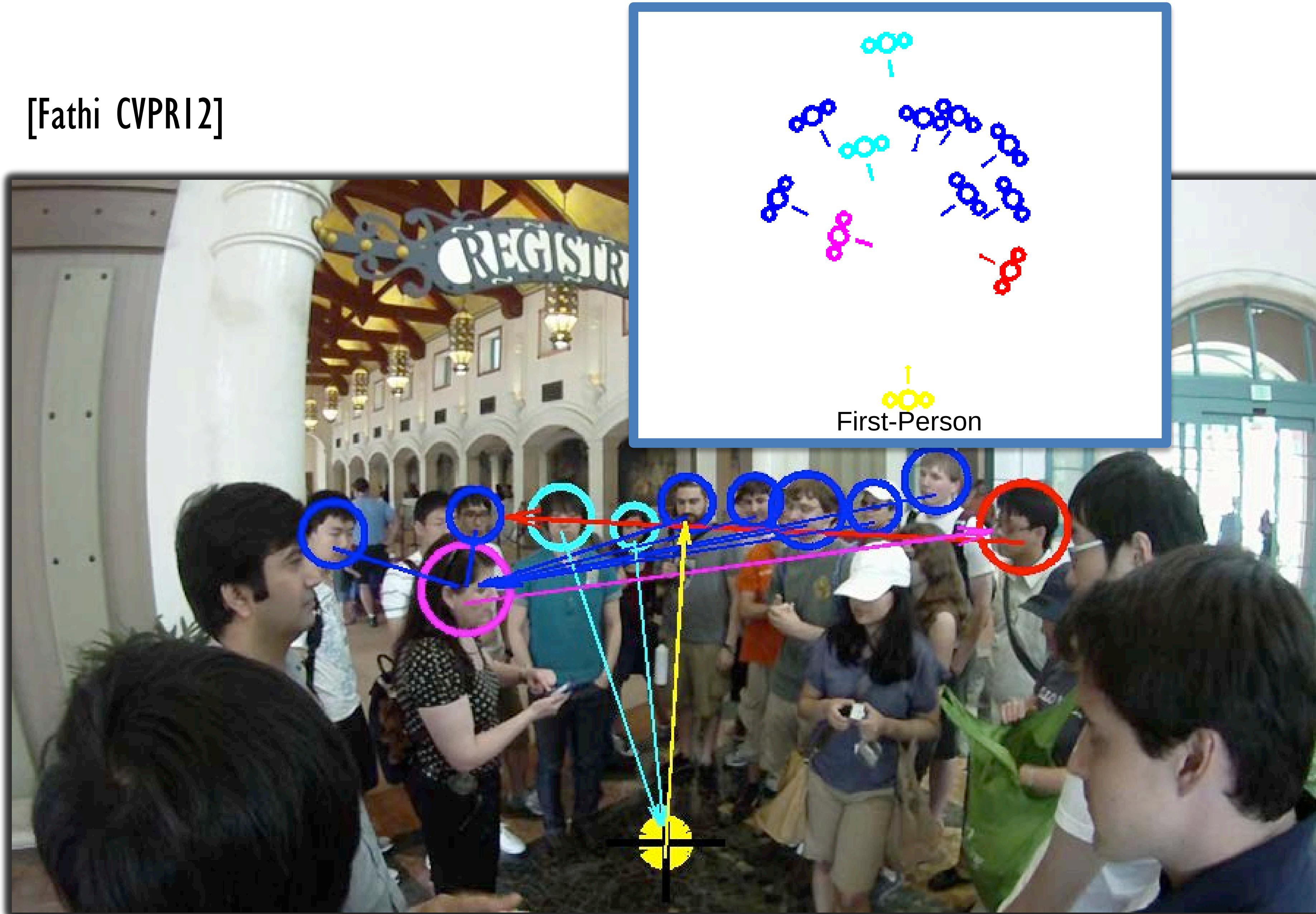
MRF Inference

Where Do They Look?

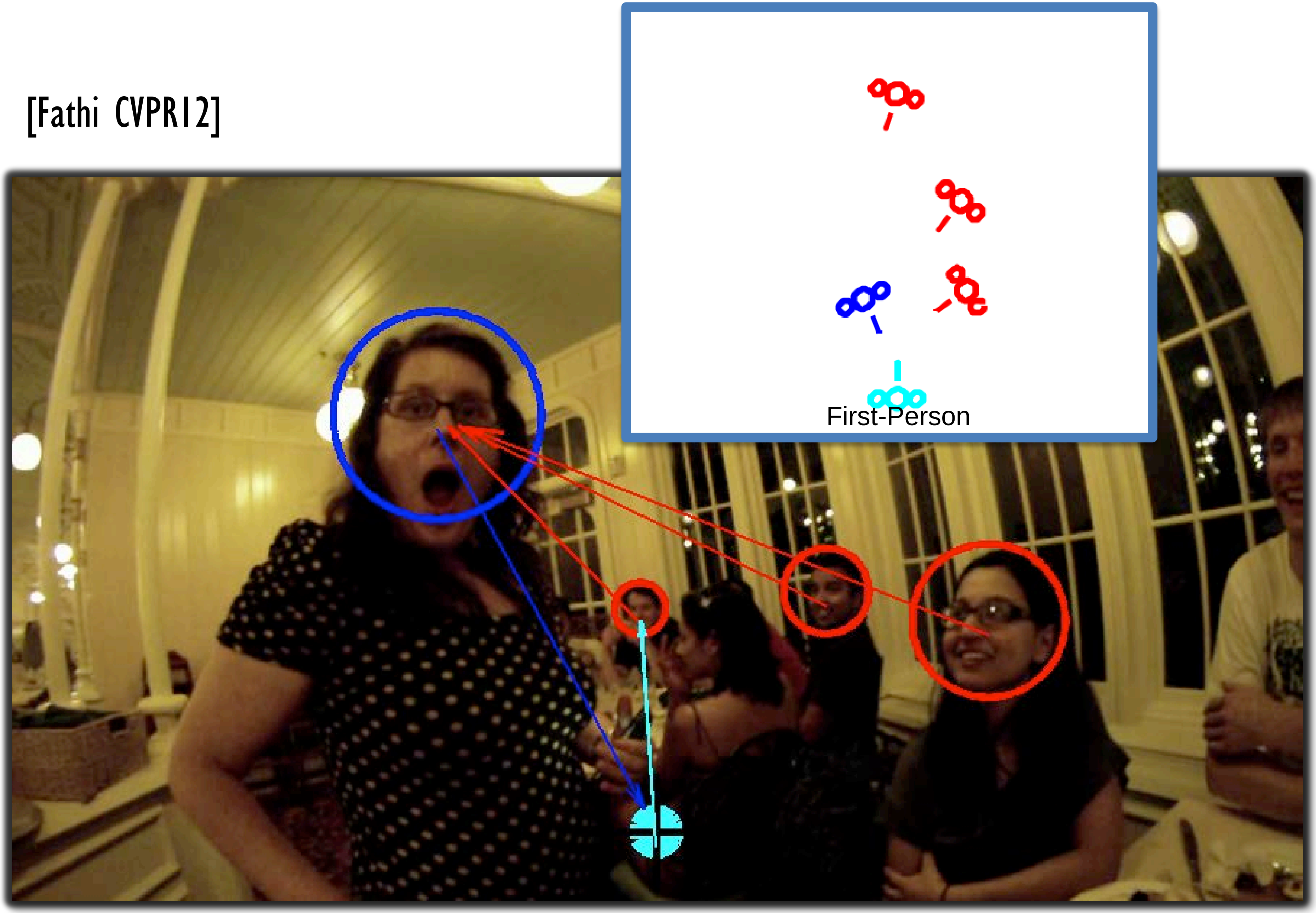
[Fathi CVPR12]



[Fathi CVPR12]



[Fathi CVPR12]



Detection of Social Interaction

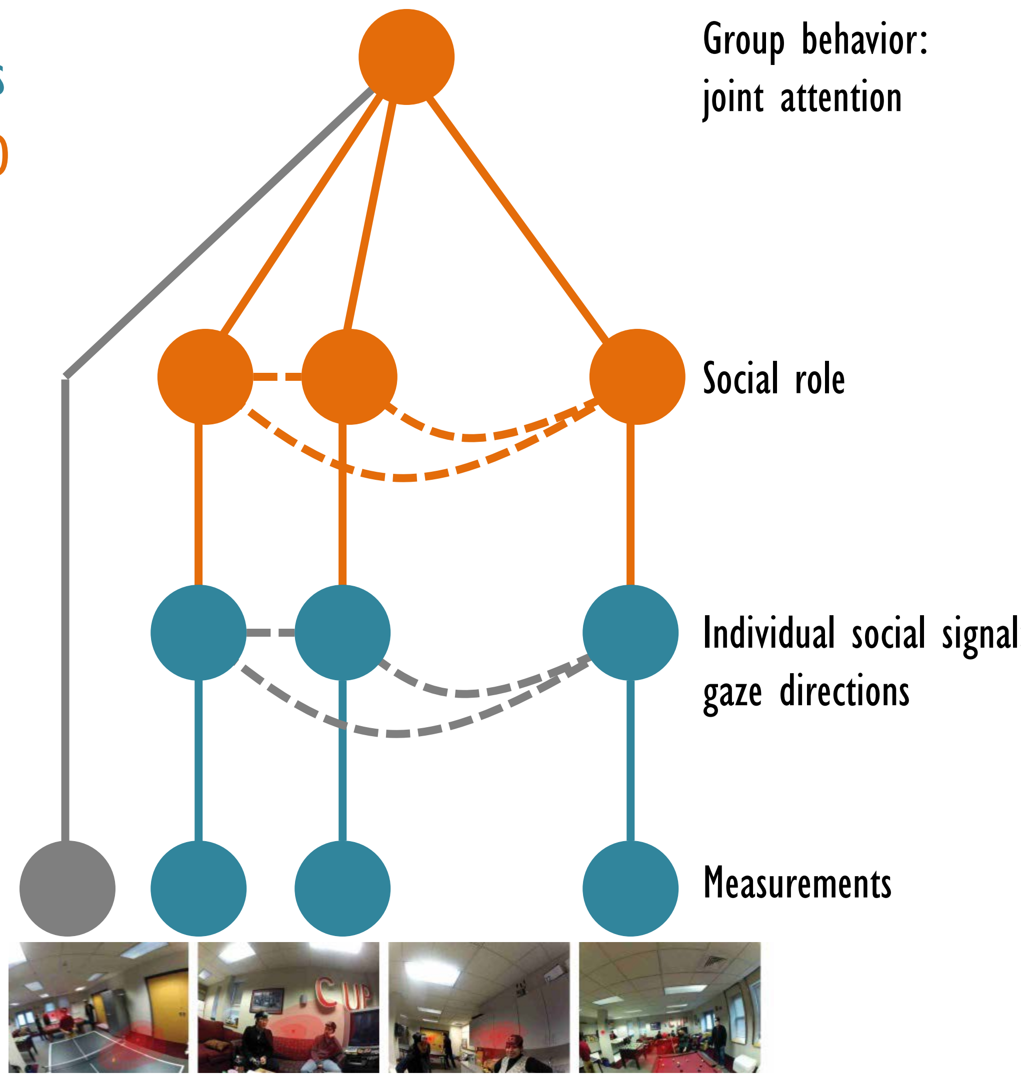
[Fathi CVPR12]

Joint Attention

[Park NIPS12]

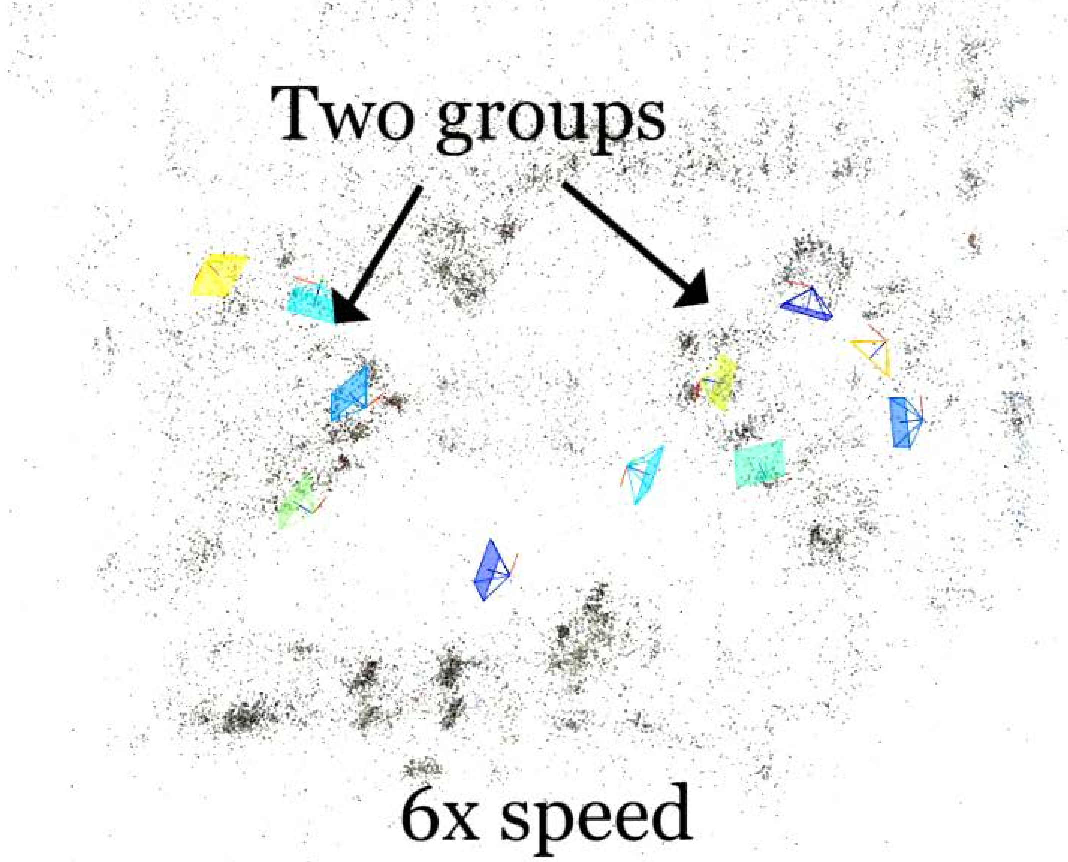
Input: images from first person cameras

Output: to localize joint attention in 3D



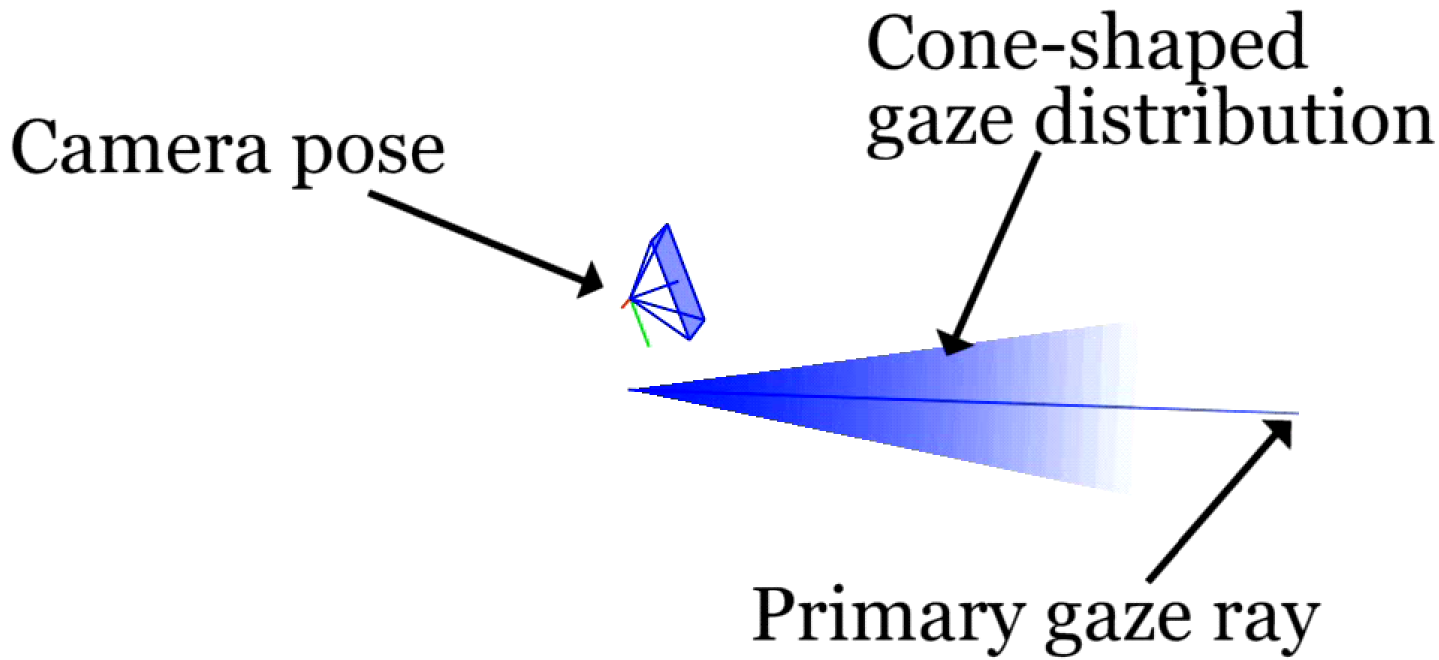
[Park NIPS12]

3D Camera Pose Estimation
(Structure from motion)



[Park NIPS12]

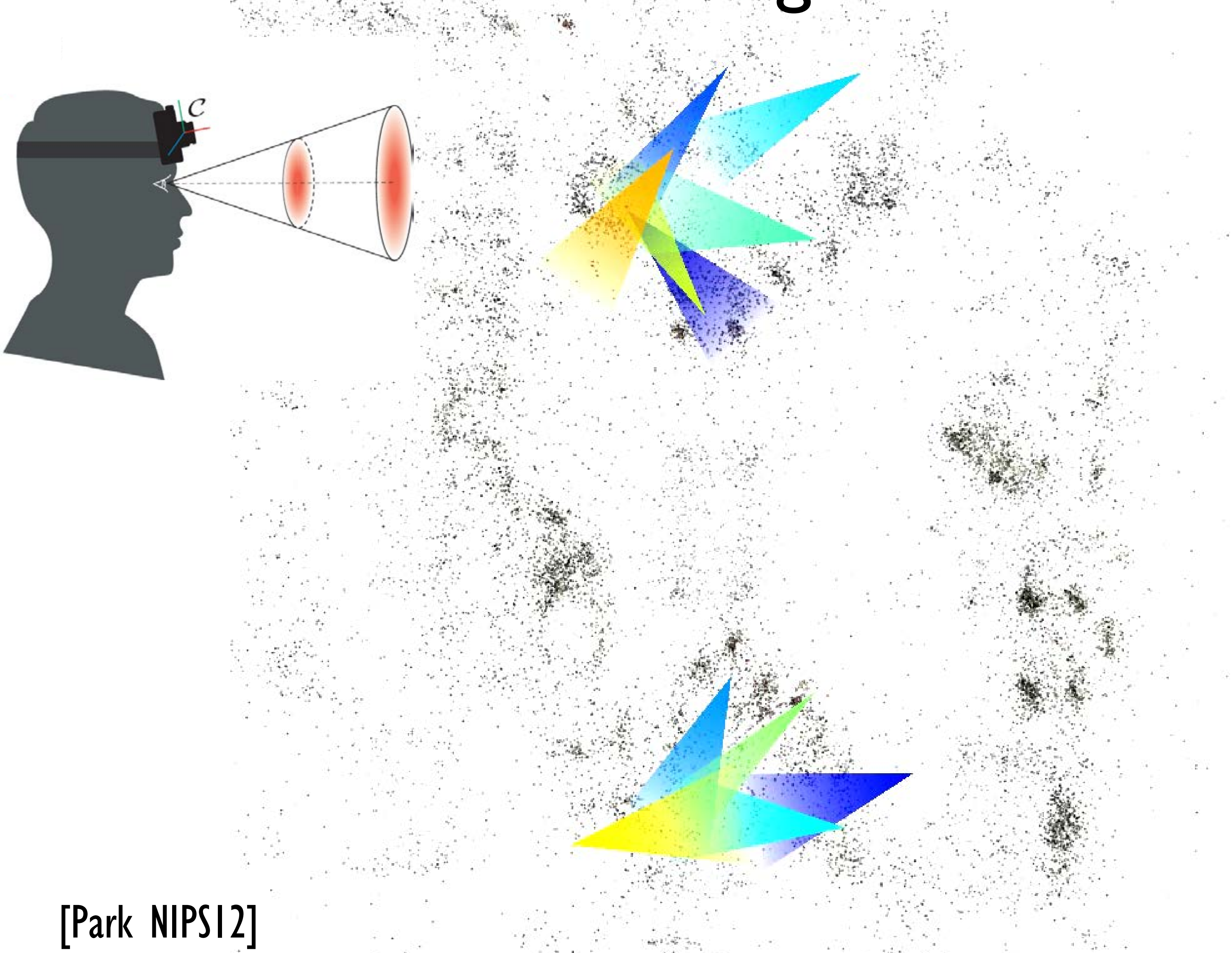
Gaze Ray Calibration



Primary gaze ray with respect to the camera pose

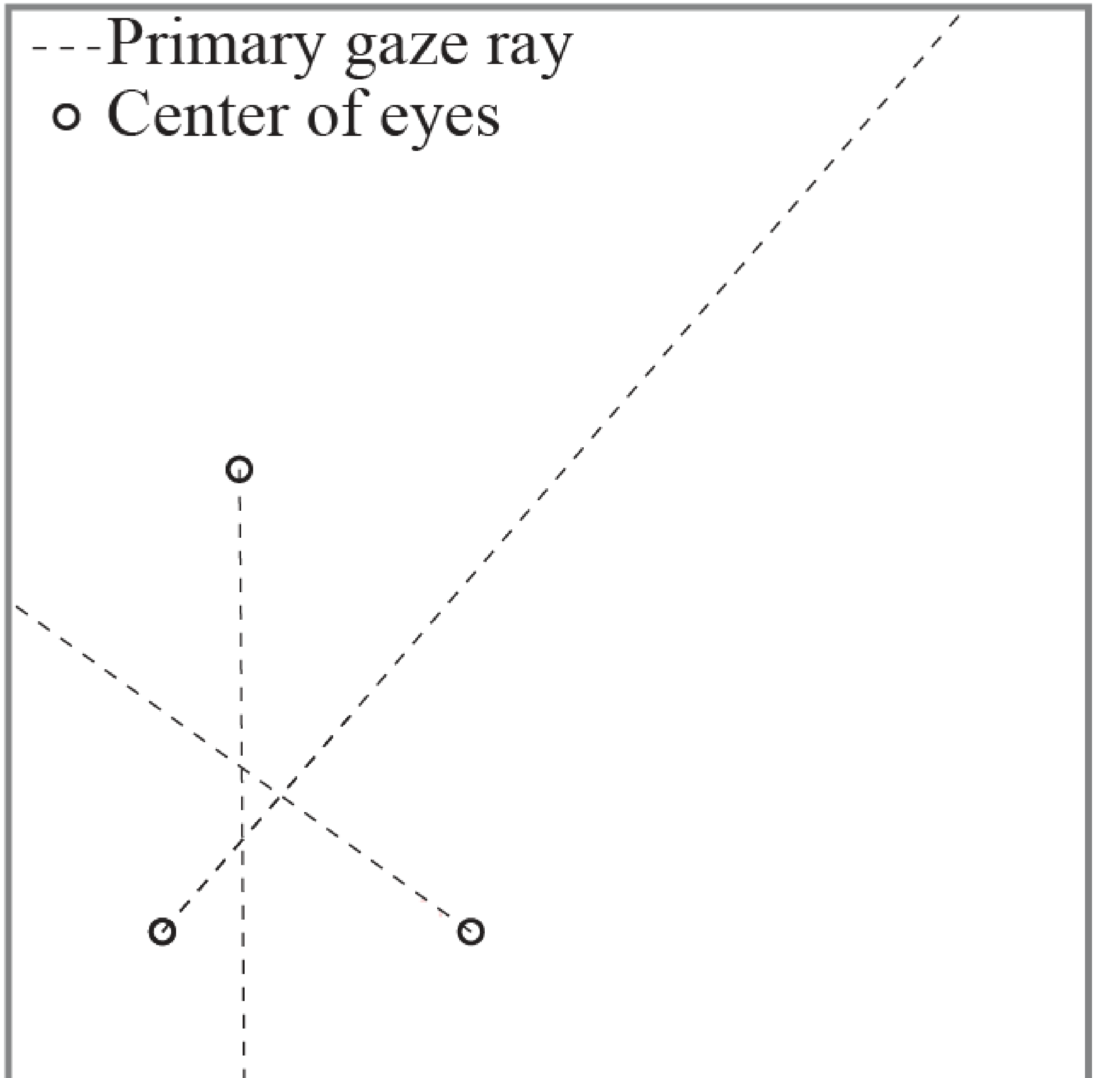
[Park NIPS12]

3D Gaze Registration



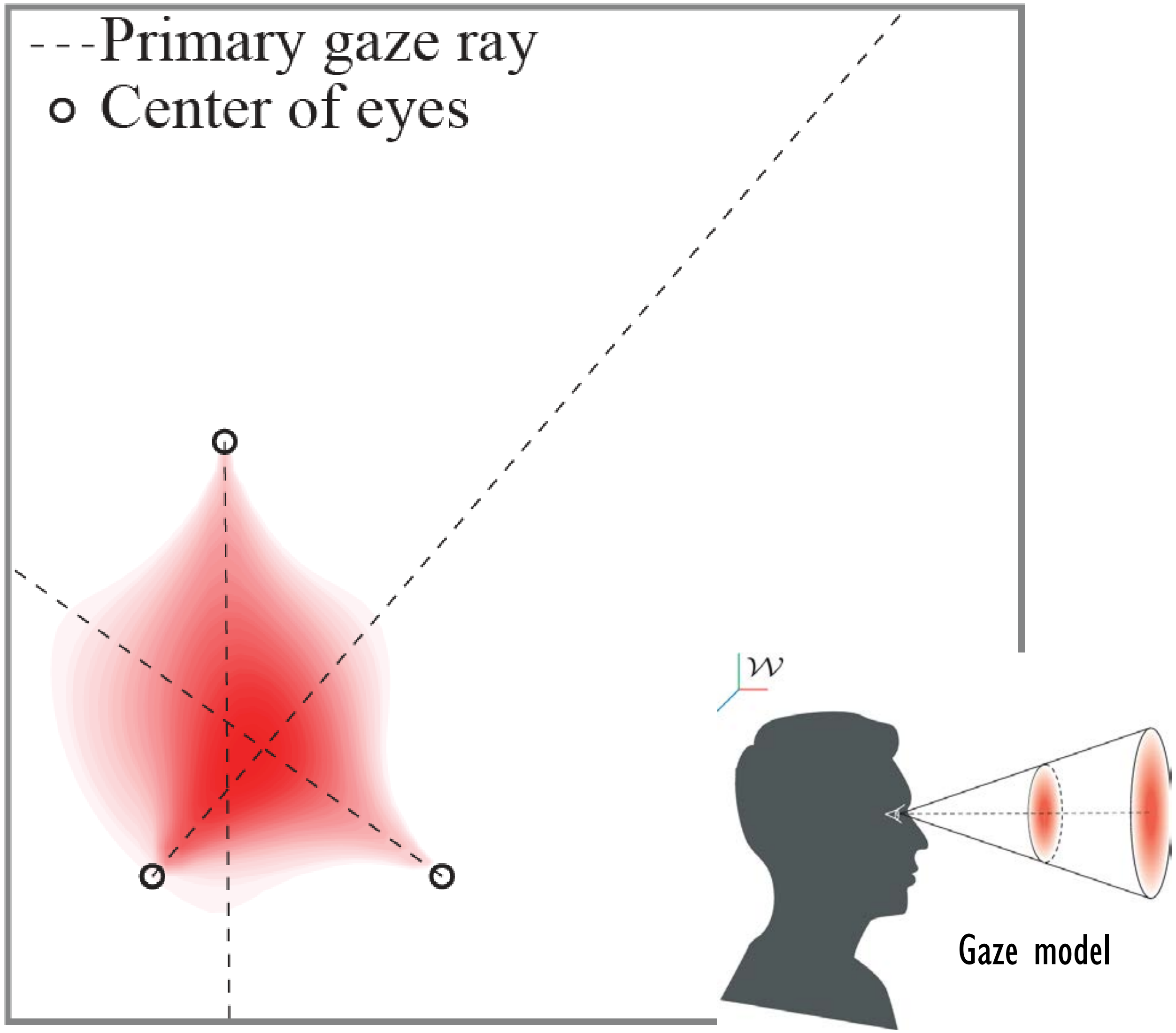
[Park NIPS12]

3D Gaze Registration



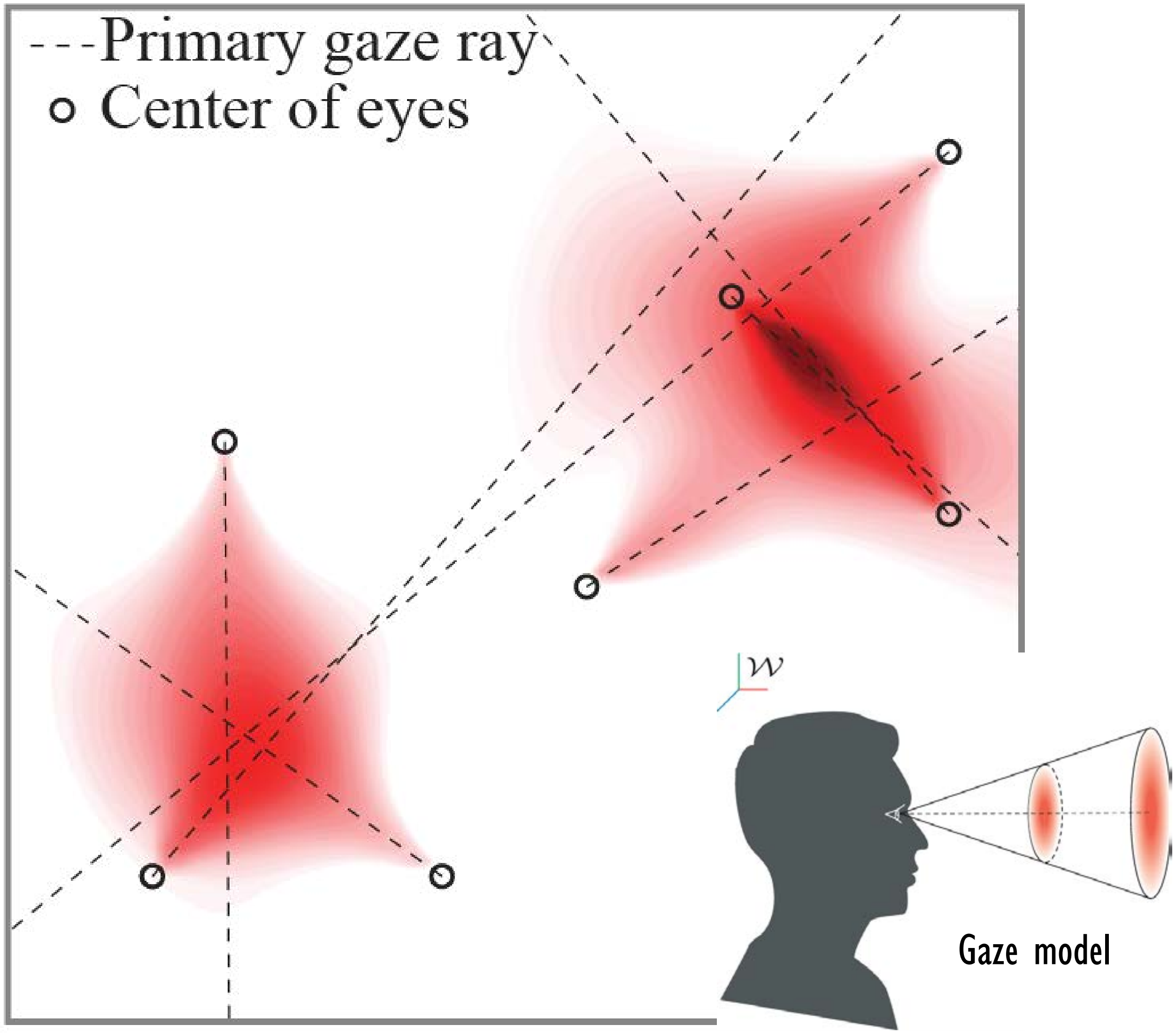
[Park NIPS12]

Social Saliency



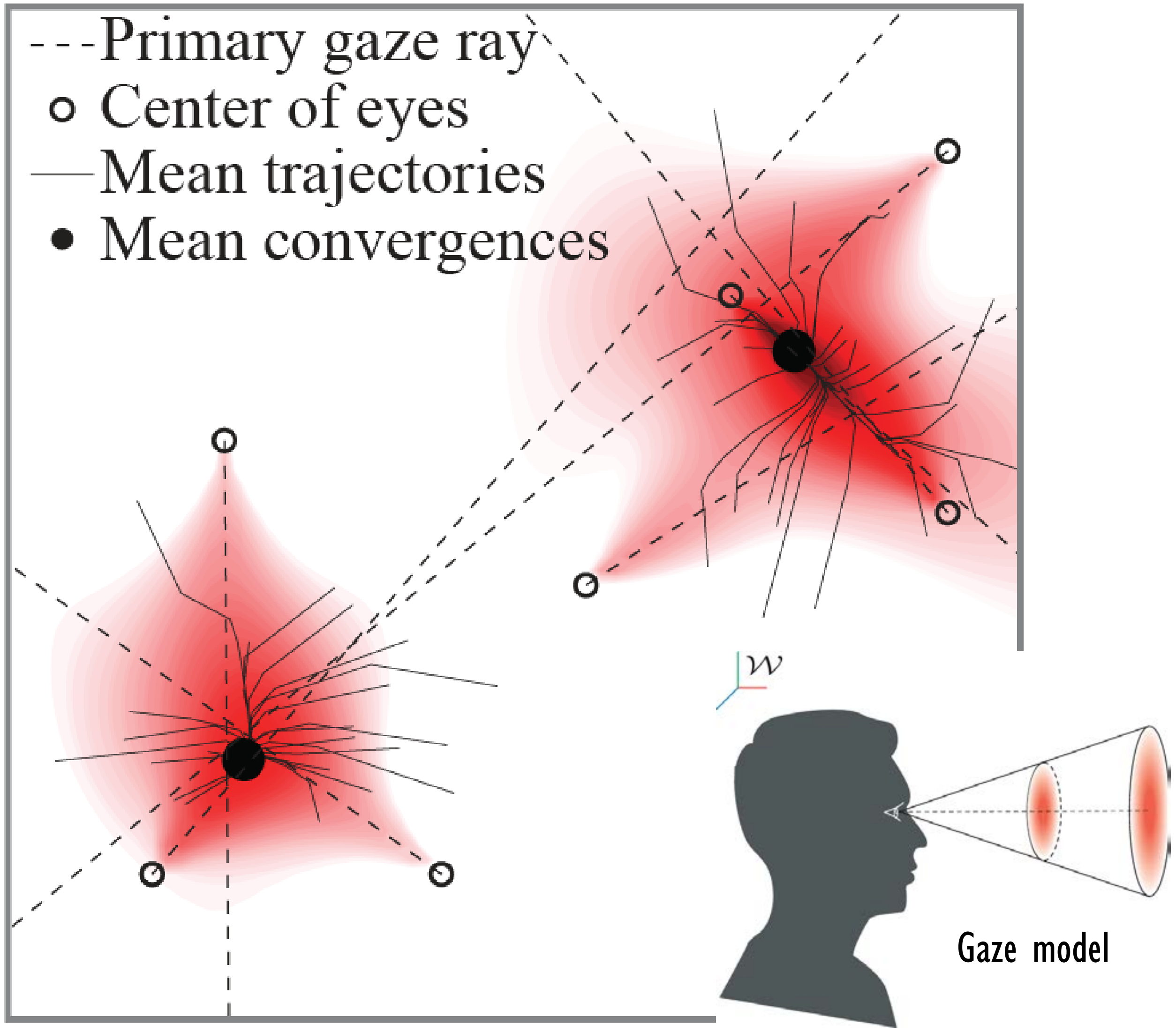
[Park NIPS12]

Social Saliency



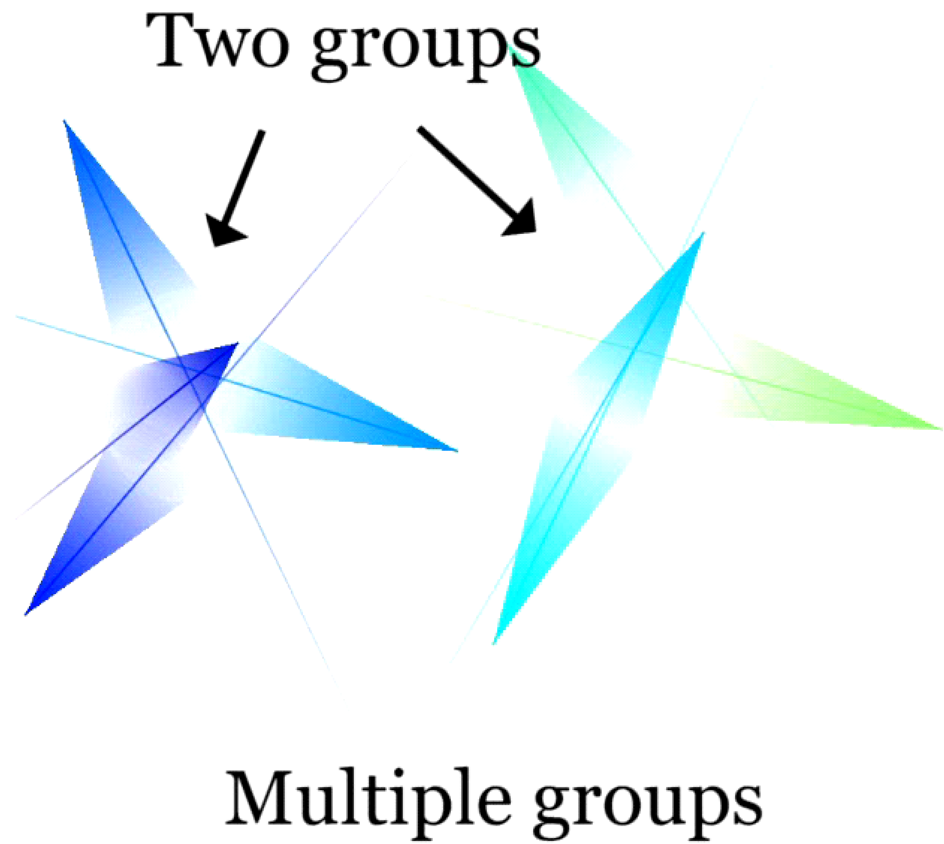
[Park NIPS12]

3D Joint Attention via Mode-seeking



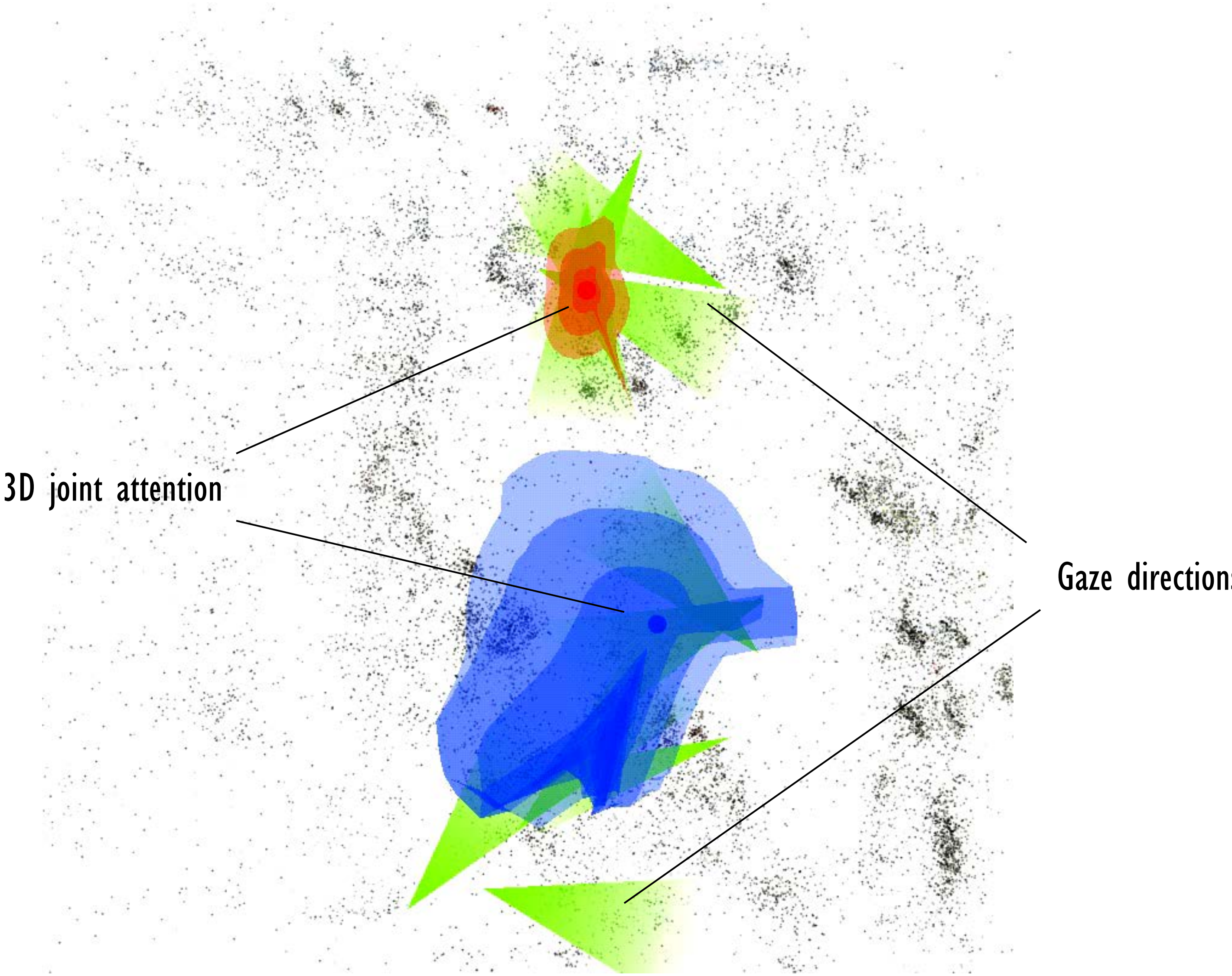
[Park NIPS12]

Mode-seeking: Gaze Concurrences

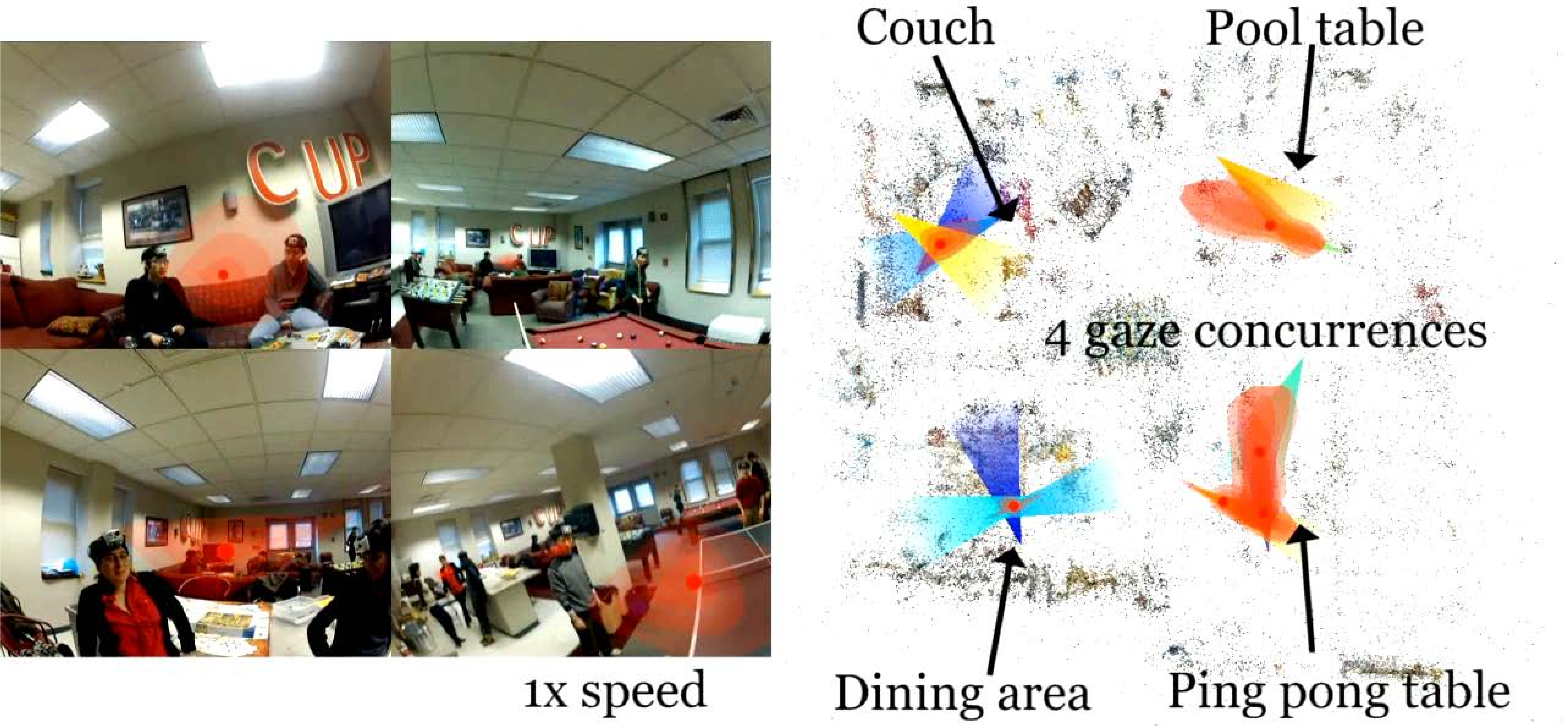


[Park NIPS12]

3D Joint Attention Reconstruction

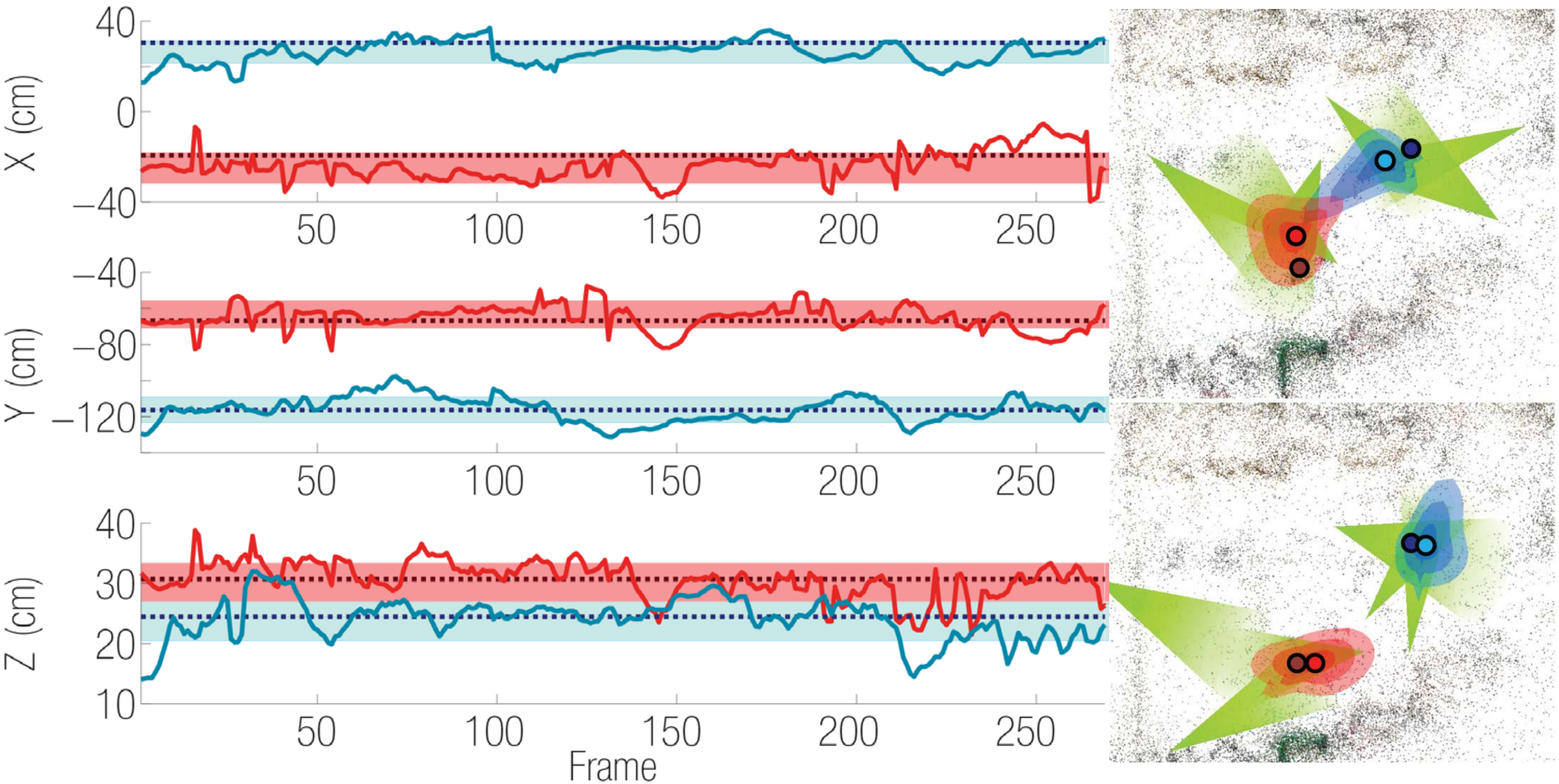


[Park NIPS12]

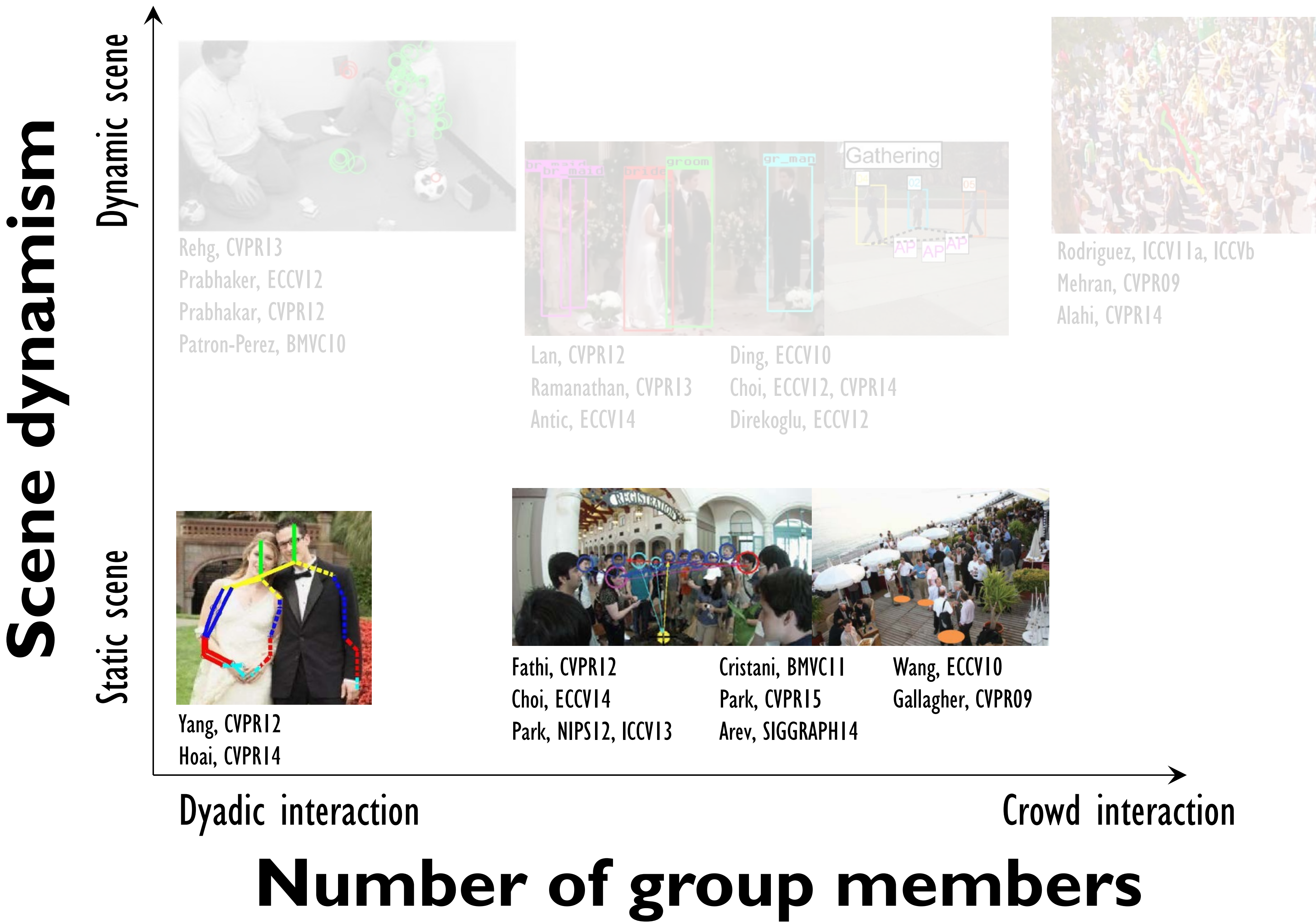


3D joint attention

[Park NIPS12]



[Park NIPS12]



Scene dynamism
Dynamic scene
Static scene
Rehg, CVPR13
Prabhaker, ECCV12
Prabhakar, CVPR12
Patron-Perez, BMVC10
Gathering
Lan, CVPR12
Ramanathan, CVPR13
Antic, ECCV14
Ding, ECCV10
Choi, ECCV12, CVPR14
Direkoglu, ECCV12
Rodriguez, ICCV11a, ICCVb
Mehran, CVPR09
Alahi, CVPR14
Yang, CVPR12
Hoai, CVPR14
Fathi, CVPR12
Choi, ECCV14
Park, NIPS12, ICCV13
Cristani, BMVC11
Park, CVPR15
Arev, SIGGRAPH14
Wang, ECCV10
Gallagher, CVPR09
Dyadic interaction
Crowd interaction
Number of group members

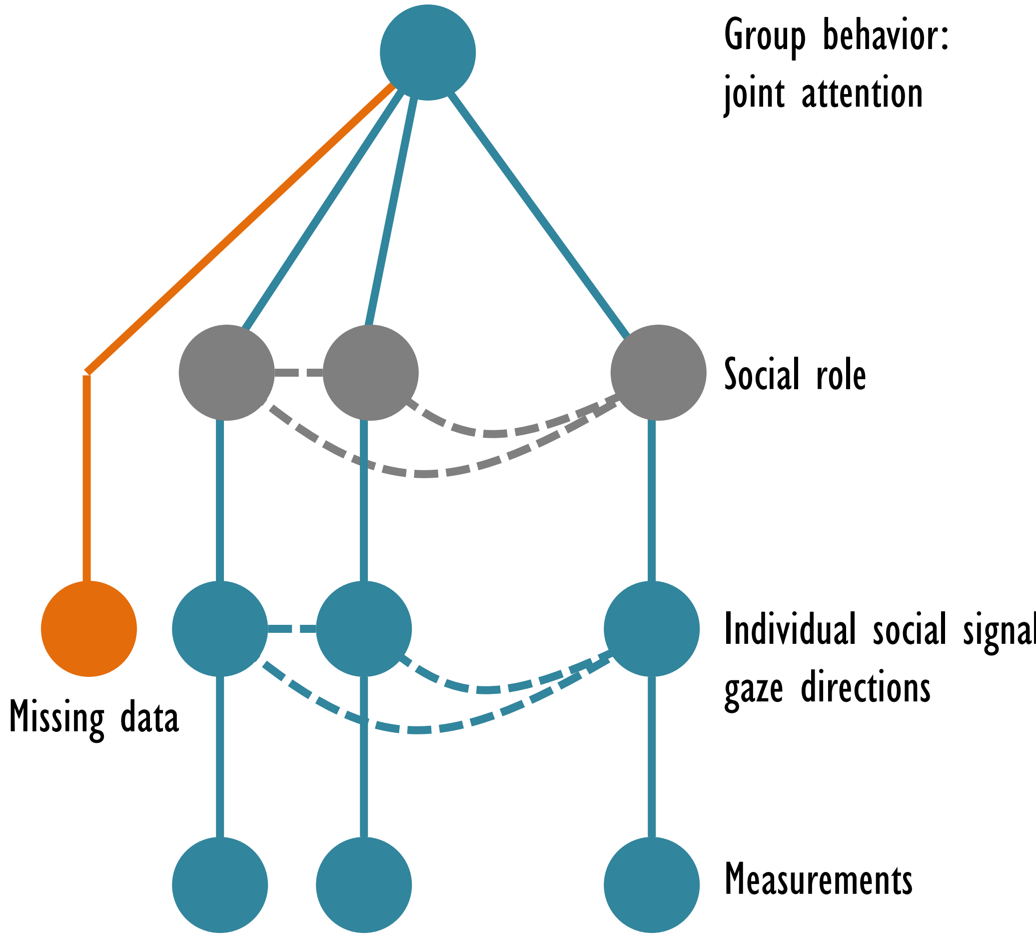
Applications of Joint Attention

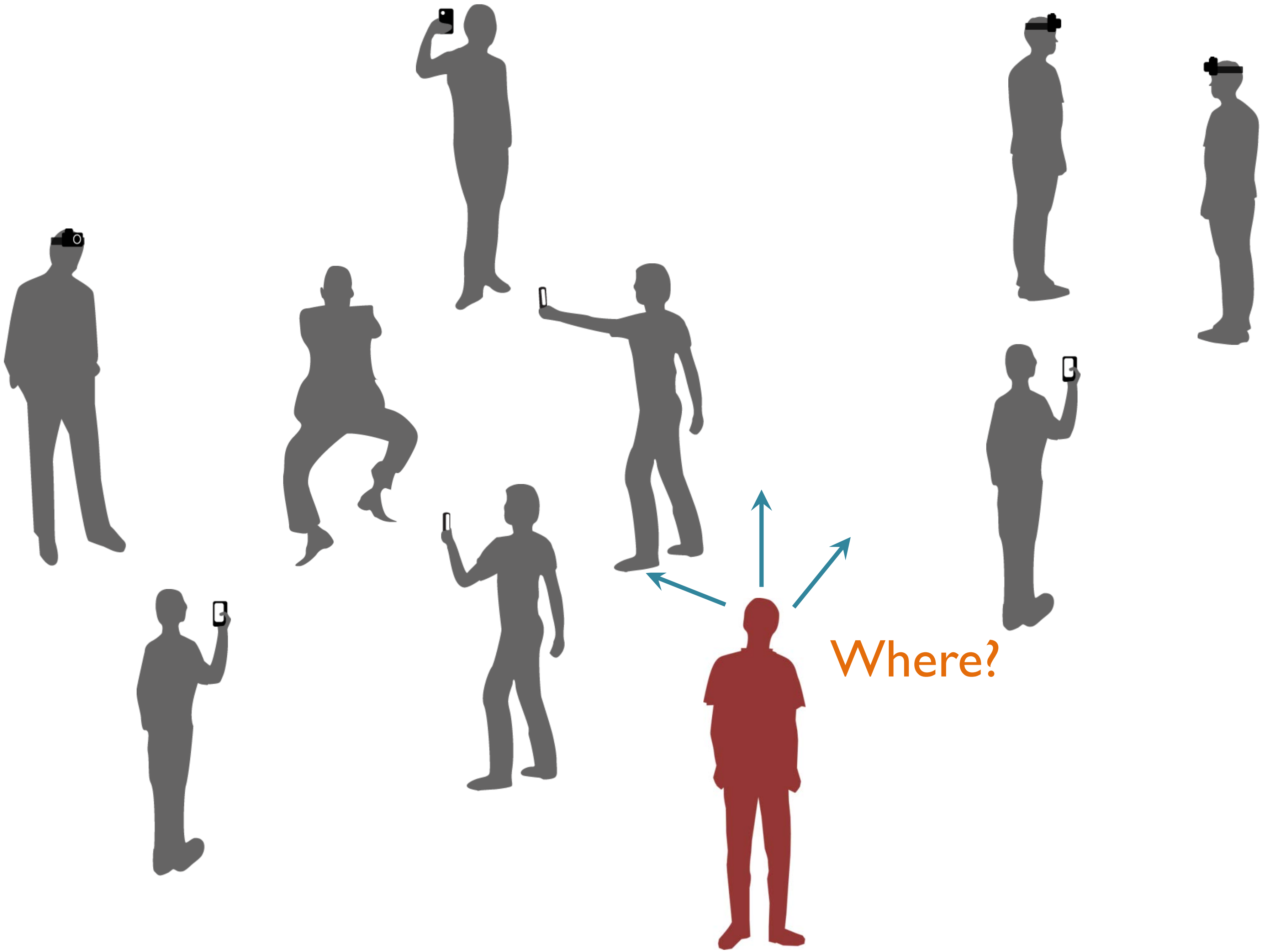
Gaze Prediction

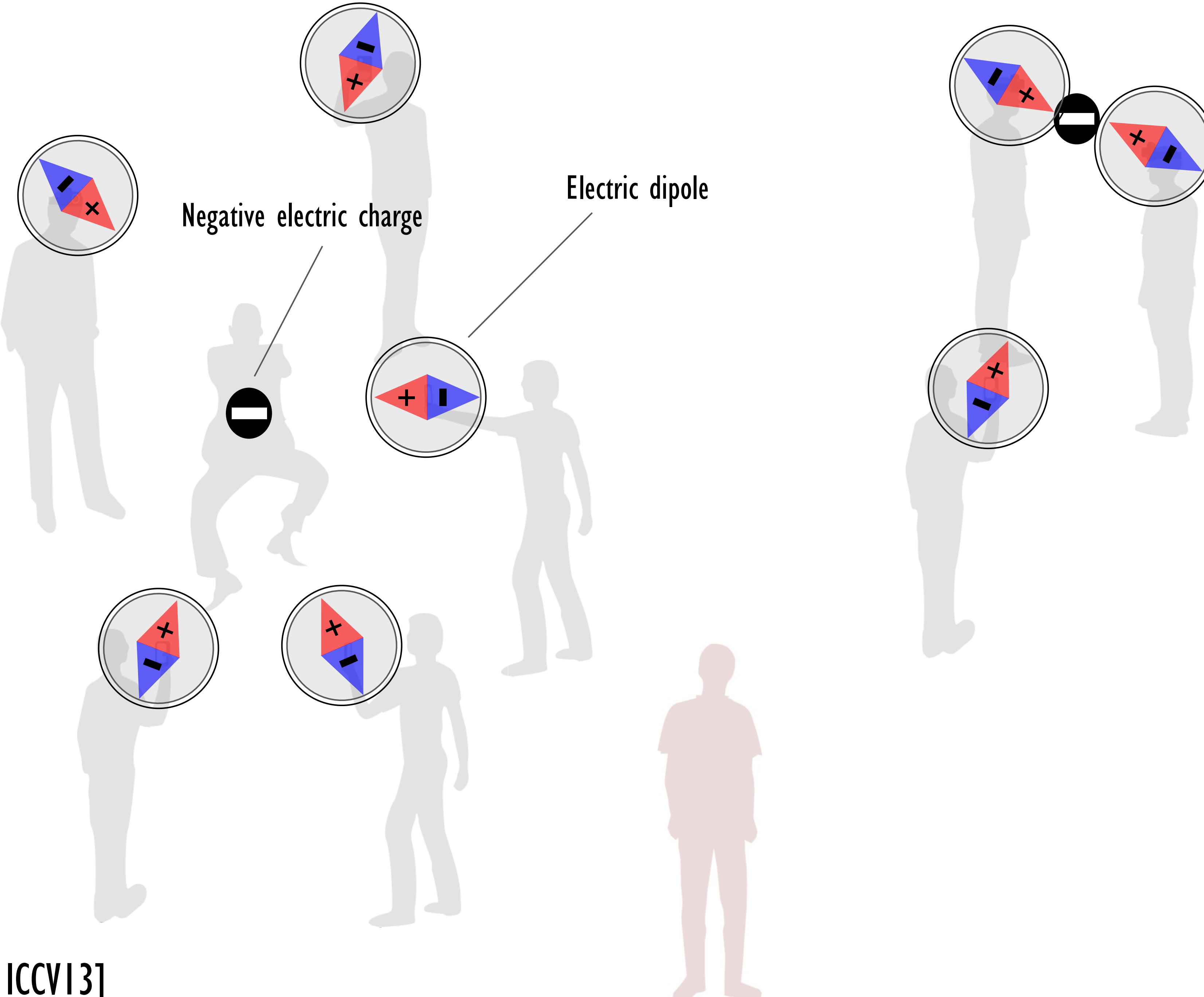
[Park ICCV13]

Input: images of social interactions

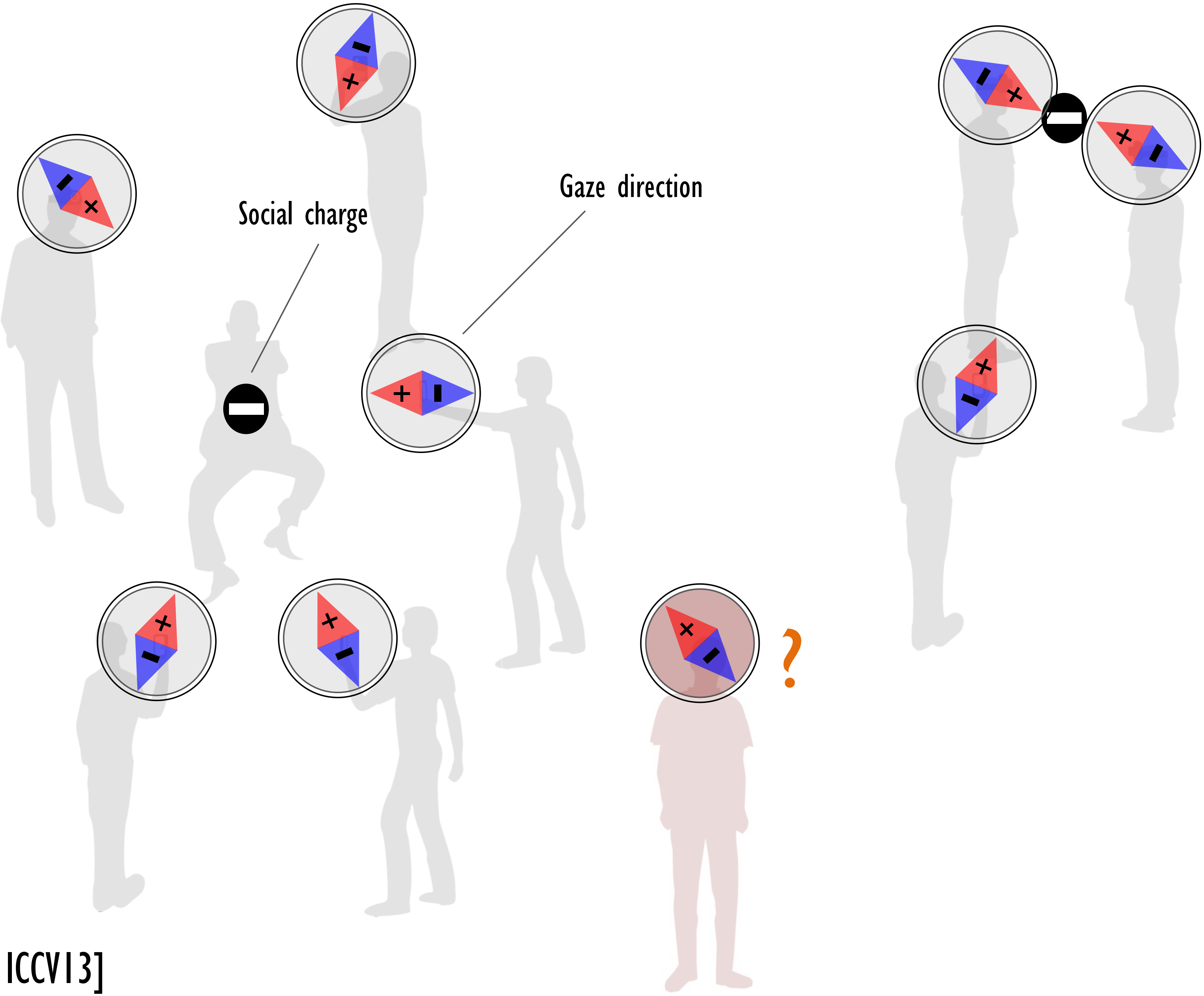
Output: to predict gaze direction




Where?

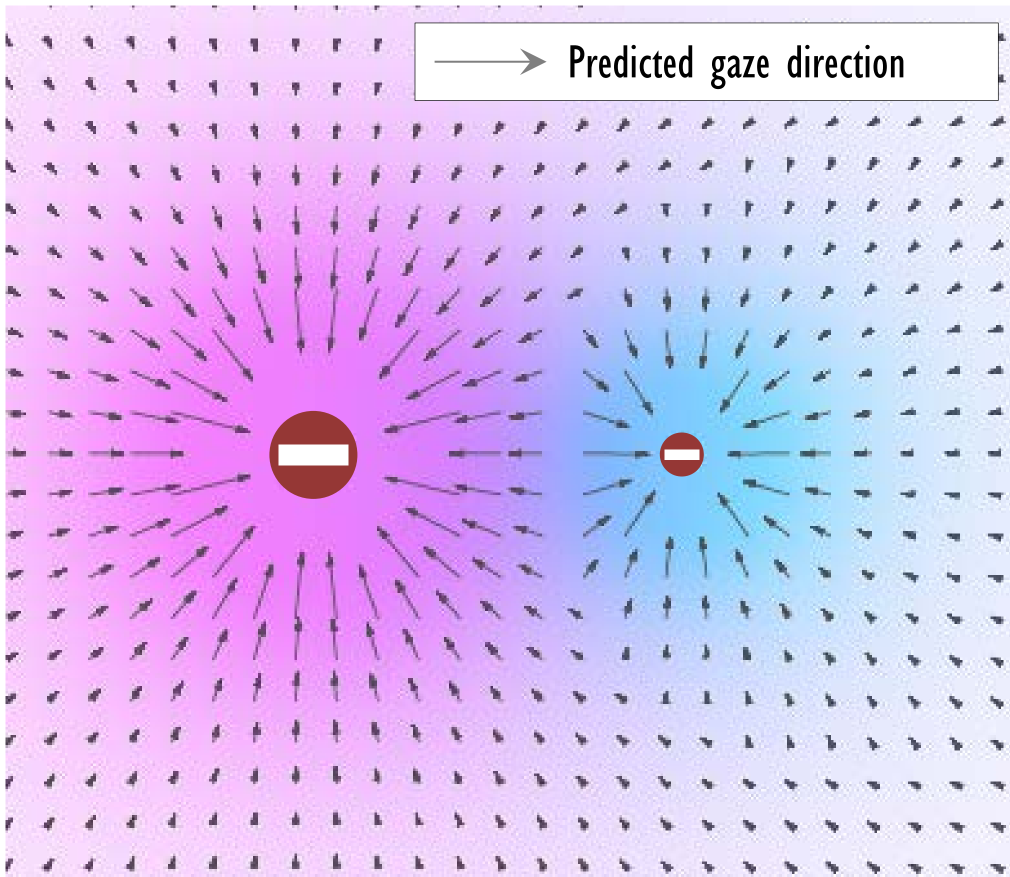


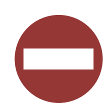
[Park ICCV13]



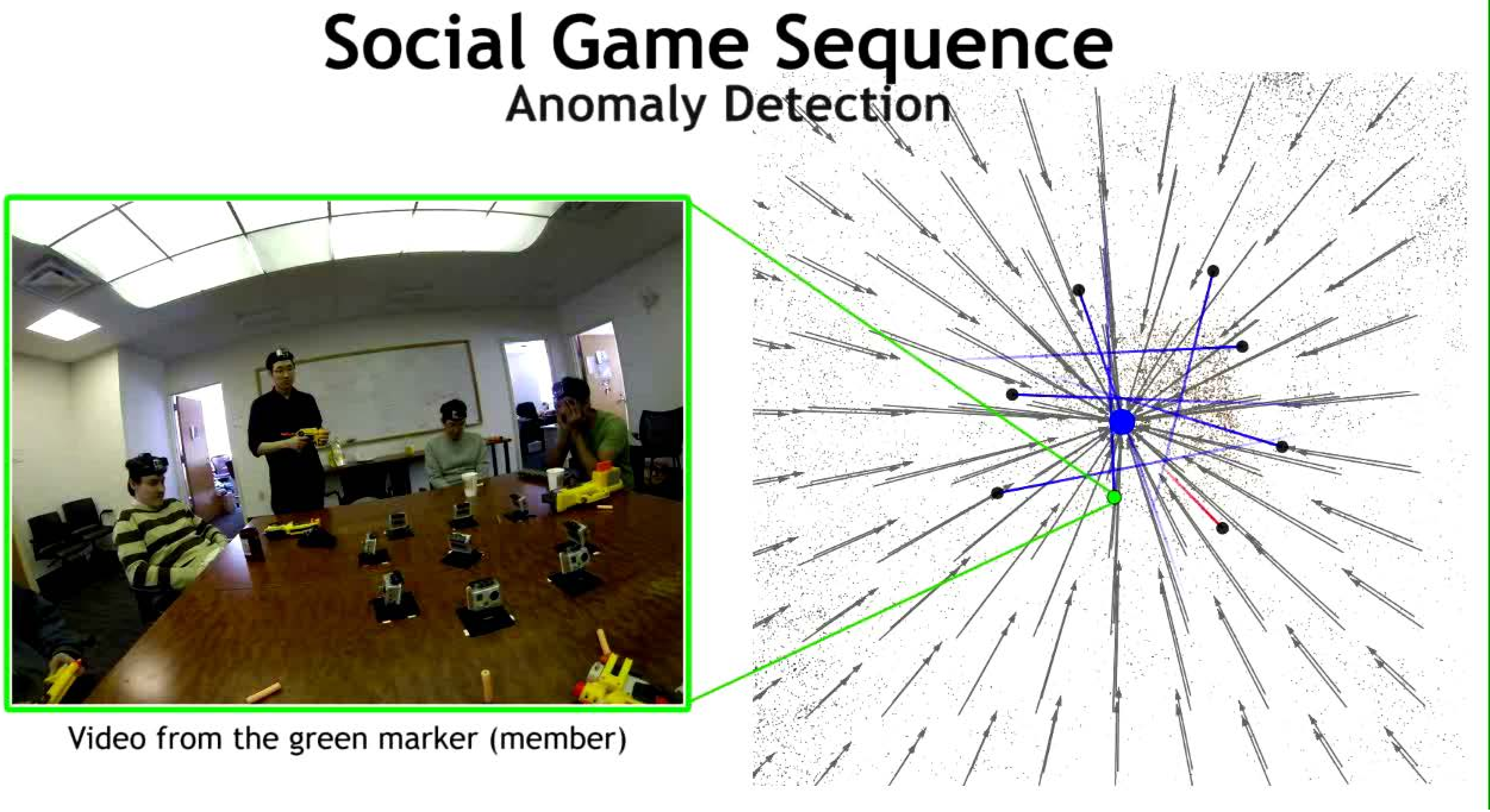
[Park ICCV13]

Gaze Field



Gaze field $\mathbf{G} = \nabla\Phi($  $)$

[Park ICCV13]


Social Game Sequence
Anomaly Detection
Video from the green marker (member)

[Park ICCV13]

Mafia Game

[Park ICCV13]

Mafia Game

Prediction for Missing Data

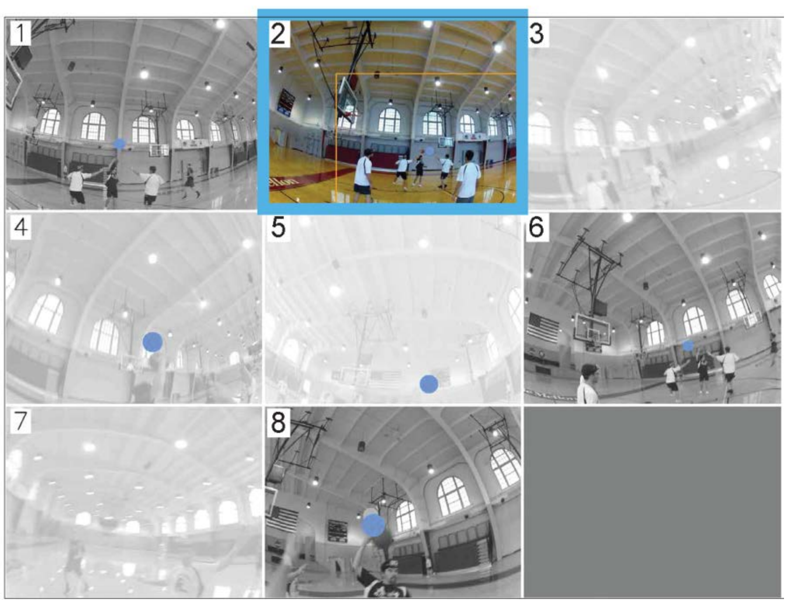
[Park ICCV13]

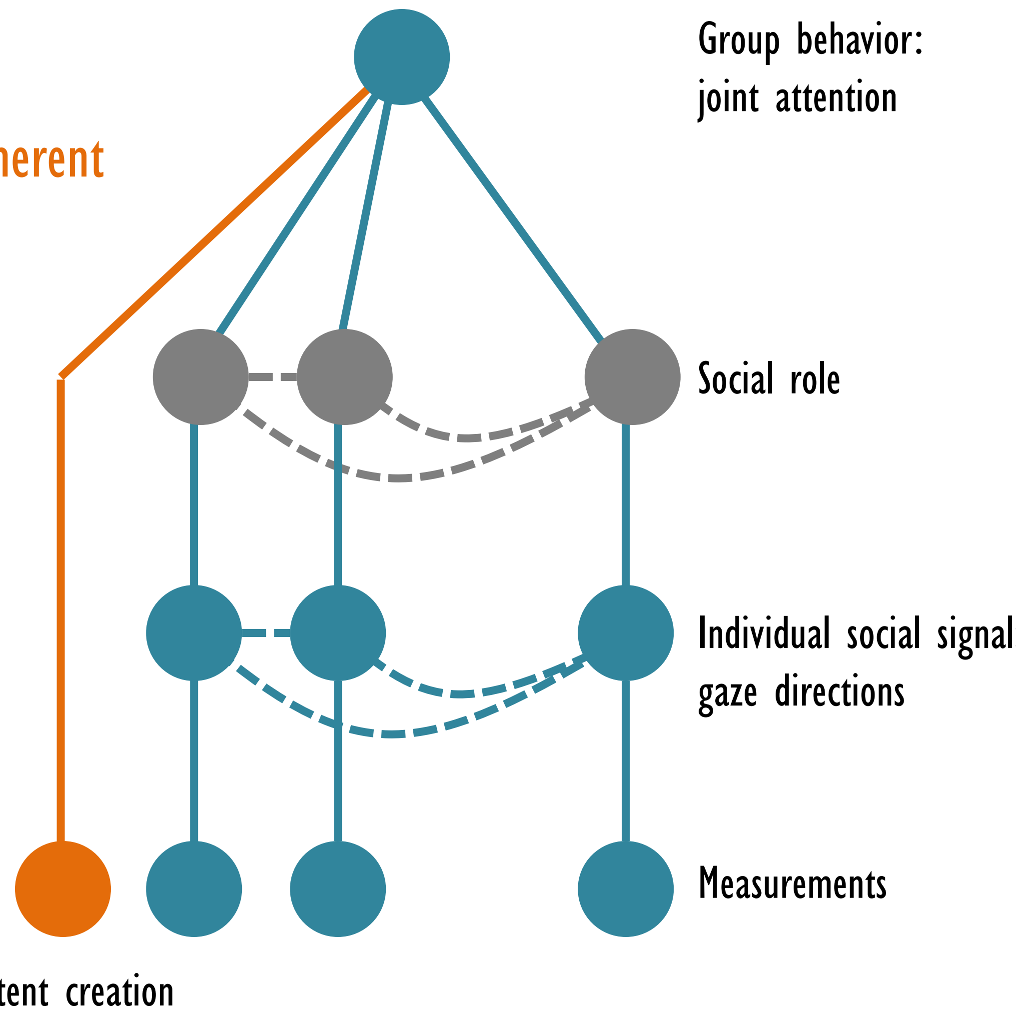
Social Footage Editing

[Arev SIGGRAPH14]

Input: videos of social interactions

Output: to edit videos to produce a coherent story of social events.





First Person Cameras

Problems of videos taken by social cameras:

- Produce too much information to digest at once
- Are biased by an intimate and personal view

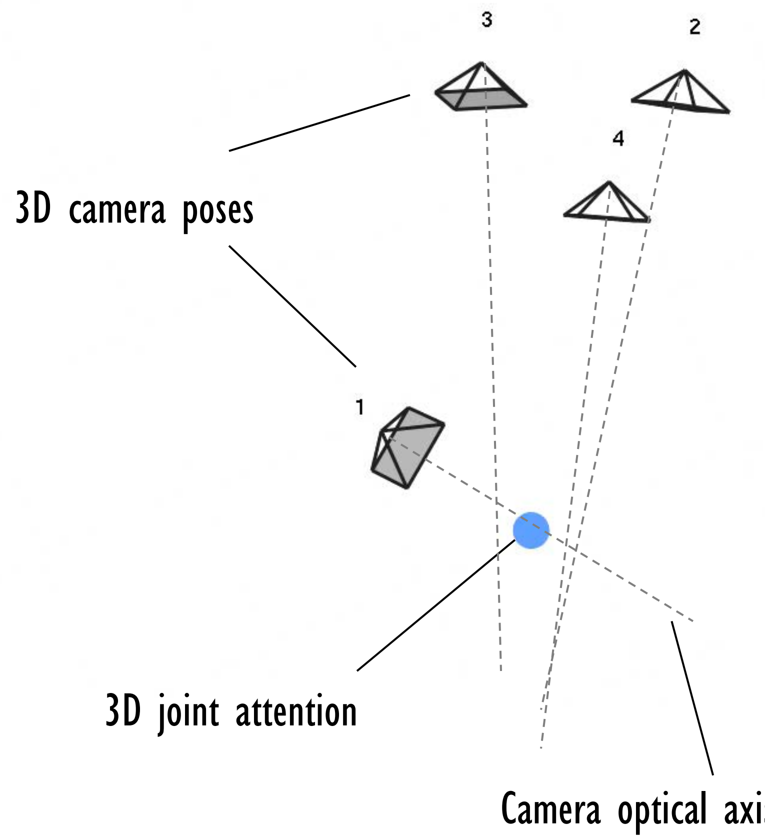
Input: Synchronized Videos

[Arev SIGGRAPH14]

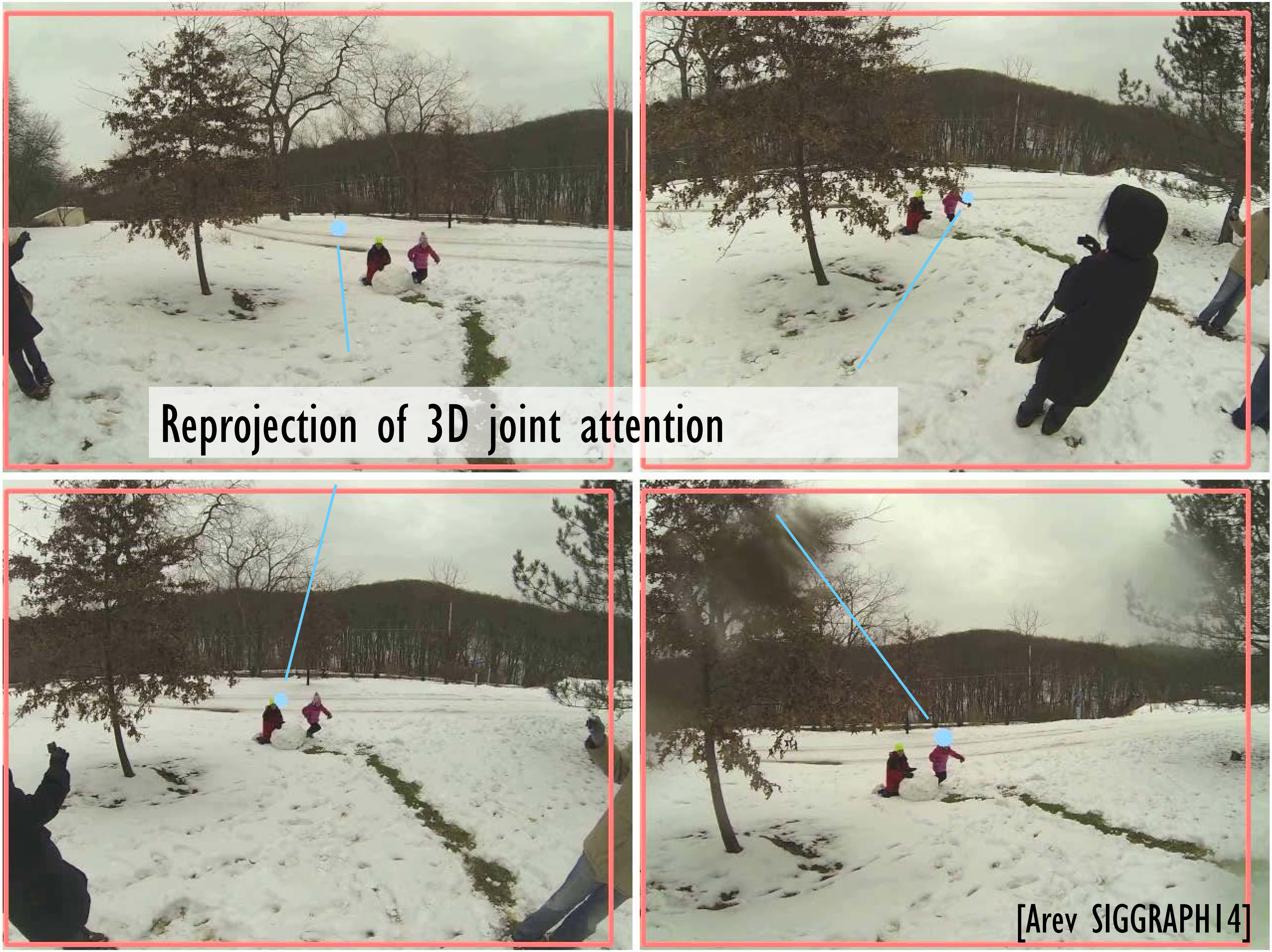
Output: Coherent Video of Event

[Arev SIGGRAPH14]

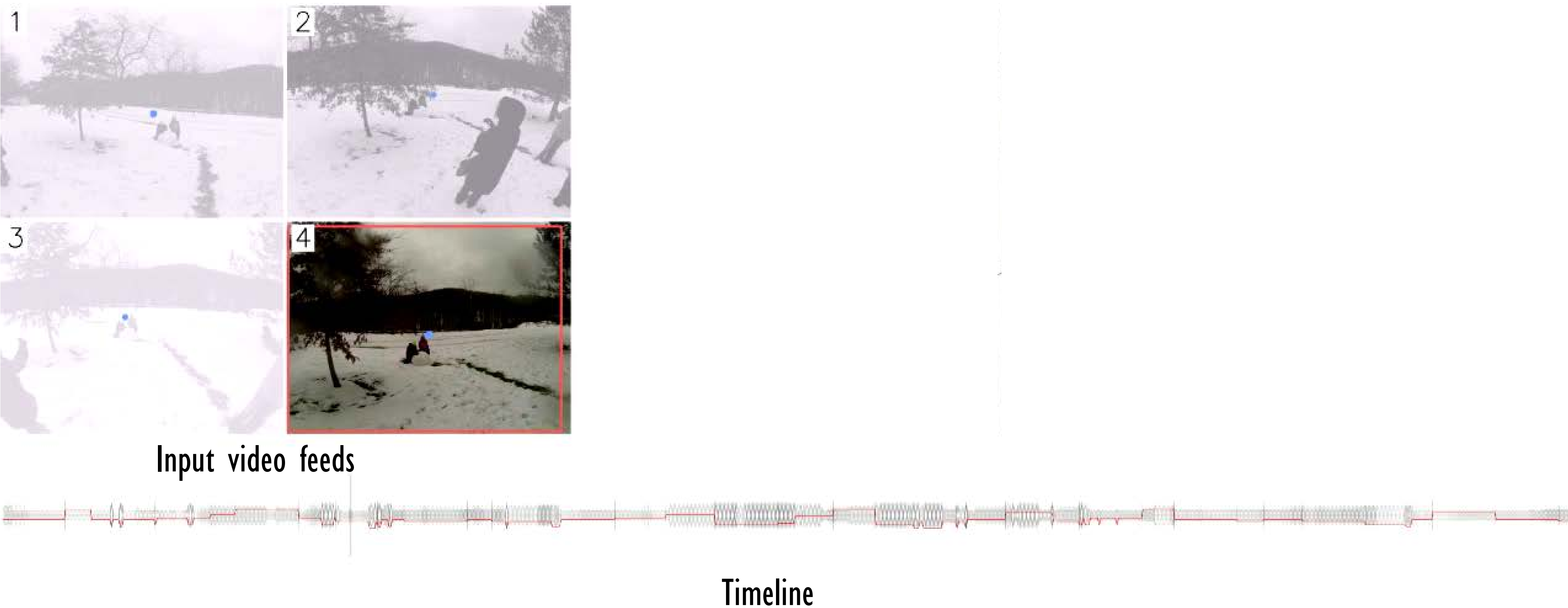
Content: 3D Joint Attention



[Arev SIGGRAPH14]


Reprojection of 3D joint attention
[Arev SIGGRAPH14]

Automatic Video Editing



[Arev SIGGRAPH14]

Basketball Scene

[Arev SIGGRAPH14]

Basketball Scene

[Arev SIGGRAPH14]

Scene Summarization

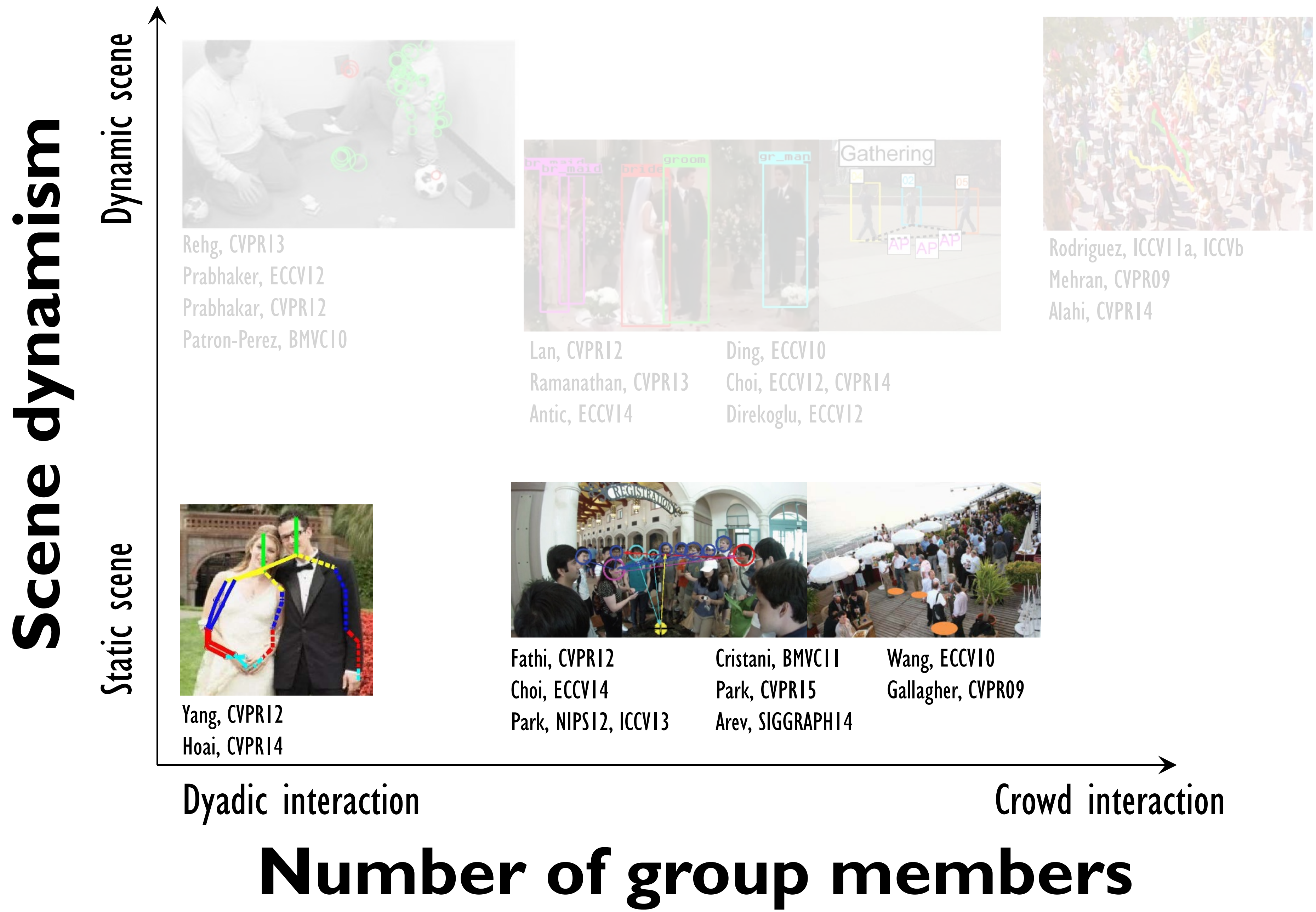
[Arev SIGGRAPH14]

Surprise Party Scene

Our method

Professional Editor

[Arev SIGGRAPH14]



Scene dynamism
Dynamic scene
Static scene
Rehg, CVPR13
Prabhaker, ECCV12
Prabhakar, CVPR12
Patron-Perez, BMVC10
Gathering
Lan, CVPR12
Ramanathan, CVPR13
Antic, ECCV14
Ding, ECCV10
Choi, ECCV12, CVPR14
Direkoglu, ECCV12
Rodriguez, ICCV11a, ICCVb
Mehran, CVPR09
Alahi, CVPR14
Yang, CVPR12
Hoai, CVPR14
REGISTRATION
Fathi, CVPR12
Choi, ECCV14
Park, NIPS12, ICCV13
Cristani, BMVC11
Park, CVPR15
Arev, SIGGRAPH14
Wang, ECCV10
Gallagher, CVPR09
Dyadic interaction
Crowd interaction
Number of group members