Social Statics
Scene dynamism

Dyadic interaction

Crowd interaction

Number of group members

Static scene

Dynamic scene

Rehg, CVPR13
Prabhaker, ECCV12
Prabhakar, CVPR12
Patron-Perez, BMVC10

Yang, CVPR12
Hoai, CVPR14

Lan, CVPR12
Ramanathan, CVPR13
Antic, ECCV14

Ding, ECCV10
Choi, ECCV12, CVPR14
Direkoglu, ECCV12

Fathi, CVPR12
Choi, ECCV14
Park, NIPS12, ICCV13

Cristani, BMVC11
Park, CVPR15
Arev, SIGGRAPH14

Wang, ECCV10
Gallagher, CVPR09

Rodriguez, ICCV11a, ICCVb
Mehran, CVPR09
Alahi, CVPR14
Measurements

Individual social signal

Group behavior

Social role

Individual social signal

Measurements
Measurements

Individual social signal

Group behavior

Social role

Individual social signal

Measurements
Measurements

Individual social signal

Group behavior

Social role

Individual social signal

Measurements
Measurements

Individual social signal

Group behavior

Social role

Input

Output

Model

Measurements

Individual social signal

Social role

Group behavior

Input

Output

Model
Gaze Estimation

Eye tracker (local coordination)

Looking-out camera (global coordination)
Gaze Estimation w/o Eye Tracker

Point of regard

Point of regard
Social Signal Perception (Gaze)

[Li ICCV13]

Input: image or video of first person view
Output: localization of fixation point.
GTEA Gaze+ Dataset
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

Left Hand | Right Hand | Hands Together | Hands Apart

Gaze Shift vs Left Hand | Gaze Shift vs Right Hand | Gaze Shift vs Intersecting Hands | Gaze Shift vs Both Hands

[Li ICCV13]
Gaze Prediction
Application to Foreground Segmentation

Foreground Object

75.2%

[Li ICCV13]
Social Signal Perception (Gaze)

Input: image or video of first person view
Output: to calibrate the fixed relationship between gaze and head-mounted camera.

[Park NIPS12]
Eye-in-head Motion

[Park NIPS12]
Gaze Distribution

von Mises-Fisher distribution

\[ P(\mathbf{v}; \mathbf{v}, \kappa) = \frac{\kappa}{4\pi \sinh \kappa} \exp(\kappa \mathbf{v}^T \mathbf{v}) \]

\[ \cos(\theta) = \mathbf{v}^T \mathbf{v} \]

- Primary gaze direction
- Center of eyes

[Park NIPS12]
Cone-shaped Gaze Model

Gaze parameterization via von Mises-Fisher distribution

\[ g = P(v; \bar{v}, \kappa) = \frac{\kappa}{4\pi \sinh \kappa} \exp(\kappa v^T \bar{v}) \]

[Park NIPS12]
Gaze Calibration

Camera pose

Primary gaze ray

Gaze parameterization via von Mises-Fisher distribution

\[ \mathbf{g} = P(\mathbf{v}; \bar{\mathbf{v}}, \kappa) = \frac{\kappa}{4\pi \sinh \kappa} \exp(\kappa \mathbf{v}^T \bar{\mathbf{v}}) \]

Camera pose

\[ C = C(R, C) \]

[Park NIPS12]
Gaze Calibration

Fixed relationship between gaze and camera
\( \overline{v}(R) \)  \( p(R,C) \)

[Park NIPS12]
Gaze Calibration

\[ X = p + \lambda v \]

Point of regard

\[ X \in \mathbb{R}^3 \]
Gaze Calibration

\[ X = p + \lambda v \]

[Park NIPS12]
Gaze Calibration

$$X = p + \lambda v$$

[Park NIPS12]
Gaze Calibration

Points of regard

[Park NIPS12]
Gaze Calibration

Points of regard

\[ p_0 \]  \[ \mathbf{v} \]

\( \mathbf{v} \): Linear regression

[Park NIPS12]
Gaze Calibration

$p = p_0 - \lambda v$

$\lambda \approx \lambda_0$

[Park NIPS12]
Gaze Calibration

\[ p = p_0 - \lambda v \]

\[ \lambda \ll \lambda_0 \]
Gaze Calibration

\[ p = p_0 - \lambda v \quad \lambda \gg \lambda_0 \]

Points of regard

[Park NIPS12]
Gaze Calibration

\[ p = p_0 - \lambda v \]

minimize \( \xi + w \lambda \)
subject to \( \frac{b_i}{a_i + \lambda} \leq \xi, \quad \forall i = 1, \ldots, n \)
\( \lambda \geq 0 \)

[Park NIPS12]
Gaze Ray Calibration

Camera pose

Cone-shaped gaze distribution

Primary gaze ray

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

Input: image or video of third person view
Output: to find head / to estimate head direction

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

[Schneiderman CVPR00, Xiong CVPR13]
Body Configuration

[Hoai CVPR14]

Input: images of TV shows
Output: to detect social interactions

Group behavior: pose configuration
Social role
Individual social signal: upper body bounding box
Measurements

[Image: Diagram showing the flow from input to output with labeled nodes for group behavior, social role, individual social signal, and measurements.]
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

Location

Scale
Common Configuration of Upper Body

More than Three People

[Hoai CVPR14]
Input image

DPM

Dense detection scores at multiple location and scales

Fast inference

Best configuration:
+ High unary scores
+ High similarity to a common configuration

Output

Learned configurations

[Hoai CVPR14]
Comparison w/ DPM
Joint Attention

[Fathi CVPR12]

Input: images from a first person camera
Output: to detect social interactions
Where Do They Look?

[Fathi CVPR12]
MRF Modeling

Space of interactions

Top view

[Fathi CVPR12]
MRF Modeling

MRF modeling: unary potential

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

- Head direction is aligned with the point of regard.
  \[ \phi_1(L_{f_i} = \ell, P_{f_i}) = \frac{1}{\sigma_1 \sqrt{2\pi}} \exp \left\{ -\frac{||V_{f_i} - (\ell - T_{f_i})||^2}{2\sigma_1^2} \right\} \]

- The point of regard cannot be the himself.
  \[ \phi_2(L_{f_i} = \ell, P_{f_i}) = \frac{1}{1 + \exp \left\{ -c_2 ||\ell - P_{f_i}|| \right\}} \]

- The point of regard is likely to be a face.
  \[ \phi_3(L_{f_i} = \ell, P_{f_1}, \ldots, P_{f_N}) = \begin{cases} c_3 & \ell = P_{f_j} \forall j \neq i \\ 1 & \text{otherwise} \end{cases} \]

[Fathi CVPR12]
MRF Modeling

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 \( f_i \)

MRF modeling: binary potential
\[ \phi_B(L_{f_i} = \ell_1, L_{f_j} = \ell_2) = \begin{cases} c_B & \text{if} (\ell_1 = \ell_2) \\ 1 - c_B & \text{if} (\ell_1 \neq \ell_2) \end{cases} \]

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]
First-Person
Detection of Social Interaction
Joint Attention

Input: images from first person cameras
Output: to localize joint attention in 3D

[Park NIPS12]
Input Video: Meeting Scene

1x speed

[Park NIPS12]
3D Camera Pose Estimation
(Structure from motion)

Two groups

6x speed

[Park NIPS12]
Gaze Ray Calibration

Camera pose

Cone-shaped gaze distribution

Primary gaze ray

Primary gaze ray with respect to the camera pose

[Park NIPS12]
3D Gaze Registration

[Park NIPS12]
3D Gaze Registration

--- Primary gaze ray
○ Center of eyes
Social Saliency

--- Primary gaze ray
○ Center of eyes

[Park NIPS12]
Social Saliency

--- Primary gaze ray
○ Center of eyes

[Park NIPS12]
3D Joint Attention via Mode-seeking

--- Primary gaze ray
- Center of eyes
- Mean trajectories
- Mean convergences

Gaze model

[Park NIPS12]
Mode-seeking: Gaze Concurrences

Two groups

Multiple groups

[Park NIPS12]
3D Joint Attention Reconstruction

[Park NIPS12]
3D joint attention

Couch

Pool table

Dining area

Ping pong table

1x speed

3D joint attention

Park NIPS12
[Park NIPS12]
Applications of Joint Attention
Gaze Prediction

Input: images of social interactions
Output: to predict gaze direction

[Park ICCV13]
[Park ICCV13]
Gaze Field

\[ \mathbf{G} = \nabla \Phi \]

\[ \text{Predicted gaze direction} \]

Gaze field
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

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

[Arev SIGGRAPH14]
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
Output: Coherent Video of Event
Content: 3D Joint Attention

3D camera poses

3D joint attention

Camera optical axis

[Arev SIGGRAPH14]
Reprojection of 3D joint attention
Automatic Video Editing
Basketball Scene
Scene Summarization
Surprise Party Scene

Our method

Professional Editor

[Arev SIGGRAPH14]