Social Interactions: A First-Person Perspective

Alireza Fathi, Jessica K. Hodgins, James M. Rehg
CVPR 2012

Bora Çelikkale

General Idea

Detecting and characterizing social interactions

Memorable Moments

“Egocentric” video : Viewed from one's own mind as a center.

Related Work

First-Person Wearable Sensors

Recognize daily activities as meal preparation - 2011

Recognize atomic actions as turn-left - 2011

SenseCam :
Automatic capture
based on lighting change
- 2004

Related Work

Social Networks

Recover social network and patterns of influence between individuals - 2004

Related Work

Activity Recognition

Recognizing atomic / group activities – 2008

Why First-Person Camera?

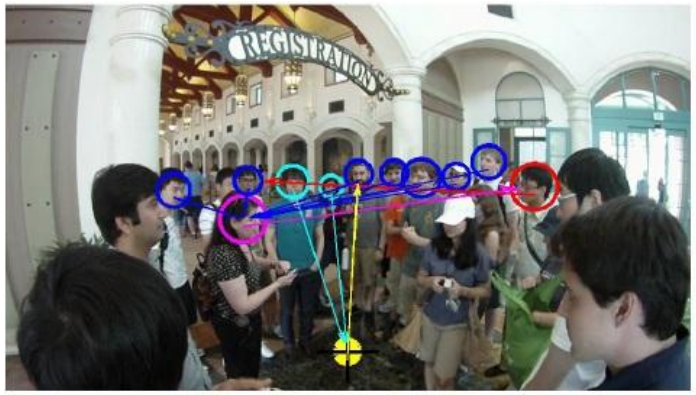
Always records where wearer is attending

Occlusions are less likely

Hard to track all individuals with static cameras

Details

Main Source of Information : Faces

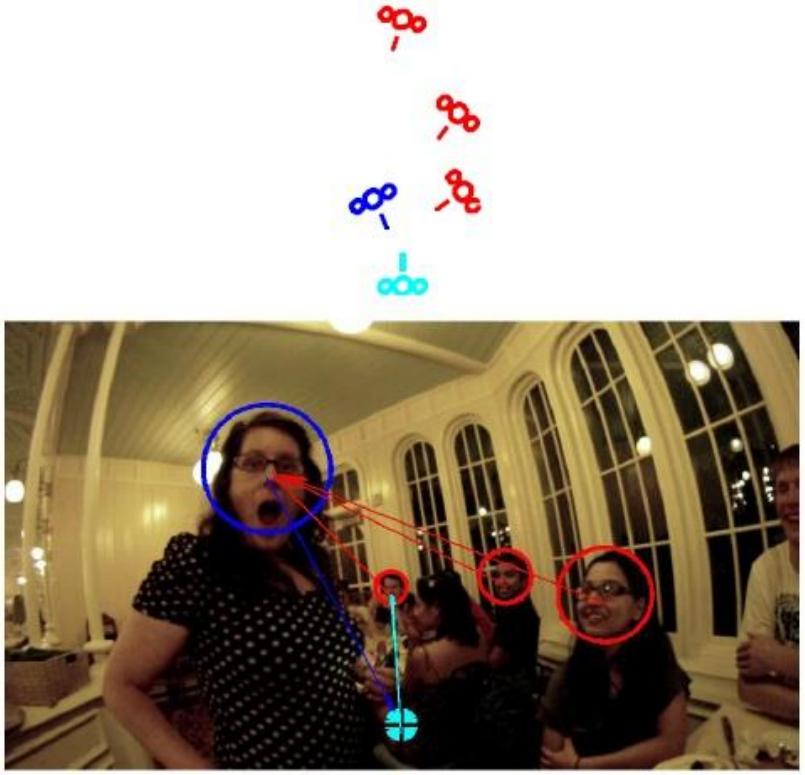


Why?

- Primary role in social interactions
- Robust face detection and recognition methods

Details

Estimation of location of faces in 3D space
(approximate estimations)



View Angle Θ from camera

$$\Theta = r/f$$

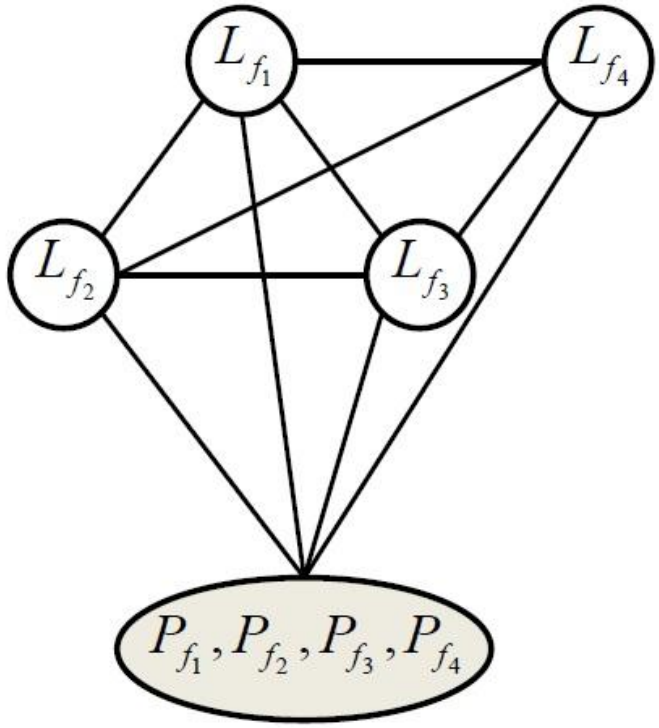
Distance d from camera

$$d = c/h$$

Details

Assumptions:

- Looking at something in the direction of orientation
- Looking at people rather than objects
- Looking at same location



Details

Unary Potentials:

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

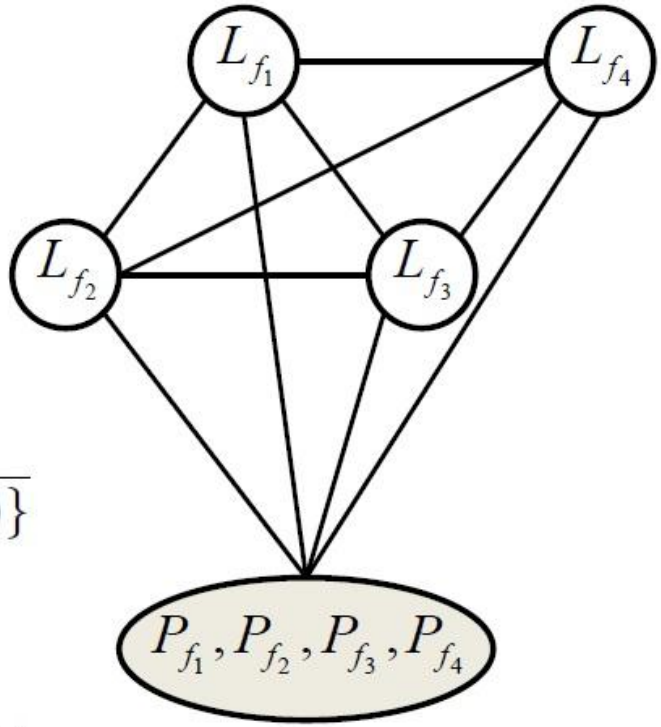
Gaussian function: Possibility of f_i looking at location ℓ based on location and orientation of f_i

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

Sigmoid function: How close F_i=ℓ can be. To avoid looking at itself

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

Bias faces to look at where other faces are

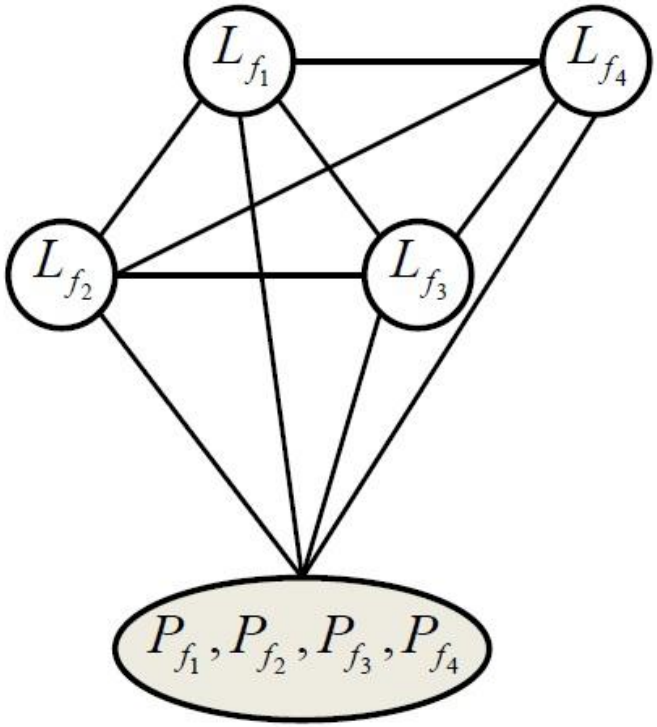


Details

Pairwise Potentials:

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

Bias faces to look at where other faces are looking at

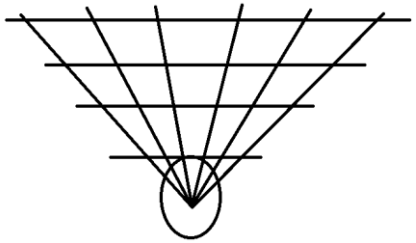


Method

Features

Location of Faces Around First Person

- 5 angular bins (-75 to 75 degrees)
- 4 distance bins (0 to 5m)
- 20 dimensional histogram



Analyze Attention and Roles Over Time

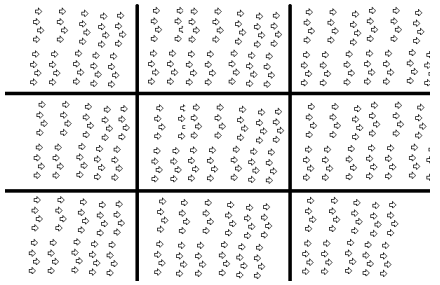
Assign roles to individuals based on:

- Faces looking at him
- Whether first person looks at him
- Mutual attention with first person
- Faces looking at same location

- 4 dimensional feature vector

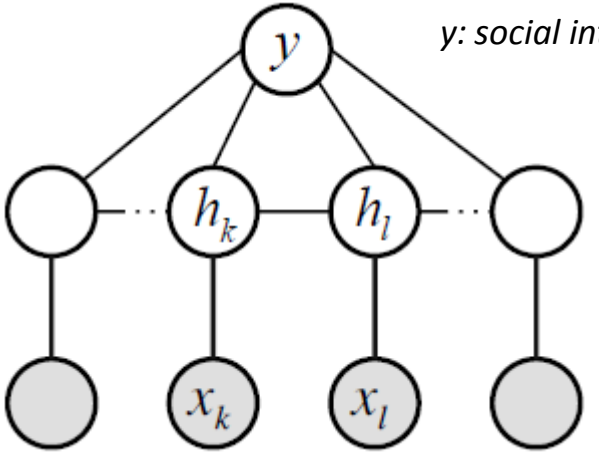
First Person Head Movement

- Head motion patterns
- Optical flow



Method

Temporal Model:



y: social interaction label. Binary in detection, multiple in training

h: hidden state labels assigned to frames

x: features

$$\Psi(y, \mathbf{h}, \mathbf{x}; w) = \sum_{i=1}^{n} w_{h_i} . \varphi_{x_i} + \sum_{i=1}^{n} w_{y,h_i} + \sum_{(k,l)\in E} w_{y,h_k,h_l}$$

φ: feature vector, w: learned parameters from model

Experiments & Results

Dataset

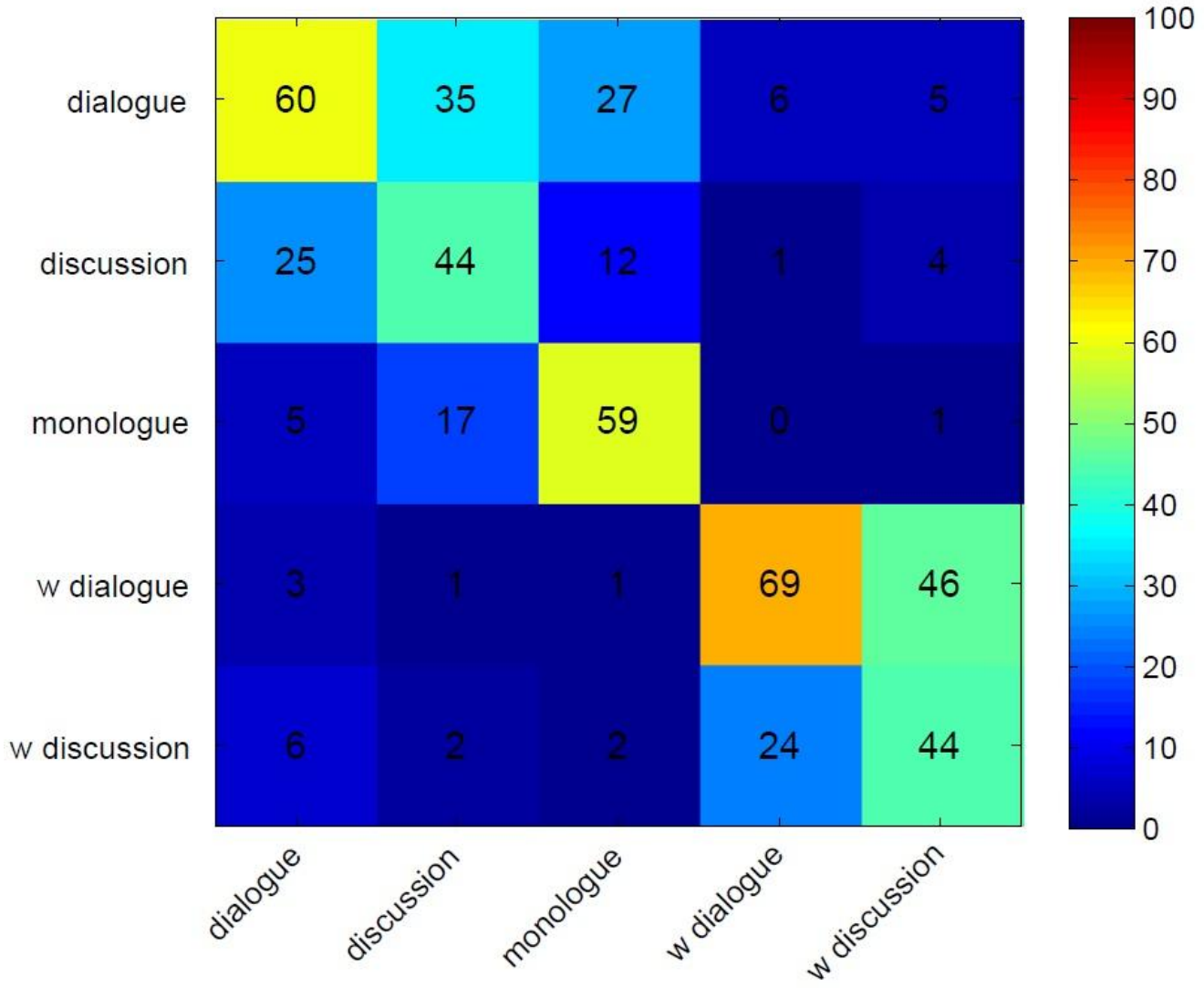
8 subjects (individuals)
3 days in amusement park
More than 42 hours of video

Social Interactions

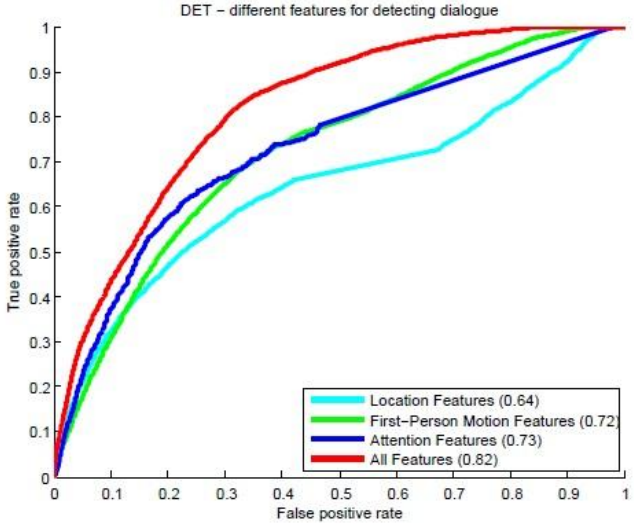
Dialogue, discussion, monologue, walk dialogue, walk discussion

Train on 5 subjects, test on 3

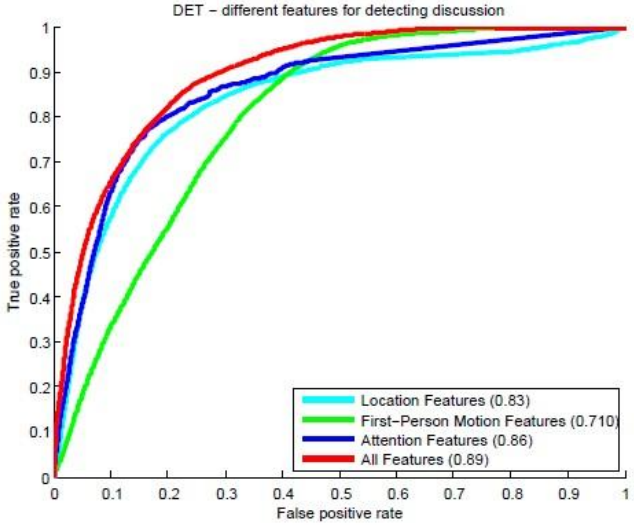
Experiments & Results



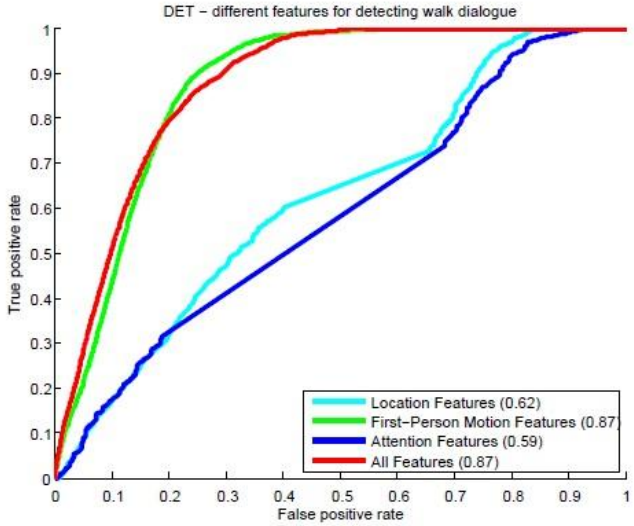
Experiments & Results

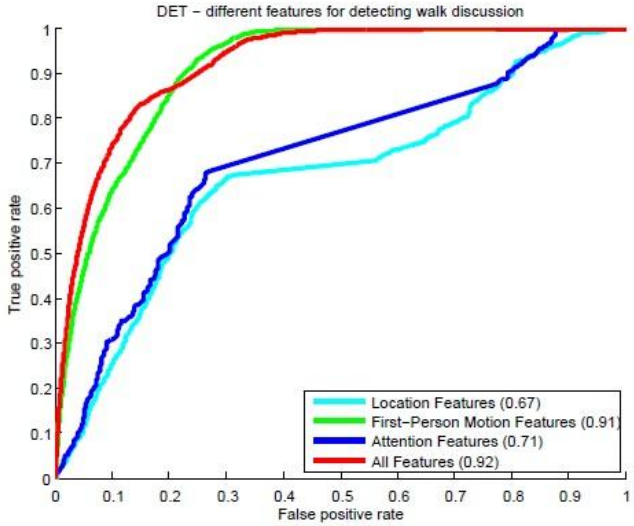


(a)



(b)





Thank You