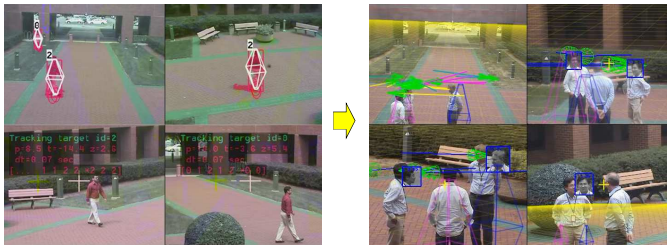
Tracking Gaze Direction from Far-Field Surveillance Cameras

Karthik Sankaranarayanan [1] **Ming-Ching Chang** [2] **Nils Krahnstoever** [2]

[2] Visualization and Computer Vision, GE Global Research Center, Niskayuna NY, USA

[1] Dept. of Computer Science and Engineering, Ohio State University, USA

Overview



Main idea: Real-time tracking of the gaze of multiple individuals using a network of far-field surveillance cameras.

- Fixed cameras performing site-wide tracking in unconstrained environments.
- PTZ cameras zooming to faces dynamically.
- Cooperative person tracking and head pose tracking:
  - Person detection, person tracking.
  - Face detection, head pose estimation, *asynchronous* gaze tracking in a centralized tracker.

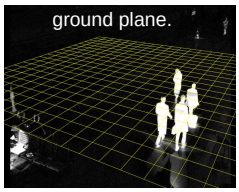
Applications:

- Improve person localization in crowded scenes.
- Real-time understanding of pose, gesture & social interaction.
- Find out what people are looking at (*e.g.*, security, retail apps).

Method

Video person tracking system:

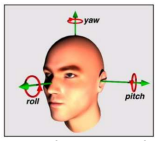
- Multiple calibrated static cameras.
- For each view a set of foreground person detection is form and then projected to a top-down, centralized Kalman filter tracker.



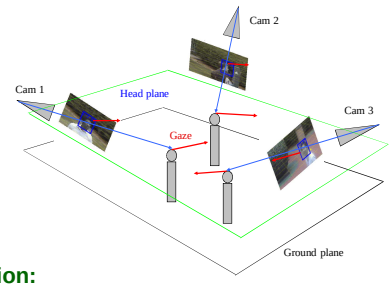
Pan Tilt Zoom (PTZ) camera control:

Real-time, continuously evolving schedule of PTZ camera actions obtained by optimizing toward a set of performance objectives:

- Estimate each PTZ camera's projection matrix based on predicted motion.
- Each schedule is assigned with a probability of maximizing the facial shots of all individuals.
- Quality of facial shot is governed by distance & angle from camera, and person tracking accuracy.



Egocentric gaze angles:
Yaw (left/right)
Pitch (chin up/down)



Face detection & projection:

- Localize face using **Pittsburgh Pattern Recognition** face detector.
- In each PTZ view, each detected face is projected to the *head plane* (1.8m height) to obtain the head position in 3D.

This project was supported by grant #2009-SQ-B9-K013 awarded by the National Institute of Justice, Office of Justice Programs, US Department of Justice.
The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the Department of Justice.

imagination at work

Gaze Analysis

Head pose detection to 3D gaze vector:

- Convert detected head pose in each camera's Cartesian coordinate $g_{im} = (x_{im}, y_{im}, z_{im})$ to a common centralized 3D world coordinate g_w using camera's rotation matrix $\mathbf{R}$.

$$\mathbf{g}_w = \mathbf{g}_{im} * (\mathbf{R}^{-1})^{\mathbf{T}}$$

- Convert g_w to each individual's egocentric angles in *yaw* α and *pitch* β.

$$\alpha = \arctan(\frac{x_g}{y_g}),$$

$$\beta = \arctan(\frac{z_g}{\sqrt{x_g^2 + y_g^2}})$$

Kalman filter gaze tracking:

Operates on individual's angular gaze coordinates (α, β).

Gaze state Θ :

$$\Theta = \begin{bmatrix} \alpha & \beta & \dot{\alpha} & \dot{\beta} \end{bmatrix}^{\mathbf{T}}$$

State transition model from time *k-1* to *k*:

$$\Theta_k = \mathbf{F} * \Theta_{k-1} + \mathbf{w}_{k-1}$$

State transition matrix $\mathbf{F}$ using constant velocity mode:

$\mathbf{w}$: Gaussian distributed process noise

$\mathbf{v}_k$: Gaussian distributed measurement noise

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Gaze measurement:

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} * \Theta_k + \mathbf{v}_k$$

Data association for gaze and person tracking:

- Association of multiple face detections with multiple person tracks, by minimizing a cost function reflecting two cues: *head location* and *gaze direction*.

Distance between head observation $\mathbf{h}_i$ and person track $\mathbf{t}_j$:

$$\eta(\mathbf{h}_i, \mathbf{t}_j) = \exp(-\frac{d(\mathbf{h}_i^{\mathbf{x}}, \mathbf{t}_j^{\mathbf{x}})}{\sigma_{\mathbf{x}}} - \frac{\lambda(\mathbf{h}_i^{\Theta}, \mathbf{t}_j^{\Theta})}{\sigma_{\Theta}})$$

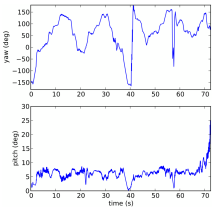
Euclidean distance between head & track's location

Geodesic distance between gaze & track's direction

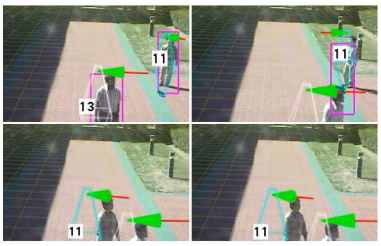
- **Hungarian algorithm** to minimize η to find the best association of detected faces and person tracks.

Experimental Results

Single person



Two people



Three people