…continued
Details and More Results

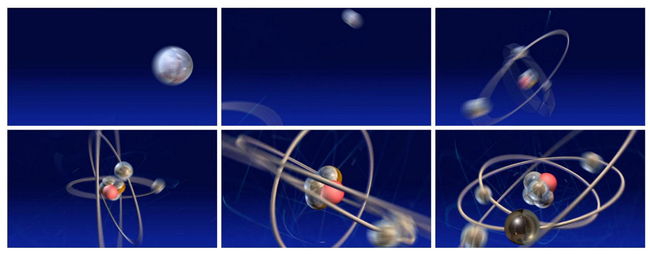
Video Analysis

Shot Boundaries Displaced Frame Difference (distance between two motion compensated frames) used as feature after normalization and filtering.

- 1981 correct detections, 2 miss, 8 false positives.

Special Sequences

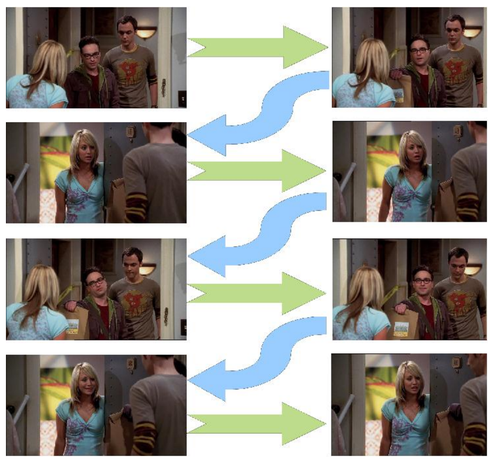
Audio-visual template matching techniques can be used in general. For TBBT, we use the dominant color descriptor to find artificial gradients.



- 19 correct detections, 0 miss, 0 false positives.

Alternating Shots

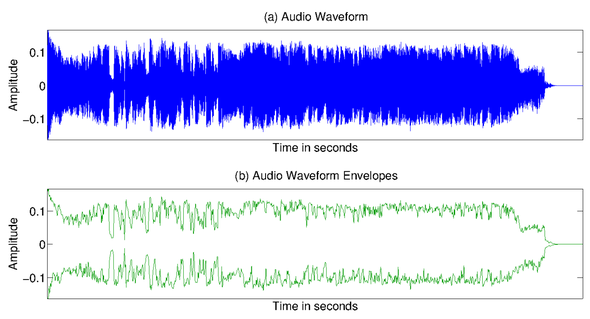
Same people occur in alternate shots! This can be used as a constraint on recognition in our model or to perform cumulative track scoring. Detected using Displaced Frame Difference on the first frames of each shot.



- 4.5% Equal Error Rate

Title Song

Detected and removed from analysis using audio template matching with audio waveform envelope.



- Up to ±20ms

Model Energy Functions

Minimize shot energy $E(i)$ to obtain optimal identities μ which match evidence (clothing and face) and satisfy constraints.

Clothing & Face: $E_C(i,j) = -\langle \mu_{ij}, c_{ij} \rangle$; $E_F(i,j) = -\langle \mu_{ij}, f_{ij} \rangle$

$E_{C/F}(i,j)$ is the clothing (face) energy term for shot i, track j and captures the similarity between the track evidence c_{ij}, f_{ij} with the optimization variable.

Regularization: $E_R(i,j) = \langle \mu_{ij}, \mu_{ij} \rangle$

$E_R(i,j)$ is a regularization term for μ_{ij}.

$$E_{FCR}(i) = \sum_j w_F E_F(i,j) + w_C E_C(i,j) + w_R E_R(i,j)$$

Presence: $\mathcal{P}^{(i)} = sigmoid(\sum_j \mu_{ij})$

Captures who is seen in the shot i.

Speaker Penalty: $E_S(i) = \langle (1 - \mathcal{P}^{(i)}),\ s_i \rangle$

Impose a penalty if the speaker s_i is not among the people present in the shot $\mathcal{P}^{(i)}$.

Uniqueness Constraint: $E_U(i) = \sum_{(j,k)\in\mathcal{T}_i} \langle \mu'_{ij}, \mu'_{ik} \rangle$

Two simultaneously occurring tracks should have different identities. μ' allows 2 unknown people to co-exist.

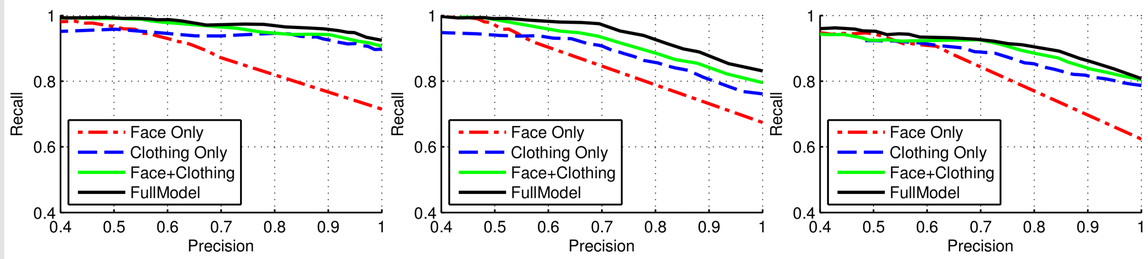
$$E(i) = E_{FCR}(i) + w_S E_S(i) + w_U E_U(i)$$

$$\mu^* = \arg\min_{\mu} E(i)$$

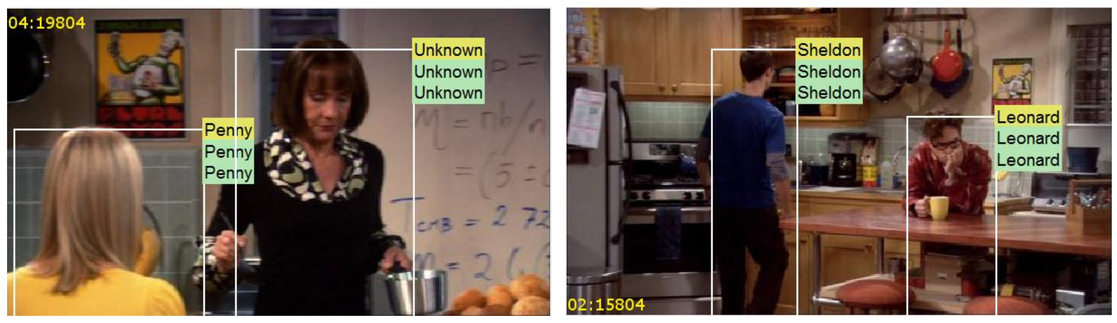
***Weights are seen to not have a large influence

More Results

Person Recognition – precision-recall curves, E1, E2, and E3



No face visible, characters correctly recognized…



Direct face result is wrong, clothing corrects

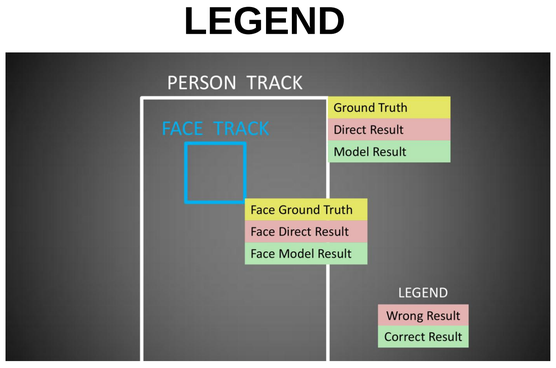
Direct clothing result is wrong, face corrects

Uniqueness constraint (can be tricky!)

The uniqueness constraint just enforces the two people are different. However, it cannot tell who is the right person! We see an example in which a correct result is flipped due to higher confidence in the other.

Sheldon speaks “Coming…”

LEGEND



Contact
{makarand.tapaswi, baeuml}@kit.edu
Computer Vision for Human Computer Interaction
http://cvhci.anthropomatik.kit.edu

Project page (tracks, ground truth, etc.)
http://cvhci.anthropomatik.kit.edu/~mtapaswi/projects/personid.html

Acknowledgment: This work was supported by the Quaero Programme, funded by OSEO; and BMBF contract no. 01ISO9052E.

KIT – University of the State of Baden-Wuerttemberg and
National Research Center of the Helmholtz Association

www.kit.edu