Gaze Estimation in Camera Networks

Peter Malmgren
Related Work: Parts based models

Their data: 

Our data: 

Source: Face Detection, Pose Estimation and Landmark Localization in the Wild
Related Work: Detection models

Tilt

Pan

Source: Pointing 2004 ICPR Workshop
Problem 1: Arbitrary Boundaries

These are now in different categories!
Problem 2: Similar appearance

- Looking left
- Looking right

Source: IIT (Italian Institute of Technology) Head Orientation Dataset
Our approach: Data-driven
How do we learn the groups?

40,000 faces ([IDIAP Head Pose Dataset](https://www.idiap.ch/dataset/))
Head-mounted tracker

80,000 non-faces
Training time

Error

Time (Days)
Determining head pose

Pan

Tilt

Yes

No
Results

Pan
-46.2° -15°
-6.8° -20°
-9.6° -10°

Tilt
9.3° -10°
-10.9° -10°
-26.8° -5°

Tested on 2000 images

<table>
<thead>
<tr>
<th>Median Tilt Error</th>
<th>Median Pan Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>29.8°</td>
<td>10.3°</td>
</tr>
</tbody>
</table>
Results from camera network
Combine camera predictions
Future work: Integrate into existing camera networks
Questions?
Modified Adaptive Boosting

<table>
<thead>
<tr>
<th>Adaptive Boosting</th>
<th>Modified Adaptive Boosting</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$</td>
<td>$H(x, \theta) = \text{sign}\left(\sum_{t=1}^{T} \Phi_t(\theta)\alpha_t h_t(x)\right)$</td>
</tr>
</tbody>
</table>

$H(x, \theta)$: Prediction at pan/tilt angle

$\Phi_t(\theta)$: \[
\begin{cases} 
1 & \text{if input is inside pan/tilt range} \\
0 & \text{if input is outside pan/tilt range} 
\end{cases}
\]
Local Binary Patterns

Threshold

Multiply

\[
\begin{align*}
7 & 1 & 12 \\
2 & 5 & 5 \\
5 & 3 & 0 \\
\end{align*}
\]

\[
\begin{align*}
1 & 0 & 1 \\
0 & 1 & 0 \\
1 & 0 & 0 \\
\end{align*}
\]

\[
\begin{align*}
1 & 2 & 4 \\
8 & 16 & 0 \\
32 & 64 & 128 \\
\end{align*}
\]

\[
\begin{align*}
1 & 0 & 4 \\
0 & 16 & 0 \\
32 & 0 & 0 \\
\end{align*}
\]

\[
LBP = 1 + 4 + 16 + 32 = 53
\]
Local Binary Patterns

Feature histogram (for every sub-window)
Unintended Feature Sharing
Evaluating Entire Gaze Space

Azimuth Error: 8.11832  Altitude Error: 13.06764  Distance: 15.38410
Spherical Coordinates

- Camera → world
- Head is origin (0,0,0)