
Peer reviewed version
Link to published version (if available): 10.1109/ICIP.2008.4711852

Link to publication record in Explore Bristol Research
PDF-document

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: http://www.bristol.ac.uk/pure/about/ebr-terms.html

Take down policy

Explore Bristol Research is a digital archive and the intention is that deposited content should not be removed. However, if you believe that this version of the work breaches copyright law please contact open-access@bristol.ac.uk and include the following information in your message:

• Your contact details
• Bibliographic details for the item, including a URL
• An outline of the nature of the complaint

On receipt of your message the Open Access Team will immediately investigate your claim, make an initial judgement of the validity of the claim and, where appropriate, withdraw the item in question from public view.
A Multi-Cue Bayesian State Estimator for Gaze Prediction in Open Signed Video

SJC Davies, D Agrafiotis, CN Canagarajah & DR Bull
Introduction

• Gaze patterns for sign language
• Open sign language
• Gaze prediction for open sign language
  • Facial Orientation
• Facial Orientation Tracker
• Results
• Conclusion
Sign Language Gaze Analysis

- Previous work has demonstrated the gaze pattern of sign language
Sign Language Gaze Analysis

- Previous work has demonstrated the gaze pattern of sign language
Open Sign Language

- Extend ‘head & shoulders’ to broadcast open sign language
Open Sign Language

• Extend ‘head & shoulders’ to broadcast open sign language
Open Sign Language

• Extend ‘head & shoulders’ to broadcast open sign language
Open Sign Language

- Extend ‘head & shoulders’ to broadcast open sign language
OSL Gaze Pattern

• Eye tracking study of people familiar with BSL
OSL Gaze Pattern

• 68% inset shot changes followed by saccade from signer to inset
• 95% of time observer looking at the signer, then signing is taking place
• 82% of time signing taking place, observers are looking at the signer
Gaze Track of Signing

Horiz. Gaze Location (px)

Frame Number

↑ Horizontally Aligned with Signer’s Face
↓ Horizontally Aligned with Programme Inset

Non-signing period
Gaze Prediction

• Can use these gaze patterns to code OSL video with variable quality
• Not feasible to eye track all broadcast material
• Need a method of predicting the gaze pattern of OSL
Orientation of the Head

• It is conventional in broadcast signing for the signer to watch a monitor to their side when not signing.
• This means their face is profile when not signing compared to frontal when signing.
• In 95% of ‘frontal’ frames the signer was signing.
• Only 0.15% of all frames were classified as ‘frontal’ and had no signing occurring.
Obtaining Facial Orientation

- Face detector based on Haar transforms (Viola & Jones)
Grid Based Likelihood Ratio Tracker

- Face detector not completely accurate
- Use 2 LR trackers in parallel - one tracking ‘frontal’ faces, the other ‘profile’ faces
- Trackers based on Bayesian principles
Tracker Process

A Priori Surface
Tracker Process

A Priori Surface

Frame Iterations resulting in LR surfaces
Tracker Process

A Priori Surface

UPDATE
Face detection results evoke localised increase in LR

Frame Iterations resulting in LR surfaces
Tracker Process

A Priori Surface

- **UPDATE**
  - Face detection results evoke localised increase in LR

- **PREDICTION**
  - Filter ‘spreads’ LR - small movement of faces likely in every direction

Frame Iterations resulting in LR surfaces
Tracker Process

A Priori Surface

**UPDATE**
Face detection results evoke localised increase in LR

**PREDICTION**
Filter ‘spreads’ LR - small movement of faces likely in every direction

Frame Iterations resulting in LR surfaces
Tracker Prediction

• For each frame have 4 different tracking direction ‘sheets’
Tracker Prediction

• For each frame have 4 different tracking direction ‘sheets’
• Represent movements in discrete directions
• Different amount per direction
Tracker Prediction

• For each frame have 4 different tracking direction ‘sheets’
• Represent movements in discrete directions
• Different amount per direction
Tracker Prediction

- For each frame have 4 different tracking direction ‘sheets’
- Represent movements in discrete directions
- Different amount per direction
Leaking

- Leak track directions between sheets
- Different factors according to directional pairs
Leaking

- Leak track directions between sheets
- Different factors according to directional pairs
Leaking

- Leak track directions between sheets
- Different factors according to directional pairs
Leaking

- Leak track directions between sheets
- Different factors according to directional pairs
Tracker Update

- Each of the detected faces causes likelihood ratio to be added to each of the sheets
- Gaussian shaped patch added equally to each sheet at detection locations

- To generate overall likelihood have to marginalise over all directional sheets by summing the components
Grid Based LR Tracker

- Showing:
  - frontal face detections
  - eye track locations
  - LR surface evolution
Orientation Tracker Results

- Want orientation of signer’s face
- Define $d$ as the difference between the max value of the frontal and profile trackers
- Possible to attain recall of 0.95 with $d$ at -5

\[
\text{signing?} = \begin{cases} 
  \text{yes} & \max(Fro) - \max(Pro) > d \\
  \text{no} & \text{otherwise}
\end{cases}
\]
Orientation Tracker Results

Clip Number vs. Accuracy (%)

Accuracy (%)

Clip Number

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30

0 10 20 30 40 50 60 70 80 90 100
Orientation Tracker Results

![Graph showing accuracy percentage for different clip numbers]
Orientation Tracker Results

![Bar chart showing accuracy of clip numbers 1 to 30]
Conclusions

- Observed gaze patterns for open sign language broadcast material
- Found that the orientation of the signer’s head is a good indicator of whether signing is taking place
- Developed a tracking system to detect presence of signing
- Can predict gaze location with an accuracy of 86%
Extension

• Have extended this to a multi-cue gaze predictor
  • Inset shot changes
  • Face locations
  • Signer’s facial orientation

• Generates probability surface of fixations
• Attain a gaze prediction accuracy of 90%, with a 20% ROI frame area

To be published Jan/Feb 2009 :: IEEE Transactions on Multimedia
Acknowledgements

• This work was funded by the British Broadcasting Corporation (BBC)