Evaluating the Robustness of an Appearance-based Gaze Estimation Method for Multimodal Interfaces

Nanxiang Li and Carlos Busso

Multimodal Signal Processing (MSP) Laboratory
Erik Jonsson School of Engineering & Computer Science
University of Texas at Dallas, Richardson, Texas, 75083, U.S.A.

ICMI 2013
Dec 9 - 13, Sydney, Australia

Motivation

Advantages of Gaze-aware Multimodal Interfaces

- Natural and fast
- Related to the users' cognitive state

Challenges

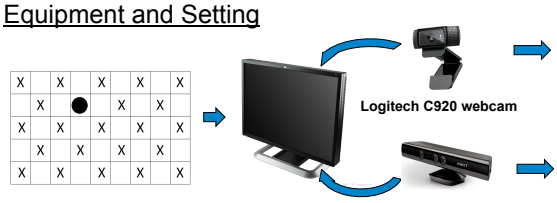
- Tedious calibration process
- Sensitive against variability in real applications

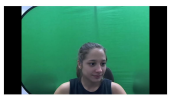
Aim of this Study

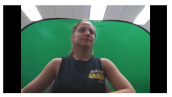
- Collect a multimodal corpus for gaze
- Evaluate an appearance-based method for gaze tracking based on PCA
- Evaluate the robustness of the proposed method against
  - Head movement
  - Calibration pattern
  - User distance to the monitor
  - Individual differences
  - Different sessions

MSP-GAZE Database

Equipment and Setting




(a) Webcam Image


(b) Kinect Image

Monitor projects a target point randomly chosen from the 23 highlighted grids as both webcam and Kinect record the subject behavior

Data Collection Protocol

- 46 subjects (gender balanced)
- Diverse ethnic representation
  - Caucasian – 16 subjects
  - Asian – 10 subjects
  - Indian – 10 subjects
  - Hispanic – 10 subjects
- Two sessions on different days
- 14 recordings per session (training – 12, testing – 2)

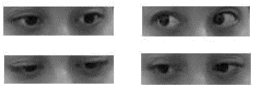
Recording	Head Movement	Distance	Pattern
1	Yes	User-defined	Testing
2	Yes	User-defined	Training
3	Yes	Near	Training
4	Yes	Medium	Training
5	Yes	Medium	Training
6	Yes	Far	Training
7	Yes	Far	Training
8	No	User-defined	Testing
9	No	User-defined	Training
10	No	Near	Training
11	No	Medium	Training
12	No	Medium	Training
13	No	Far	Training
14	No	Far	Training

Recordings conditions for each session (Near - 0.4 meter, Medium - 0.5 meters, Far - 0.6 meter)

Appearance Based Gaze Estimation

Proposed Approach

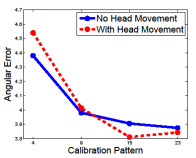
- We use patch with both eyes
  - Reliable for eye detection
  - Robust against head motion
- Eye pair image extraction using cascade object detector


Extracted eye pair images

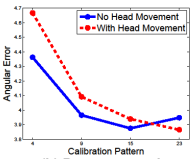
- We estimate eigenvectors from the covariant matrix of the training images
  - Select 30 principle components
- Build linear regression model
  - Independent variables – projections into the eigenspace
  - Dependent variables – the x, y coordinates on the screen

Experimental Results

- Performance Metrics
  - Correlation (ρ_x, ρ_y), Angular error (Θ_{error})
- Effect of calibration pattern (~15 grids)



(a) Within session



(b) Between sessions

- Subject Dependent Results

	Without head motion			With head motion		
Distance	ρ_x	ρ_y	Θ_{error}	ρ_x	ρ_y	Θ_{error}
Near	0.90	0.85	4.7	0.91	0.84	4.5
Medium	0.89	0.84	3.8	0.91	0.83	3.9
Far	0.88	0.83	3.5	0.90	0.83	3.4
User-Defined	0.89	0.82	3.9	0.88	0.82	3.9

- Subject Independent Results

	Without head motion			With head motion		
Distance	ρ_x	ρ_y	Θ_{error}	ρ_x	ρ_y	Θ_{error}
Near	0.85	0.76	7.0	0.87	0.75	6.8
Medium	0.86	0.75	6.0	0.85	0.74	5.9
Far	0.85	0.68	5.3	0.85	0.73	5.2
User-Defined	0.85	0.78	5.9	0.86	0.70	6.0

Discussion

- Consistent performance for subject-dependent models
  - Head motion
  - User-interface distance
  - Sessions
- Performance on the subject-independent model slightly decreases
  - It does not need calibration

Future Directions

- Use Kinect and Webcam images
- Improve performance under subject-independent conditions
  - Find subjects with similar eye appearance to PCA
  - Apply whitening transformation on the training image covariance matrix
- Implement the proposed method in mobile devices

Acknowledgements: NSF and Samsung