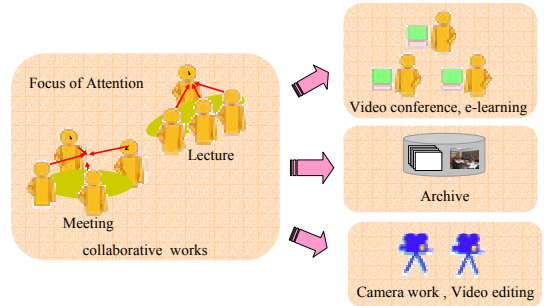


Estimation of Focus of Attention of Multiple People for Video Conferencing

NAIST Robotics Lab.
D2 Kentaro Takemura

Motivation

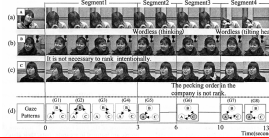


Related Works

1. Focus of Attention in Meetings, [R. Stiefelhagen(CMU& Karlsruhe), 2002]



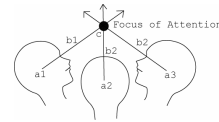
2. Video Cut Editing Rule based on Participants' Gaze [Y. Takemae(NTT), 2003]



Discrete measurement, Static environment

Approach

- If people share attentive object, all of their gaze directions should intersect at one point.



- Using Head Orientation

Head orientation and gaze pointed at the same direction in 87%
[R. Stiefelhagen, 2002]

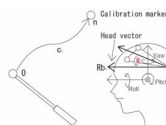
Definition of face normal vector from marker positions

Calibration

1. The person stares the calibration marker for certain length of time.
2. The error function which is the summation of the Euclidean distances from the head vector to the marker in all frames is calculated with regard to a pitch angle θ .

$$\arg \min_{\theta} \left[\sum_{i=0}^n \left\| \frac{(\mathbf{c}_i - \mathbf{a}_i) \cdot \mathbf{R}\mathbf{b}}{\|\mathbf{R}\mathbf{b}\|^2} \mathbf{R}\mathbf{b}_i + \mathbf{a}_i - \mathbf{c}_i \right\|^2 \right]$$

3. Optimal θ which minimizes the error function is searched by the steepest descent method.



Estimation of Focus of Attention

Multiple vectors in 3D space don't intersect at one point in general.

Assumption: Everyone stares an object at the same time

$$\varepsilon_i = \left\| \frac{(\mathbf{c}_i - \mathbf{a}_i) \cdot \mathbf{b}}{\|\mathbf{b}\|^2} \mathbf{b}_i + \mathbf{a}_i - \mathbf{c}_i \right\|^2$$

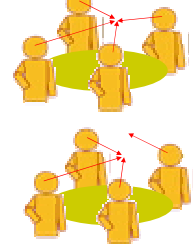
the least square method.

$$\hat{\mathbf{c}} = \arg \min_{\mathbf{c}} \sum_{i=1}^n (\varepsilon_i)^2$$

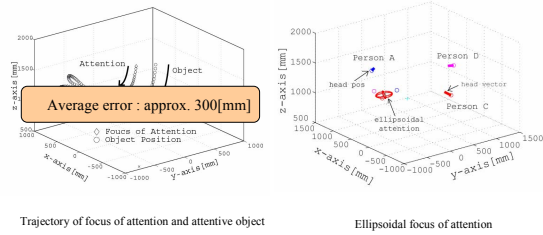
M-estimator

$$\hat{\mathbf{c}} = \arg \min_{\mathbf{c}} \sum_{i=1}^n \left\{ \sqrt{\omega(\varepsilon_i^{(k-1)})} \cdot \varepsilon_i \right\}^2$$

initial value: $\mathbf{c} = 0, \varepsilon_i^{(0)} = \varepsilon_i$
 $\omega(\cdot)$ = the Weibull function
 $\omega(\varepsilon) = \exp(-(\varepsilon / \text{const})^2)$
 ε_i = the Euclidean distance from the attentive point to face vector



Estimation of Focus of Attention

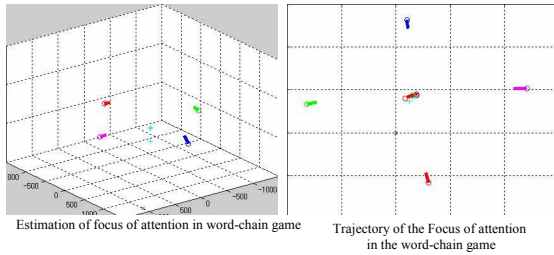


Experiment: Focus of Attention in "word-chain" game

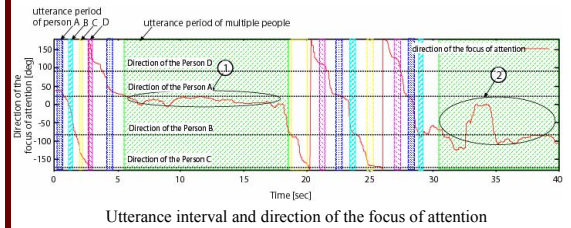
- Word-chain game (SIRITORI in Japanese)
 - For example
SIRITORI → RINGO → GOMA → MARI →



Experiment: Focus of Attention in "word-chain" game

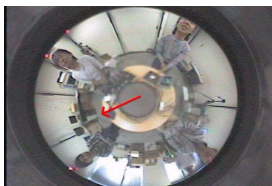


Experimental Results: Relation between focus of attention and utterance



- Stability regions correspond with angles of person.
- It is possible to estimate the focus of attention without utterance.

Video cut editing based on focus of attention

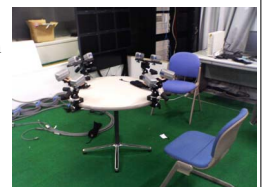


Conclusion

- A method to estimate the focus of attention of multiple people was proposed.
- I applied my method to the videoconferencing.

Future works

- Improve the accuracy of position using gaze information
- Online measurement system





Thank you for your attention
