# Estimating Gaze Points from Facial landmarks by a Remote Spherical Camera

Shigang Li
Graduate School of Information Sciences
Hiroshima City University
Japan

Norika Fujii
Faculty of Information Sciences
Hiroshima City University
Japan

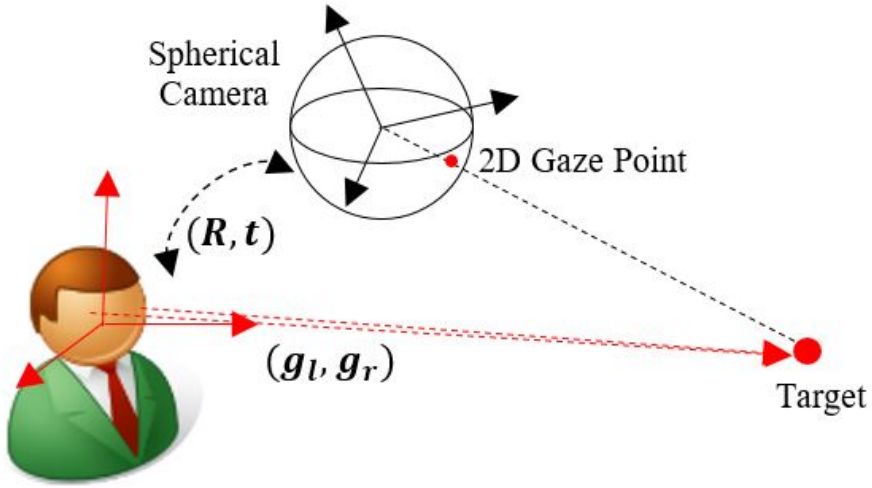
# Goal

- People sitting around a table are observed by a spherical camera.
- Because users' faces and their gaze points can be observed simultaneously, the gaze points can be estimated directly from a single spherical image.

- **Objective**: estimating gaze points from facial landmarks by a remote spherical camera

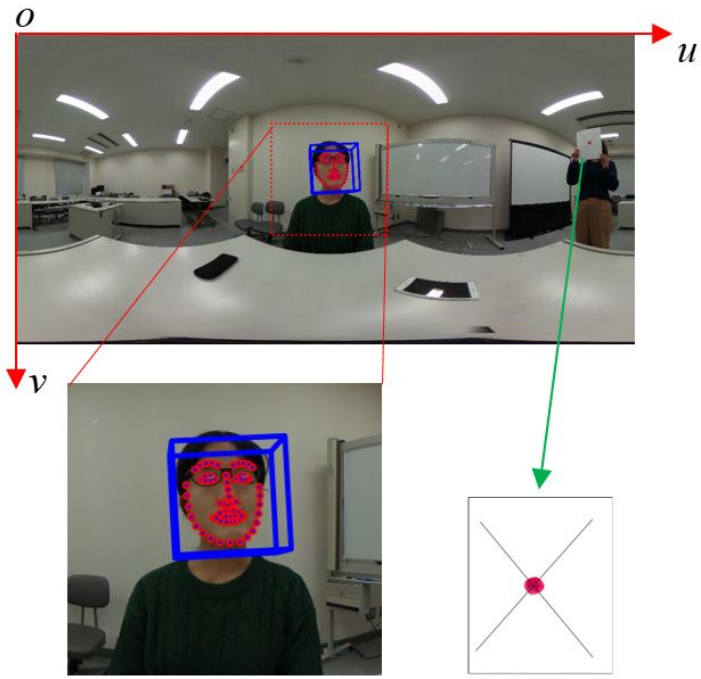
# Geometry

- Geometrical relationship among a user, a spherical camera and a gaze target

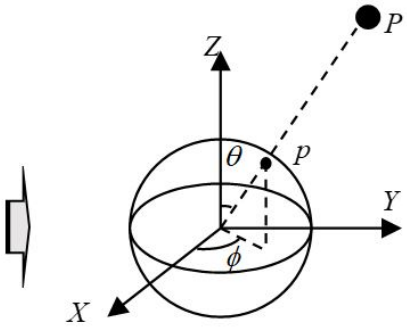



- Accurate estimation of gaze vectors and head pose by geometrical computation is not always easy in practice
- **Approach**: use a neural network to learn the geometry implicitly

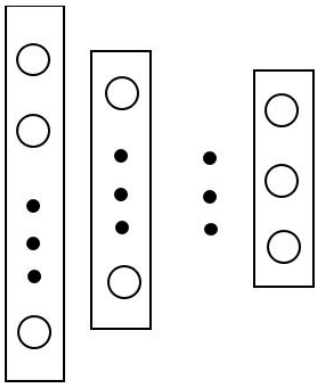
# Proposed method




Extracting facial landmarks & gaze point




Representation on a unit sphere

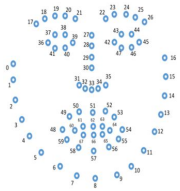


Processing by neural network

# Four neural networks

TABLE I
FOUR NEURAL NETWORKS USED FOR LEARNING GAZE POINTS

| | Eye Landmarks | Eye & Face Landmarks |
|---|---|---|
| Polar coordinates | EP-NN | EFCP-NN |
| Orthogonal coordinates | EO-NN | EFCO-NN |



Face Landmarks

Eye Landmarks

- Two representations of facial landmarks: *polar coordinates* **or** *orthogonal coordinates*
- Input landmarks to NN: *eye landmarks* **or** *Eye and face boundary landmarks*

# Error measurement & data collection

- Error measurement of gaze point:
  - angle between the vector, $p_e = (x_e, y_e, z_e)$, of estimated gaze points and that , $p_t = (x_t, y_t, z_t)$, of the truth on a sphere

$$e_g = arccos \frac{p_e \cdot p_t}{|p_e||p_t|}$$

- Data collection:
  - 4 subjects, 120 images per subject
  - data augmentation by rotating around the vertical axis randomly 72 times and adding random noise within 1 pixel to the coordinates of facial landmarks and gaze targets 3 times
  - For each subject, 120x72x3 (25,920) data are collected.
- Divide the dataset by 8:2 for training set and test set

# Experimental results

TABLE III
USING EYE LANDMARKS AND GAZE POINTS REPRESENTED AS ORTHOGONAL COORDINATES AND POLAR COORDINATES, RESPECTIVELY, FOR SINGLE-SUBJECT

| | EO-NN | | EP-NN | |
|---|---|---|---|---|
| | Mean of $e_g$ $(deg.)$ | Standard dev. of $e_g$ $(deg.)$ | Mean of $e_g$ $(deg.)$ | Standard dev. of $e_g$ $(deg.)$ |
| Subject 1 | 1.47 | 0.97 | 9.92 | 14.61 |
| Subject 2 | 1.44 | 0.99 | 10.66 | 15.87 |
| Subject 3 | 0.93 | 0.51 | 7.23 | 10.29 |
| Subject 4 | 1.48 | 0.97 | 12.79 | 14.98 |
| Average | 1.33 | 0.86 | 10.15 | 13.94 |

TABLE IV
EFCO-NN USING EYE & FACE CONTOUR LANDMARKS AND GAZE POINTS REPRESENTED AS ORTHOGONAL COORDINATES AND POLAR COORDINATES, RESPECTIVELY, FOR SINGLE-SUBJECT

| | EFCO-NN | | EFCP-NN | |
|---|---|---|---|---|
| | Mean of $e_g$ $(deg.)$ | Standard dev. of $e_g$ $(deg.)$ | Mean of $e_g$ $(deg.)$ | Standard dev. of $e_g$ $(deg.)$ |
| Subject 1 | 1.68 | 1.20 | 8.09 | 15.10 |
| Subject 2 | 1.51 | 0.75 | 11.71 | 16.08 |
| Subject 3 | 1.21 | 0.60 | 5.19 | 10.20 |
| Subject 4 | 0.86 | 0.39 | 6.68 | 11.77 |
| Average | 1.31 | 0.74 | 7.92 | 13.29 |

TABLE V
USING EYE LANDMARKS AND GAZE POINTS REPRESENTED AS ORTHOGONAL COORDINATES AND POLAR COORDINATES, RESPECTIVELY, FOR CROSS-SUBJECT

| | EO-NN | | EP-NN | |
|---|---|---|---|---|
| | Mean of $e_g$ $(deg.)$ | Standard dev. of $e_g$ $(deg.)$ | Mean of $e_g$ $(deg.)$ | Standard dev. of $e_g$ $(deg.)$ |
| Subject 1 | 7.22 | 4.51 | 15.32 | 13.11 |
| Subject 2 | 6.08 | 3.77 | 17.41 | 15.09 |
| Subject 3 | 8.05 | 4.27 | 14.01 | 7.13 |
| Subject 4 | 5.39 | 3.44 | 15.59 | 10.07 |
| Average | 6.68 | 4.00 | 15.58 | 11.35 |

TABLE VI
EFCO-NN USING EYE & FACE CONTOUR LANDMARKS AND GAZE POINTS REPRESENTED AS ORTHOGONAL COORDINATES AND POLAR COORDINATES, RESPECTIVELY, FOR CROSS-SUBJECT

| | EFCO-NN | | EFCP-NN | |
|---|---|---|---|---|
| | Mean of $e_g$ $(deg.)$ | Standard dev. of $e_g$ $(deg.)$ | Mean of $e_g$ $(deg.)$ | Standard dev. of $e_g$ $(deg.)$ |
| Subject 1 | 4.39 | 2.81 | 10.72 | 10.05 |
| Subject 2 | 3.70 | 2.14 | 9.32 | 9.45 |
| Subject 3 | 5.13 | 3.50 | 12.45 | 15.49 |
| Subject 4 | 2.99 | 1.72 | 11.02 | 8.71 |
| Average | 4.05 | 2.54 | 10.88 | 10.93 |

- NN model of single-subject outperforms that of cross-subject.
- Orthogonal coordinate representation of landmarks achieves much better performance than the polar coordinate representation.
- Combining face contour landmarks with eye landmarks can improve the accuracy of gaze point estimation a little.

# Conclusions

- Propose a method of estimating a gaze point by a remote spherical camera from facial landmarks.
- The orthogonal coordinate representation of facial landmarks results in a higher accuracy than the polar coordinate representation.

- Building a larger dataset and trying different kind of models, such SVM (support vector machine), to verify the effectiveness of the proposed method is our future work.