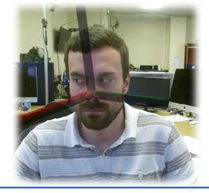
Imperial College London

PERSONAL ROBOTICS LABORATORY

www.imperial.ac.uk/PersonalRobotics

RT-GENE: Real-Time Eye Gaze Estimation in Natural Environments

Tobias Fischer Hyung Jin Chang Yiannis Demiris



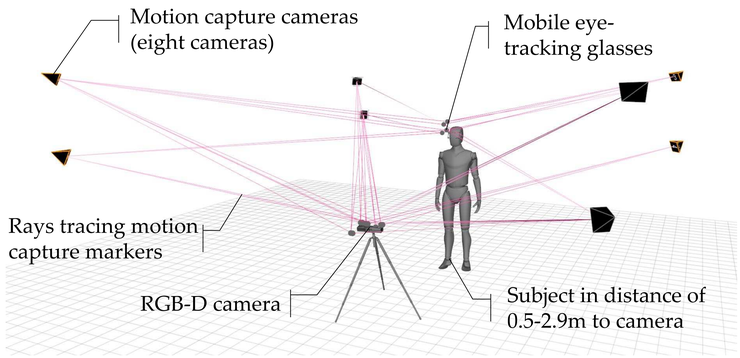
ECCV 2018

Target problem:

- We consider the problem of **robust eye gaze estimation in natural environments** with large camera-to-subject distances and high variations in head pose and eye gaze angles.

Approach & Contribution:

- Recording of **novel dataset** of varied gaze and head pose images in a natural environment. Ground truth **head pose using a motion capture system and eye gaze using mobile eyetracking glasses**.



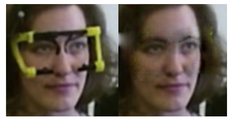
- **Removing the obtrusiveness of the eyetracking glasses** using GAN-based semantic image inpainting [58].

$$(1)\ \hat{\mathbf{z}} = \arg\min_{\mathbf{z}} \left(\lambda\, \mathcal{L}_{\text{perception}}(\mathbf{z}) + \mathcal{L}_{\text{context}}(\mathbf{z}|\mathbf{M}) \right),$$

$$\text{where}\ \ \mathcal{L}_{\text{perception}}(\mathbf{z}) = \left[D\left(G(\mathbf{z})\right) - 1 \right]^2$$

$$\mathcal{L}_{\text{context}}(\mathbf{z}|\mathbf{M},\mathbf{x}) = |\mathbf{M}' \odot \mathbf{x} - \mathbf{M}' \odot G(\mathbf{z})|$$

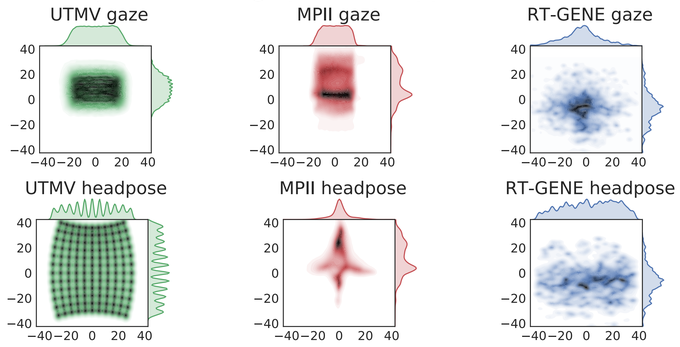
$$(2)\ \ \mathbf{x}_{\text{inpainted}} = \mathbf{M}' \odot \mathbf{x} + \mathbf{M} \odot G(\hat{\mathbf{z}})$$



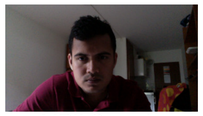
- New **real-time appearance based CNNs** with increased capacity to cope with diverse images in the dataset. Using **VGG-16** for feature extraction and **ensemble scheme** for increased robustness.

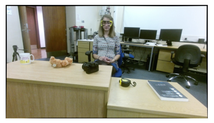
Dataset details:

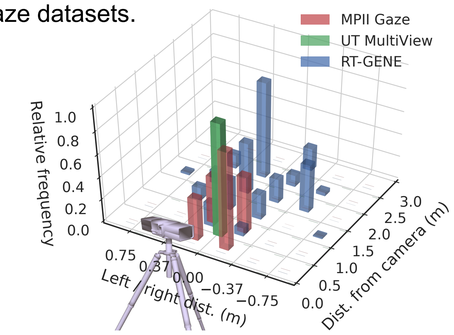
- 17 subjects with **over 100.000 training images** (RGB-D from Kinect v2). There are **high pose and appearance variations**.



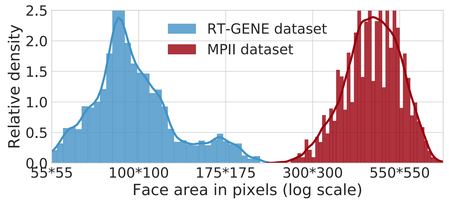
- The **camera-subject distances are considerably larger** compared to previous eye gaze datasets.


MPII Sample


RT-GENE Sample



- This leads to **challenging, low-resolution face images**.

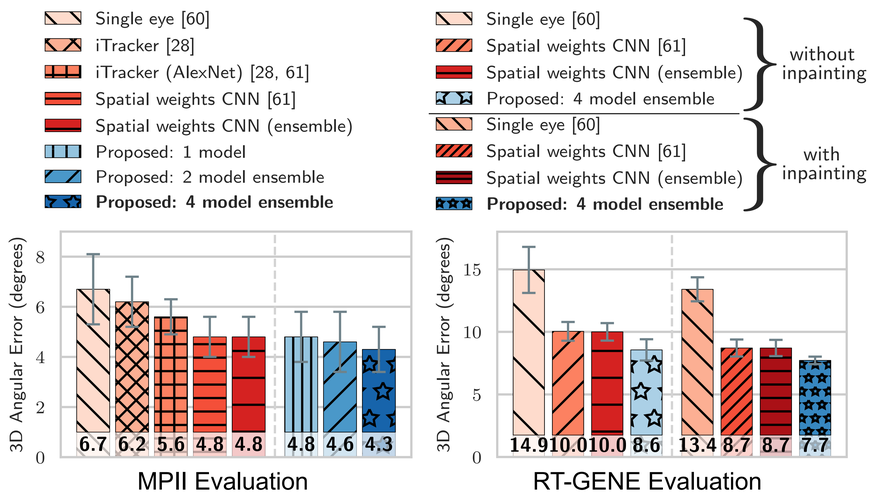


Experimental results:

- Semantic inpainting leads to increased face detection rate, and decreased landmark error. **No statistical difference between the inpainted images and natural images (no eyetracking glasses)**.

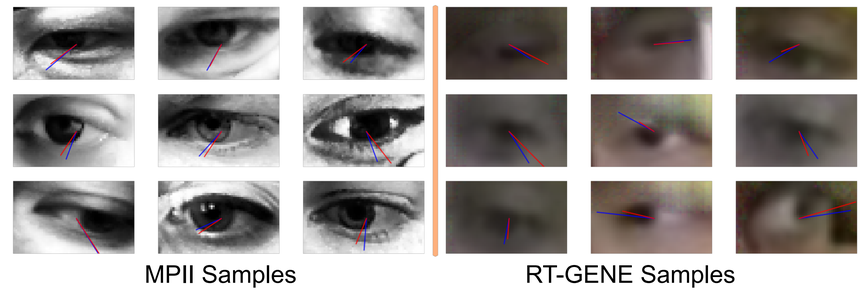
Landmark detection method	Face detection rate (%)			Landmark error (pixel)		
	Original	Uniformly filled	Inpainted	Original	Uniformly filled	Inpainted
CLNF [3]	54.6±24.7	75.4±20.9	**87.7±15.6**	6.0±2.4	5.6±2.3	**5.3±1.8**
CLNF in-the-wild [3]	54.6±24.7	75.4±20.9	**87.7±15.6**	5.8±2.3	5.3±1.8	**5.2±1.6**
ERT [25]	36.7±25.3	59.7±23.0	**84.1±17.9**	6.6±2.3	5.8±1.7	**5.1±1.3**

- Evaluation on MPII Gaze [60], UT Multi-view [53], and our newly proposed RT-GENE dataset. Our method **outperforms the state-of-the-art in all experiments by 10.4-13.6%.**

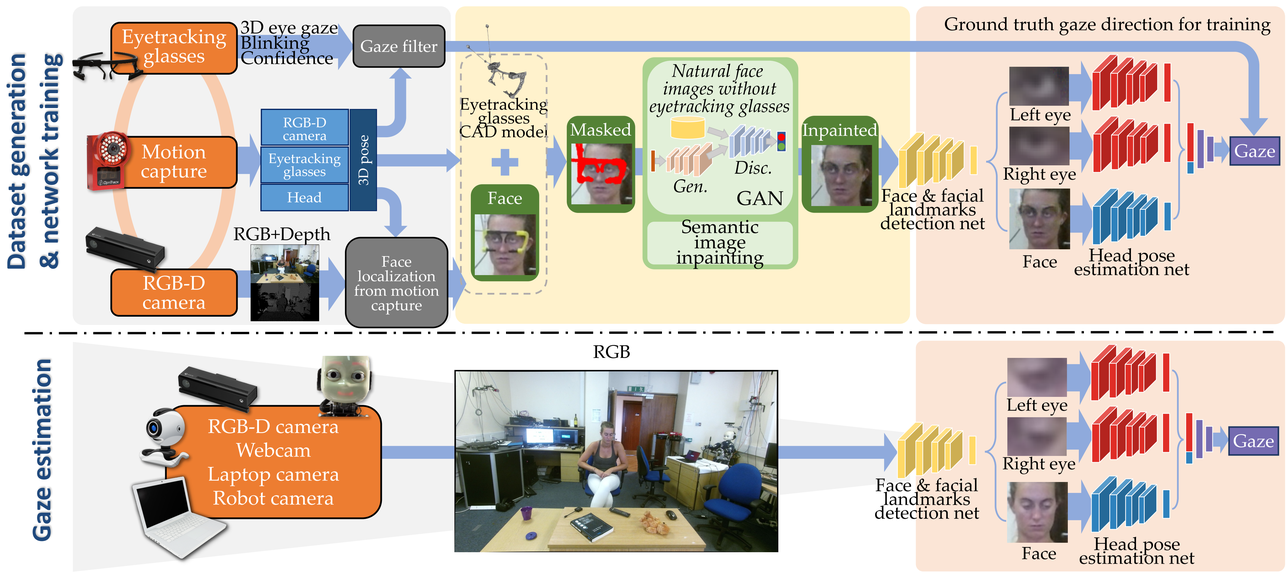


- **Cross dataset evaluation** (train on RT-GENE dataset, test on MPII dataset) leads to **22.4% improvement** over state-of-the-art [55].

- Qualitative results (red: sample estimate, blue: ground truth)



System Overview



Paper, Dataset & Code Download:

www.imperial.ac.uk/PersonalRobotics

www.tobiasfischer.info

Key References:

[3] Baltrusaitis et al. ICCVW2013: Constrained local neural fields for robust facial landmark detection
[25] Kazemi and Sullivan CVPR2014: One millisecond face alignment
[53] Sugano et al. CVPR2014: Learning-by-Synthesis for Appearance-based 3D Gaze Estimation (UTMV)
[55] Wood et al. ETRA2016: Appearance-based gaze estimator from one million synthesised images
[58] Yeh et al. CVPR2017: Semantic image inpainting with deep generative models
[60] Zhang et al. CVPR2015: Appearance-Based Gaze Estimation in the Wild (MPII dataset)
[61] Zhang et al. CVPRW2017: It's Written All Over Your Face

Acknowledgments:

This work was supported in part by the Samsung Global Research Outreach program, and in part by the EU Horizon 2020 Project PAL (643783-RIA).