C

MPIIFaceGaze: It's Written All Over Your Face: Full-Face Appearance-Based Gaze Estimation

Abstract

Eye gaze is an important non-verbal cue and promising for automatic affect analysis. We propose an appearance-based method that, in contrast to a long line of work in computer vision, only takes the full face image as input. Our method encodes the face image using a convolutional neural network with spatial weights applied on the feature maps to flexibly suppress or enhance information in different facial regions. Through extensive evaluation, we show that our full-face method significantly outperforms the state of the art for both 2D and 3D gaze estimation, achieving improvements of up to 14.3% on MPIIGaze and 27.7% on EYEDIAP for person-independent 3D gaze estimation. We further show that this improvement is consistent across different illumination conditions and gaze directions and particularly pronounced for the most challenging extreme head poses.

Download (899 Mb)

You can also download the normalized data from here, which includes the normalized face images as 448*448 pixels size and 2D gaze angle vectors. Please note the images were saved in Maltab, and there are following pre-process to make these data can be accepted by Caffe:

new_image = original_image(:,:,[3 2 1]);
new_image = flip(new_image, 2);
new_image = imrotate(new_image, 90);
These steps switched the RGB channels and flipped the image horizontally and rotate is by 90 degrees.

Contact: Andreas Bulling, andreas.bulling@vis.uni-stuttgart.de

The data is only to be used for non-commercial scientific purposes. If you use this dataset in a scientific publication, please cite the following paper:
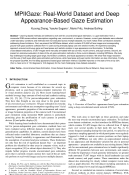
MPIIGaze: Real-World Dataset and Deep Appearance-Based Gaze Estimation

Xucong Zhang, Yusuke Sugano, Mario Fritz, Andreas Bulling

IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 41 (1), pp. 162-175, 2019.

Abstract　　　Links　　　BibTeX　　　Project

It's Written All Over Your Face: Full-Face Appearance-Based Gaze Estimation

Xucong Zhang, Yusuke Sugano, Mario Fritz, Andreas Bulling

Abstract   Links   BibTeX   Project

Content

We present the MPIIFaceGaze dataset which is based on the MPIIGaze dataset, with the additional human facial landmark annotation and the face regions are available. We added additional facial landmark and pupil center annotations for 37,667 face images. Facial landmarks annotations were conducted in a semi-automatic manner as running facial landmark detection method first and then checking by two human annotators. The pupil centers were annotated by two human annotators from scratch. For sake of privacy, we only released the face region and blocked the background in images.

Data

There are 15 participants and corresponding folder from p00 to p14. Each folder includes the images inside the different day folder.

Annotation

There are pxx.txt file in each participant folder. which saves the information:
Dimension 1: image file path and name.
Dimension 2~3: Gaze location on the screen coordinate in pixels, the actual screen size can be found in the "Calibration" folder.
Dimension 4~15: (x,y) position for the six facial landmarks, which are four eye corners and two mouth corners.
Dimension 16~21: The estimated 3D head pose in the camera coordinate system based on 6 points-based 3D face model, rotation and translation: we implement the same 6 points-based 3D face model in [1], which includes the four eye corners and two mouth corners.
Dimension 22~24 (fc): Face center in the camera coordinate system, which is averaged 3D location of the 6 focal landmarks face model. Not it is slightly different with the head translation due to the different centers of head and face.
Dimension 25~27 (gt): The 3D gaze target location in the camera coordinate system. The gaze direction can be calculated as gt - fc.
Dimension 28: Which eye (left or right) is used for the evaluation subset in [2].

Calibration

There is the "Calibration" folder for each participant, which contains
(1) Camera.mat: the intrinsic parameter of the laptop camera. "cameraMatrix": the projection matrix of the camera. "distCoeffs": camera distortion coefficients. "retval": root mean square (RMS) re-projection error. "rvecs": the rotation vectors. "tvecs": the translation vectors.
(2) monitorPose.mat: the position of image plane in camera coordinate. "rvecs": the rotation vectors. "tvecs": the translation vectors.
(3) screenSize.mat: the laptop screen size. "height_pixel": the screen height in pixel. "width_pixel": the screen width in pixel. "height_mm": the screen height in millimeter. "width_mm": the screen width in millimeter.

Full-face Patch Gaze Estimation

We proposed a full-face patch gaze estimation method with novel spatial weights CNN. To efficiently use the information from full-face images, we propose to use additional layers that learn spatial weights for the activation of the last convolutional layer. we used the concept of the three 1 × 1 convolutional layers plus rectified linear unit layers to generate a single heatmap encoding the importance across the whole face image. We then performed an element-wise multiplication of this weight map with the feature map of the previous convolutional layer.

The source code is available here, the pre-trained caffe model is available here. Please note this is still in beta version, and bug report and suggestion are welcome.
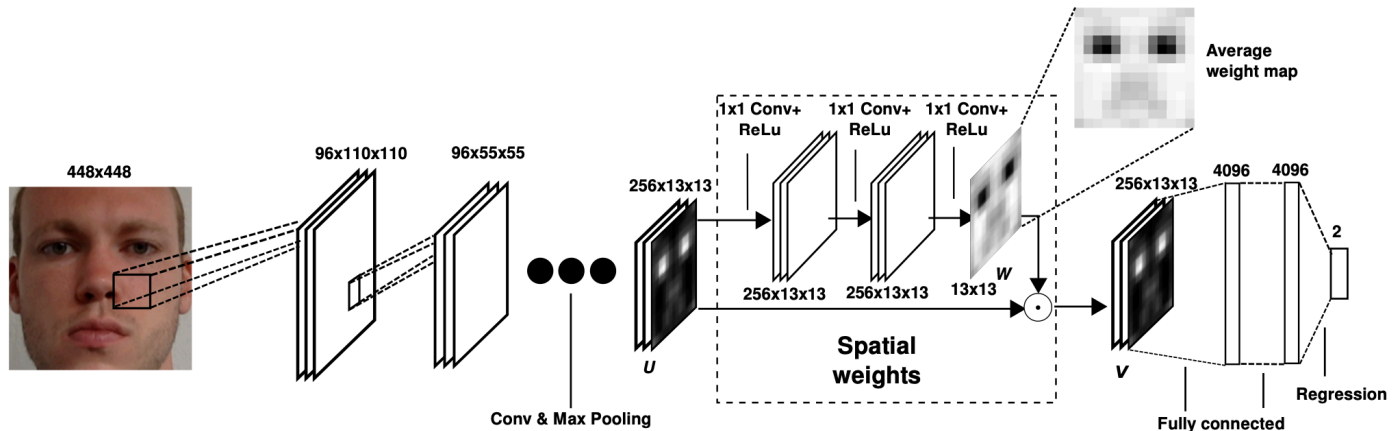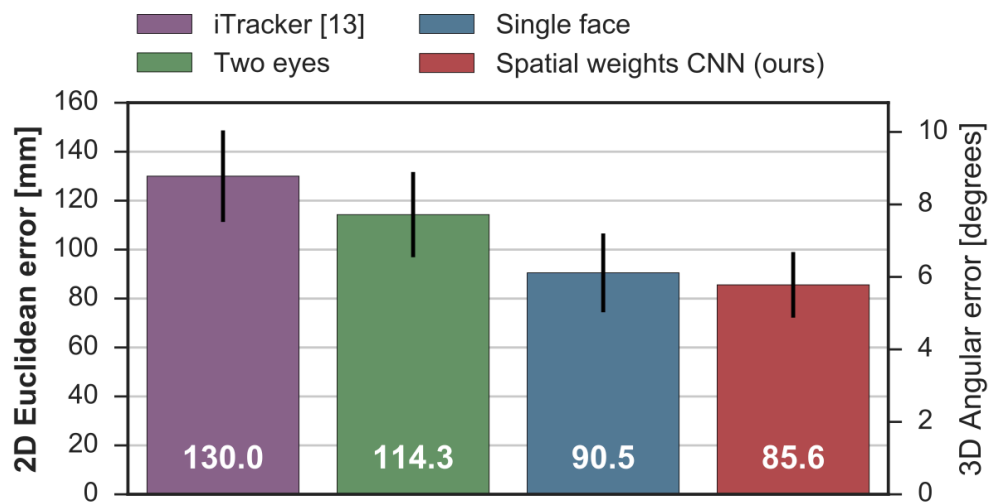

Evaluation

2D Gaze Estimation

We did the leave-one-person-out cross-validation for our MPIIFaceGaze dataset, and 5-fold cross-validataion for EYEDIAP dataset [3]. The results are shown in the below figures. The left axis shows the Euclidean error between estimated and ground-truth gaze positions in the screen coordinate system in millimetres. The right axis shows the corresponding angular error that was approximately calculated from the camera and monitor calibration information provided by the dataset and the same reference position for the 3D gaze estimation task.

For MPIIGaze dataset (below), the proposed spatial weights network achieved a statistically significant 7.2% performance improvement (paired t-test: $p < 0.01$) over the second best single face model.
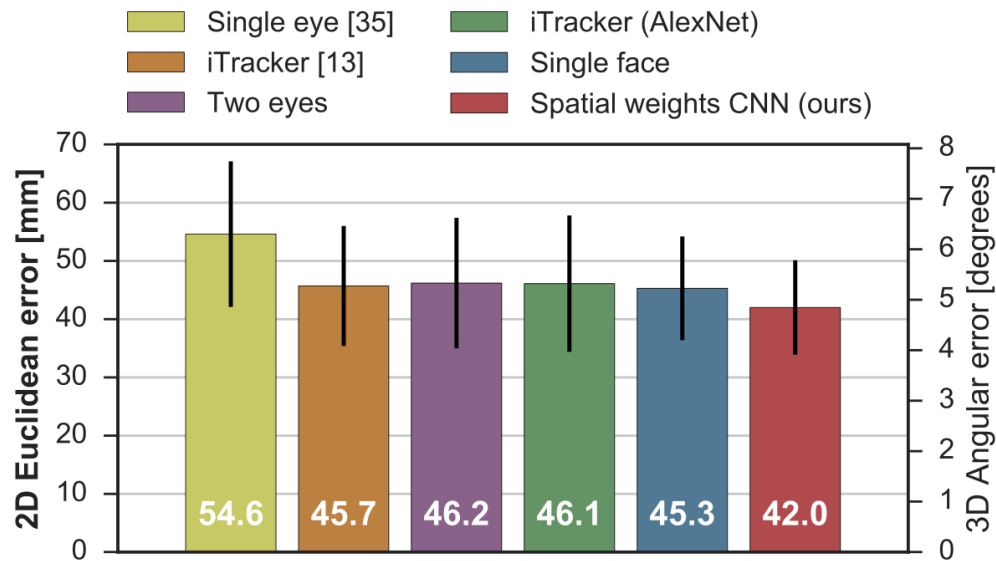
These findings are in general mirrored for the EYEDIAP dataset (below), while the overall performance is worse most likely due to the lower resolution and the limited amount of training images.
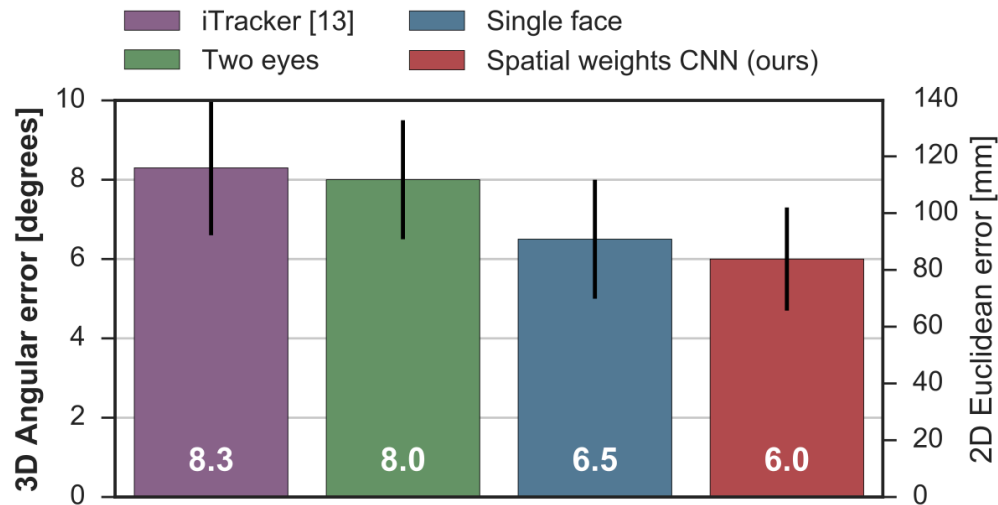

3D Gaze Estimation

As the same setting for our MPIIFaceGaze dataset and EYEDIAP dataset, we conducted the 3D gaze estimation evaluations. The left axis shows the angular error that was directly calculated from the estimated and ground-truth 3D gaze vectors. The right axis shows the corresponding Euclidean error that was approximated by intersecting the estimated 3D gaze vector with the screen plane. For our MPIIFaceGaze dataset (below), our proposed model achieved a significant performance improvement of 14.3% (paired t-test: $p > 0.01$) over iTracker, and a performance consistent with the 2D case.
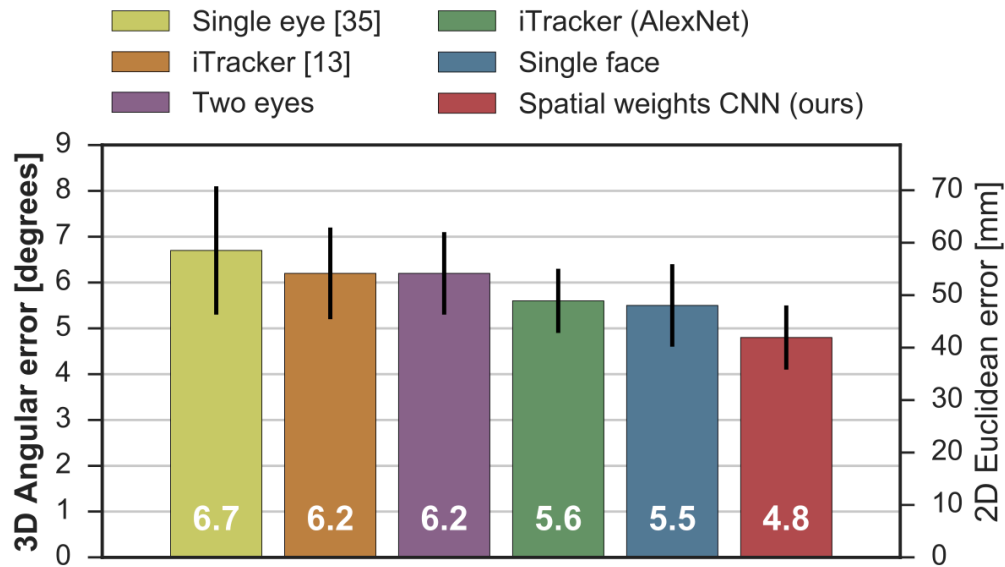
For EYEDIAP dataset (below), the proposed model also achieved the best performance for the 3D gaze estimation task on the EYEDIAP dataset.

References

[1] Eye Tracking for Everyone. K.Krafka, A. Khosla, P. Kellnhofer, H. Kannan, S. Bhandarkar, W. Matusik and A. Torralba IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[2] Appearance-based Gaze Estimation in the Wild, X. Zhang, Y. Sugano, M. Fritz and A. Bulling, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[3] EYEDIAP: A Database for the Development and Evaluation of Gaze Estimation Algorithms from RGB and RGB-D Cameras. Funes Mora, Kenneth Alberto and Monay, Florent and Odobez, Jean-Marc. Proceedings of the ACM Symposium on Eye Tracking Research and Applications (ETRA), 2014.

[4] Learning-by-synthesis for appearance-based 3d gaze estimation. Y. Sugano, Y. Matsushita, and Y. Sato. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.

Contact Us

cai-office@vis.uni-stuttgart.de

University of Stuttgart

Institute for Visualisation and Interactive Systems

Collaborative Artificial Intelligence

Pfaffenwaldring 5a

70569 Stuttgart

Germany