General

- Gaze: externally-observable indicator of human visual attention
- Problems with current solutions: high cost, custom or invasive hardware, inaccuracy under real-world conditions
- Appearance-based methods usually only use image information encoded from one or both eyes
    - Multi-region CNN architecture that takes eye and face images as input can benefit gaze estimation performance
- Model-based methods: estimate gaze direction using geometric models of the eyes and face
    - First detect visual features (such as pupil, eyeball centre, and eye corners), and then fit a geometric 3D eyeball model to them to estimate gaze
- Appearance-based methods: directly regress from eye images to gaze direction
    - Only require images, but directly learn a mapping from 2D input images to gaze directions using machine learning
    - Can typically handle input images with lower resolution and quality than model-based methods
    - Appearance-based: regression target can be in 2D or 3D
        - 2D: on-screen gaze location (target screen plane is fixed in the camera coordinate system → does not allow for free camera movement after training; less demanding and more accurate if application scenario can afford a fixed screen position)
        - 3D: 3D gaze directions in the camera coordinate system (can be applied to different device and setups without assuming a fixed camera-screen relationship → more general and practically relevant)
- Feature-based methods: use eye features for gaze direction regression, such as corneal reflections caused by reflections of an external light source on the cornea
    - Commonly used in commercial eye trackers

Current Gaze Estimation Models

- Two types of gaze-tracking algorithms: model-based and appearance based
    - Appearance-based: operate directly on the eye images

Measures of Accuracy for Eye Gaze Tracking Methods

- Angular resolution in degrees
- Gaze recognition rates in percentage
- Shifts in number of pixels between gaze and target locations

- Distance in cm/mm between gaze and target locations

Model-Based Methods

- 2D Models
    - Utilize polynomial transformation functions for mapping the gaze vector to corresponding gaze coordinates on the screen
    - Gaze vector: vector between pupil center and corneal glint
    - Typical accuracy: between two and four degrees
- 3D Models
    - Typically use a geometrical model of the human eye to estimate the center of the cornea, and the optical and visual axes of the eye
    - Gaze coordinates are estimated as points of intersection of the visual axes with the scene
    - Achieve high accuracy (1 degree) but require elaborate system setups and knowledge about geometric relations between system components
    - Recent developments: usage of depth sensors along with RGB cameras

Appearance-Based Methods

- Utilize cropped eye images of a subject gazing at known locations to generate gaze point coordinates
- Eye images are used as training data for various machine learning models
- Recently, appearance-based methods implemented using deep learning and convolutional neural network approaches have gained momentum

Gaze Estimation from Low Resolution Images

- Webcams preferred to facilitate gaze tracking in everyday settings; however, webcams offer low resolution images
- Low resolution images: strong noise effects, and distortions in the eye region contours and eye features become indistinguishable under varying illumination levels, user distance, and movements
- Approaches
    - Map gaze coordinates to low quality cropped eye images
    - Iris centers determined first using circular a Hough transform, followed by refinement using a gradient-aware random sample consensus algorithm and ellipse fitting. Eye corners determined using Gabor jets and tracked using optical flow with normalized cross-correlation. Point of gaze estimated from the iris center and eye corners using regression
    - Problem from small size of cropped eye regions from low resolution images -- overcome using 2D bilinear interpolation for reconstructing the eye image to a larger size for accurate tracing of the corneal reflection vector

Eye Gaze Estimation Using CNNs

- Deep learning techniques have been successfully used in challenging conditions such as those with variable illumination, unconstrained backgrounds, and free head motion
- Approaches
    - Two CNNs (for left and right eyes) trained independently to classify the gaze in seven directions
    - Deep features obtained from eye images using multi-scale convolutions and pooling for predicting gaze direction
        - Uses minimized cross-entropy loss, coupled with Random Forest regression as a clustering algorithm
        - Classifies areas on a device screen according to gaze locations, and operates under natural illumination and head pose
    - Utilize full face image as input with spatial weights on the feature maps to suppress or enhance information in different facial regions -- Bulling Group
        - Achieves high accuracy and robust performance under varied illumination and extreme head poses
    - Use 2 separate head pose and eye movement models with 2 CNNs, connected via a gaze transform layer
        - Achieves free head pose, 3D gaze tracking
    - CNN build to learn the mapping between 2D head angle, eye image, and gaze angle (output) using a Lenet-inspired CNN
    - GazeCapture - specifically targeted towards gaze tracking in consumer/handheld devices
        - Realtime, calibration-free
        - Trained using a large and diverse dataset of eye images taken under variable lighting, head pose, and backgrounds captured from users through a smartphone app
        - Inputs to model: eye and face images
        - Location of faces in images are obtained through a face grid, which is used to infer relative eye and head positions

MPII Gaze Dataset

- Extensive database build with more than 200,000 images under variable illumination levels, eye appearances, and head poses
- Used for training and testing