视线估计是一种旨在理解用户意图和兴趣的技术[Illumination-free gaze estimation method for first-person vision wearable device]。视线估计技术关注的是图像数据和注视方向或者注视点之间的关系[In the eye of the beholder: A survey of models for eyes and gaze]。根据从单个或者多个摄像头采集的眼睛区域的图像数据（如瞳孔或者角膜反射）来估计注视方向[A survey on eye-gazetracking techniques]

视线估计方法通常可以分为基于模型的（model-based）和基于外观的（apperence-based）[In the eye of the beholder: A survey of models for eyes and gaze]。基于模型的方法是使用一个眼睛的几何模型，利用人眼成像的原理，通过相机获取人眼眼球的特征，随后根据获取到的特征建立与注视点的映射关系，从而实现实现估计。基于模型的方法可以进一步分为基于角膜反射的方法和基于形状（shape-based）的方法。角膜反射法通过外部红外光源反射到眼睛的角膜来检测到眼睛的特征。

(1)基于外观的视线估计（APPEARANCE-BASED GAZE ESTIMATION）

基于外观的视线估计方法不依赖于眼睛特征点的检测，而是直接从眼睛光学外观来执行视线估计的[[A survey on eye-gazetracking techniques]。通常需要一台摄像机来捕捉眼睛的图像，然后用它来简历视线估计模型，将捕捉到的图像外观映射到特定的注视方向或者注视点。这些模型不需要手工设计特征进行训练，它们可以从数据中隐式的提取图像特征。最早的基于外观的视线估计工作是将眼睛图像作为神经网络的输入，用于100个分类任务，每个类别代表一个注视方向[Non-intrusive gaze tracking using artificial neural networks]。Tan等人[Appearance-based eye gaze estimation]是第一个将基于外观的视线估计作为论文工作的，该论文将校准样本视为流形中的点，并试图寻找权重来插值测试样本。本文将基于外观的视线估计的基本形式定义为特征提取和从特征到注视目标的回归映射。接下来的工作分别使用高斯过程回归[Sparse and Semi-supervised Visual Mapping with the Sˆ 3GP, in Proc]，神经网络[Appearance-based Gaze Tracking with Spectral Clustering and Semi-supervised Gaussian Process Regression]，和自适应线性[Learning Gaze Biases with Head Motion for Head Pose-free Gaze Estimation, Image and Vis]等方法，沿用原始像素作为他们的特征输入。这些早期的基于外观的视线估计方法都假定头部姿势是固定的，以此来减少外观的变化，避免复杂的讨论。

随后的研究开始允许自由的3D头部移动。Lu等人[Learning Gaze Biases with Head Motion for Head Pose-free Gaze Estimation, Image and Vis]将原始的头部自由运动问题分解为了两个子问题，分别是初始固定头部问题和随后的补偿来矫正初始估计误差，将由头部姿势引起的注视偏差作为最终视线估计结果的补偿。

尽管基于外观的视线估计技术取得了很大的进展，但这些方法比传统方法需要更多的训练数据来学习头部自由运动引起的眼睛外观的显著变化。因此，这些研究主要针对于特定领域或者个人进行评估。视线估计中的一个开放性研究挑战是学习没有用户、环境和相机限制的视线估计器。这些估计器可以在不需要额外输入的情况下推广到任何人。Schneider等人[Manifold Alignment for Person Independent Appearance-Based Gaze Estimation]使用了哥伦比亚数据集（Columbia dataset）[Gaze locking: passive eye contact detection for human-object interaction ]对流形对齐方法进行了评估，发现在6种回归方法中，使用流形对齐方法可以提高视线估计性能。Zhang等人[Rendering of Eyes for Eye-Shape Registration and Gaze Estimation]为了避免摄像头的偏移和人头部姿态对视线估计的影响，首先对眼部数据进行人脸对齐和3D头部估计、眼部图片归一化，将单眼图像作为输入，通过LeNet-5网络对图像特征进行提取，并求出的二维头部偏转信息合并到视线估计网络的全连接层，依据混合特征进行视线估计。该论文同时提出了MPIIGaze数据集，该数据集包括21万张人脸图片，涵盖15位参与者和不同光照条件。Krafka等人[eyes tracking for everyone]通过IOS应用收集了GazeCapture数据集，包含1450多人、240多万张数据，并提出iTracker网络，将人左右眼部、脸部图片和脸部网格图片分别输入到模型当中，最终预测相对于摄像头的凝视位置信息，在iPhone上实现了1.71厘米的估计误差，在iPad上达到了2.53厘米的估计误差。Deng和Zhu[Monocular Free-head 3D Gaze Tracking with Deep Learning and Geometry Constraints]提出了一个视线变换模型，将头部姿势和眼球运动模型连接起来，并收集了一个大的数据集，包括200个志愿者和不同光照条件。在另一个研究中，Kim等人[Convolutional neural network architectures for gaze estimation on mobile devices]提出了另一个基于回归的视线估计模型，他们受到了iTracker网络的启发，设计了名为Gazelle的模型，该模型需要五个输入，分别是：右眼、左眼、脸、脸网格和梯度直方图(HOG,Histogram of Gradients)，更好的捕捉用于视线估计的特征。在上述模型中，人脸图像主要提供头部姿态信息，而脸部网格则提供眼部位置信息，存在一定的信息冗余，因此，He等人[On-device Few-shot Personalization for Real-time Gaze Estimation]对上述模型的输入进行了简化，将脸部图片和脸部网格图片这两个输入替换为四个眼角的坐标位置信息，因为眼角坐标不止提供了眼睛的位置信息，还暗含了头部姿态信息，即眼角间距越小，头部姿态越大。实验结果表明，该简化后的模型在iPhone上的误差达到了1.78cm，和原始模型相差不多，同时改模型的计算效率很高，处理速度可以达到10ms/帧。随后，Zhang等人[MPIIGaze: Real-World Dataset and Deep Appearance-Based Gaze Estimation]尝试将AlexNet修改为16层的VGG网络，该研究将平均误差从13.9度提升至10.8度。2017年，Zhang等人[It’s Written All Over Your Face Full-Face Appearance-Based Gaze Estimation]认为基于学习的方法可以利用其它面部区域的附加特征，不同于之前手动提取眼部和脸部图像信息，他们尝试只将人脸图像作为输入，利用CNN对人脸图像进行编码，同时提出了一种空间权重机制，将人脸不同区域的信息有效地编码到CNN架构中，该机制在卷积层的特征图上学习空间权重，来灵活的抑制或者增强不同区域的信息。经过广泛的评估，该方法在MPIIGaze和EYEDIAP上分别有14.3%和27.7%的改进。因为数据集收集繁琐和标注困难等问题，Wood等人[Rendering of Eyes for Eye-Shape Registration and Gaze Estimation]开发了一个动态眼部区域模型，可以生成不同姿态的带有标记的可训练图像，从而大大减少了数据的收集和标注工作。在此基础上，Wood等人[Learning an appearance-based gaze estimator from one million synthesised images]提出了UnityEyes，一种可以快速合成大量可变眼域图像的新方法，该方法结合了一个人眼区域生成三维模型和一个实时绘制框架，并将该框架免费在线提供，方便研究使用。最新的方法是通过对抗训练模型[earning from simulated and unsupervised images through adversarial training]或者生成模型[A Hierarchical Generative Model for Eye Image Synthesis and Eye Gaze Estimation]来生成仿真图像。2020年，Seonwook等人[Towards End-to-end Video-based Eye-Tracking]提出了一种新的EVE数据集（End-to-end Video-based Eye-tracking）和相应的方法。该数据集收集了54名参与者，超过1200万帧，总计约105小时的视频数据。模型分为两个部分，分别是EyeNet和GazeRefineNet，EyeNet负责对每帧进行视线估计和瞳孔大小估计，GazeRefineNet通过合并屏幕内容、时间信息和EyeNet得到的初始注视点信息来优化最后的注视点预测结果。最终方法使得注视点估计正确率提高了28%，角度误差减小到了2.49度。