

Abstract

The present application provides a multimodal gaze estimation method, system, electronic device and storage medium based on a lightweight graph neural network, involving human-computer interaction applications, gaze estimation technology fields. Firstly, multimodal heterogeneous data is obtained, and the visual features are extracted from the visual appearance data. Then the visual feature, geometric orientation information, and spatial positioning coordinates are projected into the 32-dimensional space. Then, the graph structure is constructed, and the constructed graph is processed based on the multi-head attention graph neural network with a two-layer structure to obtain graph-level features. Further, the graph-level features are aggregated and fused with visual features, geometric orientation information, and spatial positioning coordinates. Finally, the gaze estimation prediction is carried out based on the fusion features by using the fully connected layer, and coordinate conversion is carried out to convert the gaze estimation prediction into a normalized three-dimensional gaze vector. The present application is suitable for mobile deployment of lightweight and accurate gaze estimation, which can effectively realize multimodal feature integration and rapid subject-specific adaptation.

Figure 1