論文の内容の要旨

論文題目

Head Pose and Gaze Estimation for Inferring Focus of Attention
(注目領域獲得のための頭部姿勢および注視点推定)

氏名　　菅野　裕介

Head pose and gaze direction play significant roles in inferring human attention, and they also help us to design more human-centered computer systems. Especially, camera-based remote sensing techniques of head pose and gaze can be led to a wide range of applications. However, although a lot of methods have been proposed, there exist some technical limitations of the estimation techniques. Accurate estimation using only a monocular camera is still a difficult task, and existing methods often require calibration actions prior to the estimation procedure. The goal of this thesis is developing head pose and gaze estimation systems with minimal requirements; all proposed methods do not need active calibration stages and additional equipments other than a camera.

The first part describes a monocular method of tracking 3D head poses and facial actions. Using a multilinear face model that treats interpersonal and intrapersonal shape variations separately, real-time parameter estimation is done by integrating two different frameworks: particle filter-based tracking for time-dependent pose and facial action estimation and incremental bundle adjustment for person-dependent shape estimation. This unique combination together with multilinear face models enables tracking of faces and facial actions in real time with no pre-learned individual face models.

In the second part, an unconstrained gaze estimation method is presented, which allows free head movements of users in a casual desktop environment using an online learning algorithm. The key assumption is that a user gazes at a cursor position when s/he presses a mouse button. The user's eye images and 3D head poses are continuously captured based on the head pose estimation method described above. By using clicked positions as exemplars of gaze positions, our system collects learning samples for estimating gazes while a user is unconscious of the system when using a PC. The samples are adaptively clustered according to the head pose and estimation parameters are incrementally updated. In this way, our method avoids the lengthy calibration stage

prior to using the gaze estimator.

One of the drawbacks of the above method is that it cannot be applied to passive displays without user interaction. To solve the problem, another novel calibration-free gaze sensing framework using visual saliency maps is presented in the last part. The method uses visual saliency maps of video frames that are computed in a bottom-up manner. By relating the saliency maps with appearances of eyes of a person watching video frames, our method automatically constructs a gaze estimator. To efficiently identify gaze points from saliency maps, saliency maps are aggregated to generate probability distributions of gaze points. Mapping between eye images and gaze points is then established by Gaussian process regression. This results in a gaze estimator that exempts users from active calibration and can be applied to any type of display devices.

Throughout these works, head pose and gaze estimation methods were made significantly more practical by reducing installation and setup costs. The proposed methods can be used with commonly-available cameras and estimation procedures without manual initialization can be seamlessly integrated into our daily computer interactions. This enhances potential for future investigations of attention-based application systems that enrich our daily lives with ubiquitous computing devices.