論文の内容の要旨

論文題目␣ Synchrony-based Audiovisual Analysis

（同期性に基づく音と映像の統合解析）

氏名　劉　玉宇

This thesis presents a computational framework to jointly analyze auditory and visual information. The integration of audiovisual information is realized based on synchrony evaluation, which is motivated by the neuroscience discovery that synchrony is a key for human beings to perceive across the senses of different modalities. The works in this thesis focus on answering two questions: how to perform and where to apply this audiovisual analysis with synchrony evaluation. To answer the first question, we develop novel effective methods to analyze the audiovisual correlation, and perform a classification and an experimental comparison of the existing techniques, including the ones we developed. Since this is the first work that classifies and experimentally compares the methods of this field, it supplies a basis for designing algorithms to computationally analyze the audiovisual correlation. To answer the second question, we apply audiovisual correlation analysis to solve three different problems. The first problem is the detection of a speaker face region in a video, whose previous solutions either require special devices like microphone array or supply only highly fragmental results. Assuming speaker is stationary within an analysis time window, we introduce a novel method to analyze the audiovisual correlation for speaker using newly introduced audiovisual differential feature and quadratic mutual information, and integrate the result of this correlation analysis into graph cut-based image segmentation to compute the speaker face region. This method not only achieves the smoothness of the detected face region, but also is robust against the change of background, view, and scale. The second problem is the localization of sound source. General sound source is diverse in types and usually non-stationary while emitting sounds. To solve this problem, we develop an audiovisual correlation maximization framework to trace the sound source movement, and introduce audiovisual inconsistency feature to extract audiovisual events for all kinds of sound sources. we also propose an incremental computation of mutual information to significantly speed up the computation. This method can successfully localize

different moving sound sources in the experiments. The third problem is the recovery of drifted audio-to-video synchronization, which used to require both special device and dedicated human effort. Considering that the correlation reaches the maximum only when audio is synchronized with video, we develop an automatic recovery method by analyzing the audiovisual correlation for a given speaker in the video clip. The recovery demonstrates high accuracy for both simulation and real data. While the theoretical justification and experimental justification are performed independently, this thesis taken as a whole lays a necessary groundwork for jointly analyzing audiovisual information based on synchrony evaluation.