PROCEEDINGS of the
23rd International Congress on Acoustics
9 to 13 September 2019 in Aachen, Germany

Inference in nonlinear dynamical systems with dynamic stream weights for audiovisual speaker tracking

Christopher SCHYMURA[1], Dorothea KOLOSSA[2]

[1]Ruhr University Bochum, Germany, christopher.schymura@rub.de
[2]Ruhr University Bochum, Germany, dorothea.kolossa@rub.de

Abstract

Audiovisual speaker tracking is a prominent example of a multimodal signal processing framework with a large variety of technical applications. Whereas acoustic speaker tracking methods suffer from decreased performance under adverse acoustic conditions, visual sensors are not affected by purely acoustic disturbances. However, the optimal combination of acoustic and visual information remains a challenging task due to the highly dynamic nature of audiovisual scenes. Recently, a framework for audiovisual speaker localization and tracking has been proposed, which extended the notion of dynamical systems with dynamic stream weights. It essentially implements an adaptive weighting for acoustic and visual observations that can be changed at each time step and allows probabilistic inference based on the Gaussian filtering paradigm. This study presents a detailed analysis of the inference procedure and proposes an extension for nonlinear dynamical systems inspired by the unscented Kalman filter. Under the constraint that the process and observation noise terms follow a zero-mean Gaussian distribution, the inference framework presented in this study allows an efficient estimation of the speaker trajectory from audiovisual observations. A systematic experimental evaluation assesses the performance of the proposed method using recorded audiovisual scenes.

Keywords: Audiovisual speaker tracking, recursive Bayesian estimation, dynamic stream weights