DOCTORAL THESIS

Lip motion tracking and analysis with application to lip-password based speaker verification

Liu, Xin

Date of Award:
2013

Lip Motion Tracking and Analysis with Application to Lip-password Based Speaker Verification

LIU Xin

A thesis submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Principal Supervisor: Prof. CHEUNG Yiu-ming

Hong Kong Baptist University

December 2012

Abstract

Lip motion tracking and analysis have received wide attention in recent years as the dynamic patterns of lip movement accompanying with the lip shape variations, tongue and teeth visibility, provide valuable information for many practical applications such as lipreading, audio-visual speech recognition, visual speaker recognition, talking face detection, and so forth. In this thesis, we first investigate two important issues concerning to the lip motion tracking (i.e., lip region tracking and lip contour tracking) and then apply these techniques associated with the discriminative motion analysis to solve lip-password based speaker verification problem.

First, we present an incremental weighted appearance learning approach with regional particle filtering scheme for efficient lip region localization and tracking. In the proposed approach, we incrementally learn a subspace representation through adaptively updating the time-varying mean and eigenbasis by considering the temporal and spatial weights, in which an Epanechnikov monotonically decreasing kernel is employed to reliably weight the important pixels within the interested lip region. The prediction scheme utilizing the regional particle filtering associated with the affine motion parameters not only has the ability to precisely locate the lip region with well estimated scale and rotation, but also contributes to a reduction of computational load and alleviates the occurrence of tracking drift as well. Meanwhile, the rank-2-optimal states of the corresponding affinely warped image patches are utilized to incrementally update the weighted lip appearance model in three CIElab color channels simultaneously, featuring adaptivity and sequentiality to model the rapid lip appearance changes.

Second, we present a local region based approach to lip contour tracking that consists of two phases: 1) lip contour extraction for the first lip frame, and followed by 2) lip contour tracking in the subsequent lip frames. A localized color active contour model is introduced provided that the foreground and background regions around the object are locally different in color space. Meanwhile, an effective illumination equalization method is given to reduce lighting asymmetry during the image pre-processing process. Accordingly, we find a combined semi-ellipse around the lip region as the initial evolving curve and compute the localized energies associated with a 16-point deformable model for lip contour extraction. Further, we present a dynamic selection of the radius of local regions associated with the extracted lip contour of the previous frame to realize lip contour tracking.

Third, we extract a group of representative visual features and propose an effective lip motion segmentation algorithm to segment the digital lip-password sequence into a small set of distinguishable subunits (i.e., password elements). Accordingly, we integrate HMMs with boosting learning framework associated with random subspace method (RSM) and data sharing scheme (DSS) to discriminatively formulate a precise decision boundary for these subunits verification. Finally, the lip-password whether spoken by the target speaker with pre-registered password or not is identified based on all subunit verification results learned from multi-boosted HMMs.

Extensive experiments are conducted to show the promising results of the proposed approaches in comparison with the existing methods.

Keywords: Lip region tracking, lip contour tracking, lip-password, speaker verification, regional particle filtering, lip motion segmentation, multi-boosted HMMs

Table of Contents