

DOCTORAL THESIS

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

Liu, Xin

Date of Award:
2013

[Link to publication](#)

General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

**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

Declaration	i
Abstract	ii
Acknowledgements	iv
Table of Contents	v
List of Tables	ix
List of Figures	x
Chapter 1 Introduction	1
1.1 Review Lip Motion based Applications	1
1.1.1 Lipreading	2
1.1.2 Audio-Visual Speech Recognition	4
1.1.3 Visual Speaker Recognition	5
1.2 Lip-password based Speaker Verification	7
1.2.1 The Limitation of the Existing Patterns	8
1.2.2 The Motivation of Lip-password based Speaker Verification . .	9
1.3 The Problems Investigated & Main Contributions in this Thesis . . .	10
1.4 Organization of the Thesis	13
Chapter 2 Review of Related Works	14
2.1 Lip Image Analysis in Color Space	14

2.2	Lip Modeling Patterns	16
2.2.1	Deformable Template	17
2.2.2	Active Contour Model	18
2.2.3	Active Shape Model and Active Appearance Model	20
2.3	Visual Feature Extraction	22
2.3.1	Shape-based Visual Features	22
2.3.2	Appearance-based Visual Features	23
2.3.3	Motion-based Visual Features	24
Chapter 3 Lip Region Localization and Tracking		27
3.1	Introduction	28
3.2	Related Works and Motivation	29
3.3	Incremental Weighted Appearance Learning	33
3.3.1	Incremental Weighted PCA with Mean Update	34
3.3.2	Spatial Weighting Scheme	38
3.3.3	The Weighted Appearance Learning in Color Space	39
3.4	Regional Particle Filtering for Tracking	40
3.5	The Pipeline of The Tracking Algorithm	46
3.6	Experimental Results	48
3.6.1	Experimental Setups	49
3.6.2	Performance Comparison	51
3.6.3	Quantitative Comparisons	54
3.6.4	Discussion	57
3.7	Summary	58
Chapter 4 Lip Contour Extraction and Tracking		59
4.1	Introduction	60
4.2	Localized Color Active Contour Model	65
4.2.1	Overview of the LACM	65
4.2.2	An Extension to LCACM	67

4.3	The Proposed Lip Contour Extraction and Tracking Algorithm	68
4.3.1	Illumination Equalization	69
4.3.2	Lip Contour Extraction	71
4.3.3	Lip Model	74
4.3.4	Lip Contour Tracking	77
4.4	Experimental Results and Discussion	78
4.4.1	Contour Extraction Results	80
4.4.2	Contour Tracking Results	84
4.5	Summary	88
Chapter 5 Approach to Lip-password based Speaker Verification		89
5.1	Introduction	90
5.2	Overview of Related Works	92
5.2.1	Discrimination Analysis	92
5.2.2	Overview of HMM-based Speaker Verification	94
5.2.3	Overview of Adaboost Learning	96
5.3	The Proposed Approach to Lip-password Based Speaker Verification .	98
5.3.1	Visual Feature Extraction	98
5.3.2	Lip Motion Segmentation	100
5.3.3	The Proposed Multi-boosted HMMs Learning Approach . . .	103
5.4	Experimental Results and Discussion	108
5.4.1	The Target Speaker Saying the Incorrect Passwords Case . .	109
5.4.2	The Imposter Saying the Correct Password Case	111
5.4.3	The Imposter Saying the Incorrect Passwords Case	113
5.5	Summary	117
Chapter 6 Conclusion and Future work		118
6.1	Conclusions	118
6.2	Future Work	121
Bibliography		123

Appendices	139
Chapter A	140
A.1 The Color Space Conversion: RGB to CIElab	140
Chapter B	142
B.1 Illumination Equalization	142
B.2 Cubic Spline Interpolation	143
Curriculum Vitae	145