Smiles vs chewing vs speech detection by similarity matching

Wenyang Cai

August 18, 2014

Abstract

This is abstract.

Acknowledgements

This is acknowledgement

Contents

Chapter 1

Introduction

This is Introduction.

1.1 Motivation

1.2 Thesis

1.3 Contribution

Chapter 2

Background

This is Background [2]. This is Background [5]. This is Background [3].

2.1 Automatic Speeech Recognition

2.2 Visual Front End

2.2.1 Face and Facial Part Detection

2.2.2 Region of Interest

2.3 Facial Expression

2.4 Nonlinguistic Vocalization

Chapter 3

Processing and Methodologies

3.1 Processing Flow

3.2 Face Alignment

The aim of face alignment is to localize the feature points on face images. The points are usually around eyes, nose, mouth, and outline. Face alignment techniques are essential on face recognition, modelling and synthesis. There are three main different approaches Parametrized Appearance Models(PAMs), Discriminative approaches, Part-based deformable models. Parametrized appearance models contains many models such as active appearance models (AAMs), morphrable models, eigentrackings, and template tracking [5]. All these models are using PCA method to parametrize a face. A face could approximately decomposed as linear combination of shape basis and appearance basis. The problem of face alignment could be refer as minimizing the difference between the constructed PAM and the face. Common approach is use Gauss-Newton methods [5]. Discriminative approaches are to learn the linear regression between the head move and appearance change. Part-based deformable model perform face alignment by maximizing the posterior likelihood of part locations given image [5].

3.2.1 Aative Appearance Model

Active Appearance Model (AAMs) is defined as a generative model of a certain visual phenomenon in [2]. AAMs are closely related to concept of morphable models, constrained models and active blobs. In face alignment it is a face model consists of linear shape model and appearance model. There are two types of AAMs, one refers as independent shape and appearance models, which model shape and appearance independently, and the other refers as combined shape and appearance models, which parameterized shape and appearance model with a single set of linear parameters [2]. Normally AAMs appears along with a fitting algorithm. However, in the following context, it only refers to a model. [2] gave a well explain about what is an AAM.
Shape of a face is definded as a mesh and is represented as coordinates of v vertices of the face points:

$$s = (x_1, y_1, x_2, y_2, ..., x_v, y_v)^T \tag{3.1}$$

s also can be expressed as a base shape s_0 plus linear combination of n shape vectors s_i:

$$s = s_i + \sum_{i=1}^{n} p_i s_i \tag{3.2}$$

For all pixels x in the mesh s_0, appearance $A(0)$ can be expressed as base appearance $A_0(x)$ and m appearance images $A_i(x)$.

$$A(x) = A_0(x) + \sum_{i=1}^{m} \lambda_i A_i(x) \qquad \forall x \in s_0 \tag{3.3}$$

Next equation defines the appearance of s. $W(x:p)$ is the warp from s_0 to s. Then the model M set the appearance of $W(x:p)$ to $A(x)$.

$$M(W(x:p)) = A(x) \tag{3.4}$$

Combined AAMs
Combined AAMs just use parameter $c = (c_1, c_2, ...)^T$ to parametrize shape:

$$s = s_0 + \sum_{i=1}^{l} c_i s_i \tag{3.5}$$

and appearance:

$$A(x) = A_0(x) + \sum_{i=1}^{l} c_i A_i(x) \tag{3.6}$$

3.2.2 Trackers

In the processing of face alignment I tried three trackers, but mainly using two trackers, one is from Intraface [5] and the other DRMF [1].

Intraface [5] implies image alignment can be posed as solving a nonlinear optimization problem. It uses Supervised Descent Method for minimising Non-linear Least Square(NLS) function, which avoids calculating the Hessian and the Jacobian that could be computationally expensive.
Examples:

DRMF DRMF uses novel discriminative regression based on Constrained Local Models(CLMs) for face alignment.
Examples

3.2.3 Comparison

[5] implies that face alignment problem are usually treated as solving continuous nonlinear optimisation problem. [5] uses supervised descent method (SDM) for minimising the Non-linear Least Square (NLS) function. [1] uses discriminative regression approach for constrained local method (CLM). However, from the computing time and alignment results, [5] is better than [1] in many aspects.
Description

3.3 Remove Head-pose

The algroithm of removing head-pose from tracking points is in [4].The following are some example of orginal track points and deformed points:

3.4 Warping

In order to have the appearance image of the face after removed head-pose, it is necessary to warp the face with head pose. Basic idea is to for each triangles builded by tracking points, the image points in the triagnles are projected to the corresponding triagnles built by deformed points. The following are some examples of face before and after warping:

3.5 Feature Extraction

The image after warping is not directly used for classfication. The data for classfication is the features of the image. There are many techniques to extract features from images, in this experiment, Local Binary Pattern are used for extracting image feature.

3.5.1 Local Binary Pattern

Effective facial representation of the original face iamges is an important part of successful facial expression recognition.

3.6 Postprocessing

Due to the time limits, in the experiment part, we only use support vector machine to do classfication.

Normalization

Scaling

Chapter 4

Experiment and Results

4.1 Database

4.1.1 Feature and Data

4.2 Methodology

RBF

4.3 Experiments

4.4 Results and Analysis

Chapter 5

Conclusion and Future Work

Bibliography

[1] Akshay Asthana, Stefanos Zafeiriou, Shiyang Cheng, and Maja Pantic. Robust discriminative response map fitting with constrained local models. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3444–3451. IEEE, 2013.

[2] Iain Matthews and Simon Baker. Active appearance models revisited. *International Journal of Computer Vision*, 60(2):135–164, 2004.

[3] Gerasimos Potamianos, Chalapathy Neti, Guillaume Gravier, Ashutosh Garg, and Andrew W Senior. Recent advances in the automatic recognition of audiovisual speech. *Proceedings of the IEEE*, 91(9):1306–1326, 2003.

[4] Jason M Saragih, Simon Lucey, and Jeffrey F Cohn. Deformable model fitting by regularized landmark mean-shift. *International Journal of Computer Vision*, 91(2):200–215, 2011.

[5] Xuehan Xiong and Fernando De la Torre. Supervised descent method and its applications to face alignment. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 532–539. IEEE, 2013.