

Facial Landmark Detection and Tracking for Facial Behavior Analysis

Yue Wu

Rensselaer Polytechnic Institute
110 8th street, Troy, NY, USA
wuy9@rpi.edu

ABSTRACT

The face is the most dominant and distinct communication tool of human beings. Automatic analysis of facial behavior allows machines to understand and interpret a human's states and needs for natural interactions. This research focuses on developing advanced computer vision techniques to process and analyze facial images for the recognition of various facial behaviors.

Specifically, this research consists of two parts: automatic facial landmark detection and tracking, and facial behavior analysis and recognition using the tracked facial landmark points. In the first part, we develop several facial landmark detection and tracking algorithms on facial images with varying conditions, such as varying facial expressions, head poses and facial occlusions. First, to handle facial expression and head pose variations, we introduce a hierarchical probabilistic face shape model and a discriminative deep face shape model to capture the spatial relationships among facial landmark points under different facial expressions and face poses to improve facial landmark detection. Second, to handle facial occlusion, we improve upon the effective cascade regression framework and propose the robust cascade regression framework for facial landmark detection, which iteratively predicts the landmark visibility probabilities and landmark locations.

The second part of this research applies our facial landmark detection and tracking algorithms to facial behavior analysis, including facial action recognition and face pose estimation. For facial action recognition, we introduce a novel regression framework for joint facial landmark detection and facial action recognition. For head pose estimation, we are working on a robust algorithm that can perform head pose estimation under facial occlusion.

CCS Concepts

•Computing methodologies → Activity recognition and understanding; Interest point and salient region detections; *Object detection; Tracking;*

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ICMR'16, June 06 - 09, 2016, New York, NY, USA

© 2016 Copyright held by the owner/author(s). Publication rights licensed to ACM. ISBN 978-1-4503-4359-6/16/06...\$15.00

DOI: <http://dx.doi.org/10.1145/2911996.2912034>

Keywords

Facial behavior analysis; Facial landmark detection and tracking; Probabilistic graphical model

1. INTRODUCTION

Humans communicate through multiple channels, including visual, auditory, olfactory, and tactile. The face plays an important role in visual communication. For example, humans can automatically extract many nonverbal messages by visualizing and analyzing facial behavior. In particular, facial behavior includes the facial deformations caused by facial expressions, head movements, and eye movements. Facial behavior reflects a human's emotion, focus of attention, and mental state. Facial behavior has been applied to multiple application areas, including human and computer interaction, entertainment, and medical applications.

In computer vision, there are a few major facial behavior analysis tasks (see Figure 1). First, given the facial images or videos as visual inputs, face detection/tracking techniques are performed to locate the face. Given the location of the face, facial landmark detection and tracking techniques are used to locate the fiducial facial key points, such as eye corners and eye centers. The facial landmark locations are used to perform advanced tasks such as facial expression and action recognition, head pose estimation, eye movement and gaze estimation, etc.

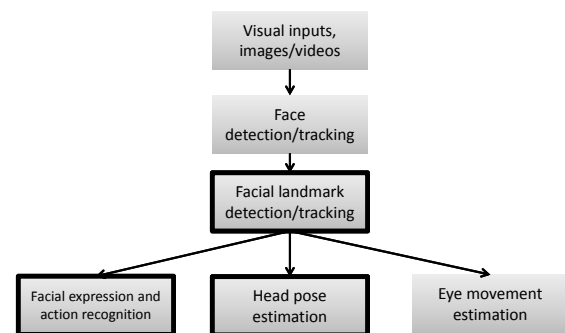


Figure 1: Major facial behavior analysis tasks in computer vision. The bold boxes highlight the topics covered in this research.

As highlighted in Figure 1, the major facial behavior analysis topics covered in this research include facial landmark detection/tracking, facial expression and action recognition,

and head pose estimation. In particular, the first part of the research focuses on facial landmark detection in the wild, which refers to the landmark detection on images with varying facial expressions, head poses, and facial occlusion. In the second part of this research, we apply robust facial landmark detection and tracking methods for facial expression and action unit recognition and head pose estimation. In particular, our work extends the existing research and contributes in the following aspects:

1. For facial landmark detection under varying facial expressions and head poses, we improve upon the CLM by proposing two effective facial shape models.
2. For facial landmark detection under facial occlusion, we improve upon the general cascade regression methods and propose a robust cascade regression method, which iteratively predicts facial landmark locations and facial occlusions.
3. To improve the facial expression and facial action unit recognition, we propose the joint cascade regression framework for simultaneous facial action recognition and facial landmark detection.
4. We are working on a robust method to improve head pose estimation under facial occlusion.

The remaining part of this manuscript is organized as follows: In section 2, we introduce the facial landmark detection algorithms. In section 3, we introduce the facial action unit recognition method. In section 4, we introduce the robust head pose estimation method. We summarize this manuscript in section 5.

2. FACIAL LANDMARK DETECTION AND TRACKING

2.1 Overview

Facial landmarks refer to the fiducial facial key points around the facial components and contour, as shown in Figure 2. They are usually with unique local facial appearance, and they are utilized as anchor points on the face. The goal of facial landmark detection and tracking techniques is to estimate the 2D coordinates of the facial landmarks on facial images or videos. Estimation of facial landmark locations is important, since these locations can jointly represent the face shape, which is usually required by facial behavior analysis tasks as shown in the flowchart in Figure 1.

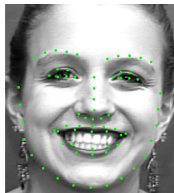


Figure 2: Facial landmark on a sample image.

In general, the facial landmark detection and tracking algorithms can be classified into three major categories depending on how the algorithm captures the facial appearance and facial shape information. These categories include

holistic methods [1], Constrained Local Methods (CLM) [2], and regression-based methods[13]. Holistic methods explicitly build appearance and shape prior models to capture the global appearance and shape variations of facial images during training. CLMs build global face shape models and local appearance models for each facial landmark independently. The regression-based methods use appearance and shape information implicitly, and they do not explicitly build any appearance or shape model.

The limitation of the existing methods is that there is lack of robust algorithms that can handle facial images with “in-the-wild” conditions with significant facial expressions, head poses, facial occlusions, etc. This is mainly due to the following reasons:

- There is no effective model that can capture the face shape variations caused by the significant facial expression and head pose, which are coupled together to generate the 2D face shape. The existing algorithms usually rely on the simple linear Active Shape Model, which may not be effective for such a difficult task.
- There is no systematic way to handle facial occlusions caused by extreme head poses and object occlusions. Most of the existing facial landmark detection and tracking algorithms would assume that all the facial landmarks are visible and consider them equally. Thus, they tend to fail on images with facial occlusions.




To tackle these limitations and facilitate facial landmark detection and tracking in real-life conditions, we propose several facial landmark detection and tracking methods as discussed in the following sections.

2.2 Facial landmark detection and tracking under varying facial expressions and head poses

We propose two face shape models [11][12][7] to handle facial expression and head pose variations for robust facial landmark detection. The first face shape model is a directed hierarchical probabilistic face shape model [12]. The model incorporates two levels of information with the explicit help of the facial expression and head pose information. In the lower level, the model captures the facial shape variations of each facial component. In the higher level, the model automatically exploits the relationships among facial components, facial expression, and head pose through automatic model structure learning and parameter learning. Model structure learning and parameter learning are non-trivial due to the existence of latent variables. To alleviate this problem, we designed effective learning algorithms based on Structural EM algorithm [4].

Although the proposed hierarchical probabilistic face shape model is effective at improving the performance of facial landmark detection algorithms, its training requires the explicit annotations of the facial expression and head pose for each training image, which limits its application areas. To solve this problem, we propose the second face shape model, which is an undirected discriminative deep face shape model based on the factored three-way Restricted Boltzmann Machine model and deep learning techniques [11][7]. Instead of explicitly utilizing the expression and head pose labels,

Table 1: List of a selection of facial action units.

AU id	1	7	27
Description	Inner Brow Raiser	Outer Brow Raiser	Lid Tightener
Example image			

the second model implicitly decouples the face shape variations into expression-related parts and head pose-related parts with the help of additional frontal training images. Effective algorithms are proposed to facilitate model learning with incomplete data and model inference for the prediction of facial landmark locations.

We evaluated the proposed two probabilistic face shape models on both controlled images and general “in-the-wild” facial images. The experiments show the effectiveness of the proposed probabilistic face shape models for facial landmark detection comparing to state-of-the-art works [11][12][7].

2.3 Facial landmark detection under facial occlusion

To perform facial landmark detection under facial occlusion, we improve upon the cascade regression framework and propose an improved robust cascade regression method [8] [9]. The proposed method iteratively predicts the facial landmark locations and facial landmark visibility probabilities. Unlike existing methods, which use local appearance information from all facial landmarks to predict landmark locations, the proposed method relies on visible landmarks rather than the occluded landmarks based on their visibility probabilities. When predicting landmark occlusions, we explicitly learn the occlusion pattern with the Auto-Encoder (AE) model and embed it as a constraint to aid prediction. By considering the facial occlusion caused by extreme head poses as a special case of object occlusion, the proposed model can handle facial images with both object occlusions and extreme head poses, outperforming the existing methods, which can only handle one of them. Experiments on benchmark databases with object occlusion and extreme head poses demonstrate the state-of-the-art performance of the proposed method.

3. FACIAL ACTION UNIT RECOGNITION

3.1 Overview

Facial expression refers to the global facial deformation caused by different emotions. Typical facial expressions include anger, sadness, disgust, happiness, surprise, and fear. In contrast to the global facial expressions, the facial Action Units (AUs) defined in the Facial Action Coding System (FACS) [3] are more local and they characterize the local facial muscle movements. For example, Table 1 shows some typical AUs, their descriptions, and sample images. The goal of AU recognition is to detect the activation of particular AUs on facial images.

Facial action unit recognition methods contain two sequential steps: feature extraction and facial action recognition. For the first step, effective features such as the appearance features, the geometric features, or their combina-

tions are extracted to represent the facial images. In the estimation step, the methods recognize AU based on the features [5][6]. Early AU recognition methods consider each AU independently, while recent works focus on joint AU recognition by exploiting the relationships among AUs.

Even though facial action unit recognition has been studied for a few decades, a fully automatic system has not been developed. The existing facial action unit recognition algorithms suffer from the following limitations:

- Facial action unit recognition relies on the location of facial landmarks. In existing works, facial landmark detection is usually performed before facial action analysis. This sequential approach is suboptimal, as it ignores the interactions and joint relationships among facial action units and face shape.
- It is difficult for the existing algorithms to handle facial action unit recognition on facial images with spontaneous and natural movements. This is partially due to the failure of the facial landmark detection algorithms in those challenging cases.

To tackle these limitations of existing AU recognition algorithms, we propose a robust method that jointly performs facial landmark detection and action recognition, which we further discuss in the following section.

3.2 Joint facial landmark detection and facial action recognition

We propose the constrained joint cascade regression framework to improve facial action unit recognition [10]. Unlike existing methods which perform facial landmark detection and facial action unit recognition sequentially, the proposed method performs them simultaneously and jointly. The model first learns the relationship among facial action units and face shape as a constraint. It then iteratively updates the facial landmark locations and the AU activation probabilities until convergence. The experimental results on benchmark posed and spontaneous databases show that the proposed method can improve the performance of both facial landmark detection and facial action unit recognition.

4. HEAD POSE ESTIMATION

4.1 Overview

In computer vision, head pose estimation refers to the prediction of the head orientation and position with respect to the camera coordinate system. The orientation of head pose can be characterized by the pitch, yaw, and roll angles. In this research, we classify head pose estimation methods into *learning-based methods* and *model-based methods*. Learning-based methods utilize pattern recognition and machine learning techniques to estimate the head pose from facial data. Model-based methods utilize 3D computer vision techniques and projection models to estimate head pose, and they usually require facial landmark location information.

The existing head pose estimation methods still suffer from the following limitations:

- The existing head pose estimation algorithms have difficulty to handle facial occlusion. Existing methods usually rely on the holistic facial appearance or the detected facial landmark locations for all points, while

ignoring the fact that the facial appearance and the detected facial landmark locations on the occluded facial part are unreliable.

- Since the rigid head movements and the non-rigid facial motion are coupled together to generate the 2D facial images, non-rigid facial motion may affect accuracy of head pose estimation for existing methods. The field currently lacks an effective method to decouple the non-rigid motion before head pose estimation.
- The existing head pose estimation methods are either learning-based or model-based methods. There is no effective way to combine them to boost the performance of head pose estimation.

To tackle these limitations of existing automatic head pose estimation methods, we propose a robust landmark-based head pose estimation method, which is further discussed in the following section.

4.2 Robust head pose estimation

We are currently working on a robust algorithm that can perform head pose estimation under facial occlusion. The method is based on the previous work in [14], which uses linear and nonlinear methods to decouple rigid head motion from non-rigid facial motion and estimates the head poses based on the detected 2D facial landmark points. We propose to add landmark occlusion/confidence information to improve the robustness of the head pose estimation algorithm. The decoupled non-rigid facial motion information can also be used for facial expression recognition. In addition, we propose to combine the learning-based and model-based methods to improve the robustness and accuracy of head pose estimation, which is rarely exploited in the literature.

5. CONCLUSION

In this research, we proposed several robust and accurate algorithms for facial behavior analysis. We focused on three major research topics: facial landmark detection and tracking, facial action recognition, and head pose estimation. In the first part of the research, we proposed two probabilistic face shape models, including the hierarchical probabilistic face shape model and the discriminative deep face shape model to handle facial expressions and head poses for facial landmark detection. We also proposed a robust cascade regression framework for facial landmark detection under facial occlusion. In the second part of this research, we applied the landmark detection algorithms for facial action recognition and head pose estimation. In particular, we proposed a constrained joint cascade regression framework for simultaneous facial action recognition and facial landmark detection. We are working on a robust head pose estimation algorithm to handle facial occlusion.

In the future, the goal is to extend the proposed algorithms and models for other applications and research areas. Some potential research topics include: automatic facial landmark detection in the wild, vision-based human behavior analysis, human event and action recognition, and probabilistic graphical model based applications.

6. REFERENCES

- [1] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, jun 2001.
- [2] D. Cristinacce and T. F. Cootes. Feature detection and tracking with constrained local models. In *Proceedings of the British Machine Vision Conference*, 2006.
- [3] P. Ekman and E. L. Rosenberg, editors. *What the face reveals: basic and applied studies of spontaneous expression using the facial action coding system(FACS)*. Series in affective science. Oxford University Press, first edition, 1997.
- [4] N. Friedman. Learning belief networks in the presence of missing values and hidden variables. In *Proceedings of the Fourteenth International Conference on Machine Learning*, pages 125–133, 1997.
- [5] B. Jiang, M. Valstar, and M. Pantic. Action unit detection using sparse appearance descriptors in space-time video volumes. In *Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 314–321, March 2011.
- [6] Z. Wang, Y. Li, S. Wang, and Q. Ji. Capturing global semantic relationships for facial action unit recognition. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 3304–3311, Dec 2013.
- [7] Y. Wu and Q. Ji. Discriminative deep face shape model for facial point detection. *International Journal of Computer Vision*, 113(1):37–53, 2015.
- [8] Y. Wu and Q. Ji. Robust facial landmark detection under significant head poses and occlusion. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2015.
- [9] Y. Wu and Q. Ji. Shape augmented regression method for face alignment. In *The IEEE International Conference on Computer Vision (ICCV) Workshops*, December 2015.
- [10] Y. Wu and Q. Ji. Constrained joint cascade regression framework for simultaneous facial action unit recognition and facial landmark detection. In *Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on*, 2016.
- [11] Y. Wu, Z. Wang, and Q. Ji. Facial feature tracking under varying facial expressions and face poses based on restricted boltzmann machines. In *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3452–3459, 2013.
- [12] Y. Wu, Z. Wang, and Q. Ji. A hierarchical probabilistic model for facial feature detection. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 1781–1788, June 2014.
- [13] X. Xiong and F. De la Torre Frade. Supervised descent method and its applications to face alignment. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, May 2013.
- [14] Z. Zhu and Q. Ji. Robust real-time face pose and facial expression recovery. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 681–688. IEEE, 2006.