



Contents lists available at ScienceDirect

## Computers &amp; Graphics

journal homepage: [www.elsevier.com/locate/cag](http://www.elsevier.com/locate/cag)

Special Issue on CAD/Graphics 2017

## Better initialization for regression-based face alignment

Hengliang Zhu<sup>a,\*</sup>, Bin Sheng<sup>a</sup>, Zhiwen Shao<sup>a</sup>, Yangyang Hao<sup>a</sup>, Xiaonan Hou<sup>a</sup>, Lizhuang Ma<sup>a,b</sup><sup>a</sup> Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China<sup>b</sup> Department of Computer Science and Software Engineering, East China Normal University, Shanghai, China

## ARTICLE INFO

## Article history:

Received 15 June 2017

Revised 27 July 2017

Accepted 29 July 2017

Available online xxx

## Keywords:

Neighborhood representation prior

Occlusions

Projected initial shape

Cascade regression

## ABSTRACT

Regression-based face alignment algorithms predict facial landmarks by iteratively updating an initial shape, and hence are always limited by the initialization. Usually, the initial shape is obtained from the average face or by randomly picking a face from the training set. In this study, we discuss how to improve initialization by studying a neighborhood representation prior, leveraging neighboring faces to obtain a high-quality initial shape. In order to further improve the estimation precision of each facial landmark, we propose a face-like landmark adjustment algorithm to refine the face shape. Extensive experiments demonstrate our algorithm achieves favorable results compared to the state-of-the-art algorithms. Moreover, our algorithm achieves a smaller normalized mean error than the human performance (5.54% vs. 5.6%) on the challenging dataset the Caltech Occluded Faces in the Wild (COFW).

© 2017 Elsevier Ltd. All rights reserved.

## 1. Introduction

Face alignment is defined as the localization of facial landmarks such as eyebrows, eyes, nose tip and mouth corners. Efficient face alignment is critical in multimedia applications and often regarded as a pre-processing step for many vision tasks, such as face recognition [1,2], expression analysis [3–5], face makeup [6,7], and 3D face modeling [8–10]. However, accurate and robust landmark localization remains a big challenge for in-the-wild images that contain severe occlusions and large head rotations. In most regression-based algorithms, the bottleneck of this problem is the initialization, since a poor initial shape may subject face alignment into local optimum. The conventional algorithms trivially make use of the mean shape as initialization during testing, or randomly choose shapes from the training images during training [11–13]. Due to large occlusions and pose variations, poor initialization may cause the failure of face alignment. In this study, we aim to exploit a simple yet effective initialization algorithm that deals with these complex face images.

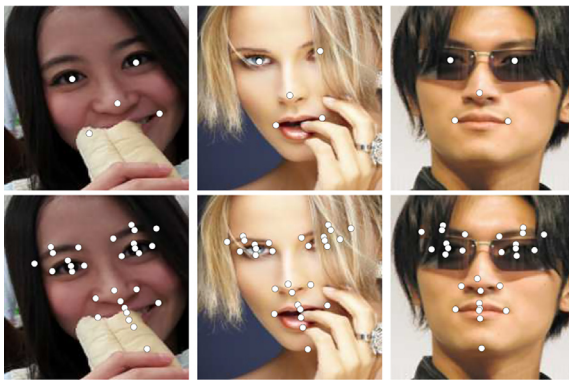
Recently, some algorithms have been proposed to improve the initial face shape [14–19], but they fail in the case of heavy occlusions and complex variations of poses and expression. These algorithms [14,16] use 3D information or other involved processes to get the initial shape. However, these algorithms cannot meet

real-time applications requirements due to complexity and time consumption. Therefore, efficient initialization is still a challenge in cascade regression. Since a face can be approximated by a linear combination of similar faces (Fig. 2(a)), finding a good initialization algorithm means to search for the right facial neighborhood and compute how it should be weighted. However, it is difficult to find the neighborhood of a face, since its landmarks positions are unknown initially. To deal with this problem, we contribute our study, which leverages a sparse set of points to approximately find neighboring faces, which estimate 5 key points: two pupils, a nose tip, and two mouth corners. It is worth noticing that these facial points can approximate the face shape and pose (Fig. 1). A subset of multiple points (e.g. the aforementioned points) is easier to obtain and more robust to occlusions [20]. As shown in Table 2, the mean errors of the 5-point set are lower than the 68-point set. The reason is that these sparse points are the most prominent on a face and can achieve good performance even in variations of occlusion, pose and expressions. Furthermore, we found that the neighboring faces and weight coefficients on the proposed subset can be well applied to the full set of facial landmarks (Fig. 2(b)). Given those experiments, we propose a neighborhood representation prior: *Based on the face similarity computed on the subset, we can approximate the multi-point shape well using its neighboring faces.*

More specifically, the point information of the subset can be used to generate a high-quality initial shape, which reduces the errors of facial landmark detection. Based on the above observations, we propose a simple yet efficient algorithm to produce a desirable initialization, called the projected initial shape (PIS). We also

\* Corresponding author.

E-mail addresses: [hengliang\\_zhu@sjtu.edu.cn](mailto:hengliang_zhu@sjtu.edu.cn) (H. Zhu), [ma-lz@cs.sjtu.edu.cn](mailto:ma-lz@cs.sjtu.edu.cn) (L. Ma).



**Fig. 1.** Some experimental results on the COFW dataset and the images with heavy occlusions.

propose a face-like landmark adjustment algorithm for shape fine-tuning and final initialization. After that, the location of key facial points is improved in detail. The PIS can significantly improve both accuracy and robustness of face alignment. Experimental results show our new strategy is more efficient in dealing with heavy occlusions and large pose variations.

In this paper, we handle the problem of finding initial face for cascade regression-based algorithms. Our main contributions are summarized as follows:

- We propose a neighborhood representation prior to approximate the target face shape. The experimental results demonstrate the effectiveness of our assumptions.
- An efficient face alignment algorithm, utilizing the neighborhood representation prior, and a face-like landmark adjustment algorithm, is used to generate a better initial shape.
- Extensive experiments demonstrate the efficiency and robustness of our initialization scheme.

The paper is organized as follows. Section 2 reviews the related work. Then, the detail of the neighborhood representation prior is presented in Section 3. Section 4 depicts the details of our algorithm, and Section 5 demonstrates the experimental results and discussions. Finally, the conclusion is given in Section 6.

## 2. Related work

In this section, we mainly review research related to our work. In recent years, a number of regression-based algorithms have been proposed and have become popular for detection of facial landmarks [21–23]. These algorithms learn the feature mapping function from image appearance to the final shape. The classic active appearance model (AAM) [21] uses the difference between the

**Table 1**

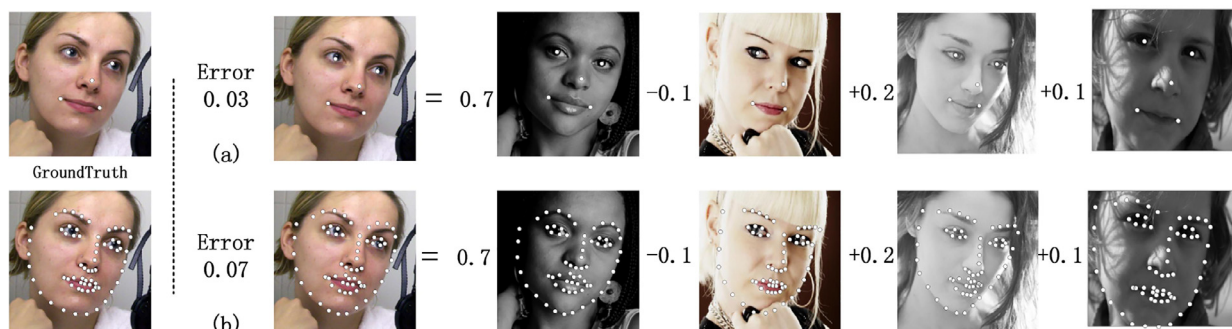
The speed and mean errors by the inter-ocular distance on 300W.

Result	Mean errors	Total time (ms)
Baseline + 5 points	5.58	4.48
Baseline + 8 points	6.12	4.80
Baseline + 14 points	6.10	5.08

current appearance estimate and the target image to drive an optimization problem. However, AAM is unfeasible for images with occlusions and large pose variations. Later, the fast AAM for face alignment was proposed [22]. Xiong and De la Torre [24] proposed a supervised descent algorithm to solve nonlinear least square problems based on scale-invariant feature transform (SIFT) [25]. Then a global supervised descent algorithm is proposed and performs well in facial tracking [26]. These algorithms may be difficult to handle complex scenarios, and fail to predict the landmarks accurately. Since feature descriptors that are extracted at occluded areas will greatly affect the update of the face shape at each iteration. It might result in a shape that is far away from the true landmarks.

A new cascaded shape regression (CSR) in face alignment is highly efficient in both training and testing. CSR algorithms use the image features to estimate the facial points in a cascaded way. For example, an explicit shape regression algorithm for face landmark location contains two-level regressors for shape estimation [11]. Kazemi and Sullivan [12] detected the landmarks by using an ensemble of gradient regression trees. Ren et al. [13] further improved this algorithm [11] and designed local binary features for shape regression. Burgos-Artizzu et al. [27] designed an occlusion-invariant face alignment algorithm that used shape indexed features and detected occlusions explicitly. Lee et al. [28] used cascade Gaussian process regression trees (cGPRT) for face alignment by using shape-indexed difference of Gaussian features to achieve robustness against geometric variances of faces. Wu and Ji [29] proposed a robust cascaded regressor to handle large pose and severe occlusions. Based on explicit head pose estimation, Yang et al. [16] presented a supervised initialization scheme for cascaded face alignment. Deng et al. [30] proposed a multi-view, multi-scale, and multi-component cascade shape regression (M3CSR) model for landmark prediction. Yang et al. [19] developed a spatio-temporal cascade shape regression (STCSR) model for robust facial tracking.

Most of the aforementioned algorithms start from a mean shape, and optimize the face shape iteratively. However, with a poor or wrong initialization, regression-based algorithms usually trap into local optimum [31]. By taking advantages of the neighborhood coherence in face similarity, our algorithm is fast, and can generate a high-quality initial shape closer to the true location. In order to alleviate the sensitivity to initialization, we propose a face



**Fig. 2.** The information (neighboring faces and weights) attained from the subset can be applied to the full set and still be a good approximation. The front numbers are weight coefficients. The errors are the distance between the predicted points and the ground truth, normalized by the inter-ocular distance.

Download English Version:

<https://daneshyari.com/en/article/6876875>

Download Persian Version:

<https://daneshyari.com/article/6876875>

[Daneshyari.com](https://daneshyari.com)