Computer Vision and Image Understanding 162 (2017) 1–22

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Face alignment in-the-wild: A Survey

Xin Jin[a,b], Xiaoyang Tan[a,b,*]

[a] Department of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics, #29 Yudao Street, Nanjing 210016, PR China
[b] Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

ARTICLE INFO

Article history:
Received 14 August 2016
Revised 20 July 2017
Accepted 16 August 2017
Available online 24 August 2017

Keywords:
Face alignment
Active appearance model
Constrained local model
Cascaded regression
Deep convolutional neural networks

ABSTRACT

Over the last two decades, face alignment or localizing fiducial facial points on 2D images has received increasing attention owing to its comprehensive applications in automatic face analysis. However, such a task has proven extremely challenging in unconstrained environments due to many confounding factors, such as pose, occlusions, expression and illumination. While numerous techniques have been developed to address these challenges, this problem is still far away from being solved. In this survey, we present an up-to-date critical review of the existing literatures on face alignment, focusing on those methods addressing overall difficulties and challenges of this topic under uncontrolled conditions. Specifically, we categorize existing face alignment techniques, present detailed descriptions of the prominent algorithms within each category, and discuss their advantages and disadvantages. Furthermore, we organize special discussions on the practical aspects of face alignment *in-the-wild*, towards the development of a robust face alignment system. In addition, we show performance statistics of the state of the art, and conclude this paper with several promising directions for future research.

1. Introduction

Fiducial facial points refer to the predefined landmarks on a face graph, which are mainly located around or centered at the facial components such as eyes, mouth, nose and chin (see Fig. 1). Localizing these facial points, which is also known as face alignment, has recently received significant attention in computer vision, especially during the last decade. At least two reasons account for this. Firstly, many important tasks, such as face recognition, face tracking, facial expression recognition, head pose estimation, can benefit from precise facial point localization. Secondly, although some level of success has been achieved in recent years, face alignment in unconstrained environments is so challenging that it remains an open problem in computer vision, and continues to attract researchers to attack it.

While face detection is generally regarded as the starting point for all face analysis tasks (Ding and Martinez, 2010; Zafeiriou et al., 2015), face alignment can be regarded as *an important and essential intermediary step* for many subsequent face analyses that range from biometric recognition to mental state understanding. Concrete tasks may differ in the number and type of the needed facial points, as well as the way these points are used. Below we give some details on three typical tasks where face alignment plays a prominent role:

- *Face recognition:* Face alignment is widely used by face recognition algorithms to improve their robustness against pose variations. For example, in the stage of face registration, the first step is usually to locate some major facial points and use them as anchor points for affine warping, while other face recognition algorithms, such as feature-based (structural) matching (Campadelli et al., 2003; Zhao et al., 2003), rely on accurate face alignment to build the correspondence among local features (e.g, eyes, nose, mouth, etc.) to be matched.
- *Attribute computing:* Face alignment is also beneficial to facial attribute computing, since many facial attributes such as eyeglasses and nose shape are closely related to specific spatial positions of a face. In Kumar et al. (2009), six facial points are localized to compute qualitative attributes and similes that are then used for robust face verification in unconstrained conditions.
- *Expression recognition:* The configurations of facial points (typically between 20–60) are reliable indicative of the deformations caused by expressions, and the subsequent analysis will reveal the particular type of expression that may lead to such deformation. Many works (Bailenson et al., 2008; Li et al., 2015; Rudovic et al., 2010; Senechal et al., 2011; Valstar and Pantic,

* Corresponding author at: Department of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics, #29 Yudao Street, Nanjing 210016, PR China.
E-mail addresses: x.jin@nuaa.edu.cn (X. Jin), x.tan@nuaa.edu.cn (X. Tan).

http://dx.doi.org/10.1016/j.cviu.2017.08.008
1077-3142/

Fig. 1. Illustration of some example face images with 68 manually annotated points from the IBUG database (Sagonas et al., 2013a).

2012) follow this idea and use various features extracted from these points for expression recognition.

The above-mentioned applications, as well as numerous ones yet to be conceived, urge the need for developing robust and accurate face alignment techniques in real-life scenarios.

Under constrained environments or on less challenging databases, the problem of face alignment has been well addressed, and some algorithms even achieve performance that is close to that of human beings (Belhumeur et al., 2011; Dantone et al., 2012). Under unconstrained conditions, however, this task is extremely challenging and far from being solved, due to the high degree of facial appearance variability caused either by intrinsic dynamic features of the facial components such as eyes and mouth, or by ambient environment changes. In particular, the following factors have significant influence on facial appearance and the states of local facial features:

- *Pose:* The appearance of local facial features differ greatly between different camera-object poses (e.g., frontal, profile, upside down), and some facial components such as the one side of the face contour, can even be completely occluded in a profile face.
- *Occlusion:* For face images captured in unconstrained conditions, occlusion frequently happens and brings great challenges to face alignment. For example, the eyes may be occluded by hair, sunglasses, or myopia glasses with black frames.
- *Expression:* Some local facial features such as eyes and mouth are sensitive to the change of various expressions. For example, laughing may cause the eyes to close completely, and largely deform the shape of the mouth.
- *Illumination:* Lighting (varying in spectra, source distribution, and intensity), may significantly change the appearance of the whole face, and make the detailed textures of some facial components missing.

These challenges are illustrated in Fig. 2 by the IBUG database (Sagonas et al., 2013a). An ideal face alignment system should be robust to these facial variations on one hand; while on the other hand, as efficient as possible to satisfy the need of practical applications (e.g., real-time face tracking).

Over the last two decades, numerous techniques have been developed for face alignment with varying degrees of success. Çeliktutan et al. (2013) surveyed many traditional methods, but some recent state-of-the-art methods are not covered. Wang et al. (2014) gave a more comprehensive survey of face alignment methods over the last two decades, but the overall difficulties and challenges in unconstrained environments have not been highlighted. More recently, Yang et al. (2015) provided an empirical study of recent face alignment methods, aiming to draw some empirical yet useful conclusions and make insightful suggestions for practical applications.

The significant contribution of this paper is to give a comprehensive and critical survey of the ad hoc face alignment methods on 2D images, addressing the difficulties and challenges in unconstrained environments. We believe that it would be a useful complement to Çeliktutan et al. (2013), Wang et al. (2014) and Yang et al. (2015). But to be self-contained, some traditional methods covered in Çeliktutan et al. (2013) and Wang et al. (2014) are also included. However, contrary to the previous works, we add some state-of-the-art algorithms emerged recently (e.g., 3D face alignment methods), and pay special attention to study and summarize the motivation and successful experiences behind the state-of-the-art. Furthermore, we organize special discussions on the practical aspects of constructing a face alignment system, which in our opinion is a very important topic in practice, but is mostly ignored in previous studies. In addition, we show comparative performance statistics of the state of the art, and propose several promising directions for future research.

In Section 2, we briefly describe the main idea of face alignment and categorize existing methods into two main categories. Then, the prominent methods within each category are reviewed and analyzed in Sections 3 and 4. In Section 5, we investigate some practical aspects of developing of a robust face alignment system. In Section 6, we discuss a few issues concerning performance evaluation. Finally, we conclude this paper with a discussion of several promising directions for further research in Section 7.

2. Overview

The problem of face alignment on 2D images has a long history in computer vision. A large number of approaches have been proposed to tackle it with varying degrees of success. From an overall perspective, face alignment can be formulated as a problem of searching over a face image for the pre-defined facial points (also called facial landmarks, or face shape). It typically starts from a coarse initial shape, and proceeds by refining the shape estimate step by step until convergence. During the search process, two different sources of information are typically used: facial appearance and shape information. The latter aims to explicitly model the spatial relations between the locations of facial points to ensure that the estimated facial points can form a valid face shape. Although some methods make no explicit use of the shape information, it is common to combine these two sources of information.

Before describing specific and prominent algorithms, a clear and high-level categorization will help to provide a holistic understanding of the commonality and differences of existing methods in using the appearance and shape information. For this, we follow the basic modeling principles in pattern recognition, and roughly divide existing methods into two categories: *generative* and *discriminative*.

- *Generative methods:* These methods build generative models for both the face shape and appearance. They typically formulate face alignment as an optimization problem to find the shape and appearance parameters that generate an appearance model instance giving best fit to the test face. Note that the facial appearance can be represented either by the whole (warped) face, or by the local image patches centered at the facial points.
- *Discriminative methods:* These methods directly infer the target location from the facial appearance. This is typically done by learning independent local detector or regressor for each facial point and employing a global shape model to regularize their predictions, or by directly learning a vectorial regression function to infer the whole face shape, during which the shape constraint is implicitly encoded.

Table 1 summarizes algorithms and representative works for face alignment, where we further divide the generative methods and discriminative methods into several subcategories. A few methods overlap category boundaries, and are discussed at the end of the section where they are introduced. Below, we discuss the