Int J Comput Vis
DOI 10.1007/s11263-013-0667-3

Face Alignment by Explicit Shape Regression

Xudong Cao · Yichen Wei · Fang Wen · Jian Sun

Received: 28 December 2012 / Accepted: 7 October 2013

Abstract We present a very efficient, highly accurate, "Explicit Shape Regression" approach for face alignment. Unlike previous regression-based approaches, we directly learn a vectorial regression function to infer the whole facial shape (a set of facial landmarks) from the image and *explicitly* minimize the alignment errors over the training data. The inherent shape constraint is naturally encoded into the regressor in a cascaded learning framework and applied from coarse to fine during the test, without using a fixed parametric shape model as in most previous methods. To make the regression more effective and efficient, we design a two-level boosted regression, shape indexed features and a correlation-based feature selection method. This combination enables us to learn accurate models from large training data in a short time (20 min for 2,000 training images), and run regression extremely fast in test (15 ms for a 87 landmarks shape). Experiments on challenging data show that our approach significantly outperforms the state-of-the-art in terms of both accuracy and efficiency.

Keywords Face alignment · Shape indexed feature · Correlation based feature selection · Non-parametric shape constraint · Tow-level boosted regression

X. Cao (✉) · Y. Wei · F. Wen · J. Sun
Visual Computing Group, Microsoft Research Asia, Beijing, People's Republic of China
e-mail: xudongca@microsoft.com

Y. Wei
e-mail: yichenw@microsoft.com

F. Wen
e-mail: fangwen@microsoft.com

J. Sun
e-mail: jiansun@microsoft.com

1 Introduction

Face alignment or locating semantic *facial landmarks* such as eyes, nose, mouth and chin, is essential for tasks like face recognition, face tracking, face animation and 3D face modeling. With the explosive increase in personal and web photos nowadays, a fully automatic, highly efficient and robust face alignment method is in demand. Such requirements are still challenging for current approaches in unconstrained environments, due to large variations on facial appearance, illumination, and partial occlusions.

A face shape $S = [x_1, y_1, ..., x_{N_{\mathrm{fp}}}, y_{N_{\mathrm{fp}}}]^T$ consists of N_{fp} facial landmarks. Given a face image, the goal of face alignment is to estimate a shape S that is as close as possible to the true shape $\hat{S}$, i.e., minimizing

$$||S - \hat{S}||_2. \quad (1)$$

The alignment error in Eq. (1) is usually used to guide the training and evaluate the performance. However, during testing, we cannot directly minimize it as $\hat{S}$ is unknown. According to how S is estimated, most alignment approaches can be classified into two categories: *optimization-based* and *regression-based*.

Optimization-based methods minimize another error function that is correlated to (1) instead. Such methods depend on the goodness of the error function and whether it can be optimized well. For example, the AAM approach (Matthews and Baker 2004; Sauer and Cootes 2011; Saragih and Goecke 2007; Cootes et al. 2001) reconstructs the entire face using an appearance model and estimates the shape by minimizing the texture residual. Because the learned appearance models have limited expressive power to capture complex and subtle face image variations in pose, expression, and illumination, it may not work well on unseen faces. It is also