Track Facial Points in Unconstrained Videos

Xi Peng[1]
xipeng.cs@rutgers.edu
Qiong Hu[1]
qionghu.cs@rutgers.edu
Junzhou Huang[2]
jzhuang@uta.edu
Dimitris N. Metaxas[1]
dnm@cs.rutgers.edu

[1] Department of Computer Science
Rutgers University
New Jersey, USA

[2] Department of Computer Science
The University of Texas at Arlington
Texas, USA

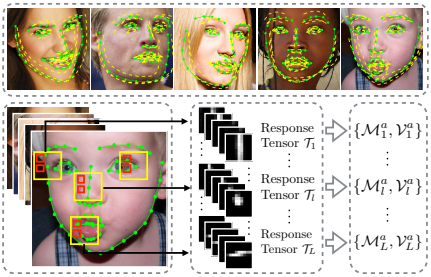

Figure 1: Part-based representation.

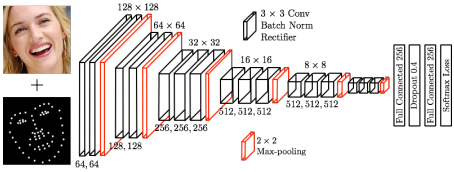

Figure 2: Deep fitting evaluation.

Tracking Facial Points in unconstrained videos is challenging due to the non-rigid deformation that changes over time. In this paper, we propose to exploit incremental learning for person-specific alignment in wild conditions.

Our approach takes advantage of part-based representation, as illustrated in Figure 1 and cascade regression for robust and efficient alignment on each frame. Unlike existing methods that usually rely on models trained offline, we incrementally update the representation subspace and the cascade of regressors in a unified framework to achieve personalized modeling.

Blind model adaptation without correction would inevitably result in model drifting. How to effectively detect misalignment is still a challenging question that is seldom investigated. To address this issue, we propose a deep neural network for robust fitting evaluation to pick out well-aligned faces from misalignment. The architecture of the network is shown is Figure 2. These well-aligned faces are then used to incrementally update the representation subspace and fitting strategy for person-specific modeling on the fly. In summary, our work makes the following contributions:

(1) We propose a novel approach for sequential face alignment. To the best of our knowledge, this is the first time that person-specific modeling is investigated to jointly learn the representation subspace and the fitting parameters in a unified framework.

(2) The proposed part-based representation together with the cascade regression guarantees robust alignment in unconstrained conditions. More importantly, they are critical to efficiently construct personalized models for real-time or large-scale applications.

(3) We propose to leverage deep neural networks for efficient and robust fitting evaluation. It significantly alleviates the drifting issue which would severely deteriorate learned models and inevitably lead to failure.

To validate our approach, we provide a detailed experimental analysis of each component of our approach, as well as performance comparisons with existing approaches. Four image datasets (MultiPIE, LFPW, Helen, and AFLW), four video datasets (FGNET, ASLV, 300-VW, and YtbVW), and four state of the arts (RLMS, DRMF, IFA, and ESR) are employed to conduct the experiments. The results demonstrate that the proposed incremental learning can significantly improve the fitting accuracy with an affordable computational cost, especially in unconstrained videos with extensive variations.