Unifying Holistic and Parts-Based Deformable Model Fitting

Joan Alabort-i-Medina[1], Stefanos Zafeiriou[1]
[1]Department of Computing, Imperial College London, United Kingdom.

Deformable model fitting has been an active area of research in computer vision for over 20 years. Fitting a deformable model consists of registering a parametric shape model to an image such that its landmarks accurately describe the shape of the object being modelled. Although a large variety of models and fitting strategies have been proposed throughout the years [4, 6, 8, 10, 11, 12, 13, 14], in general, research in this area can be divided into two different groups: (i) Holistic Deformable Models (HDMs) and (ii) Parts-Based Deformable Models (PBDMs). The main difference between both groups is the approach used to model object texture.

HDMs, such as Active Appearance Models (AAMs) [8, 10], define texture globally, typically, by means of a generative representation. Consequently, HDMs fitting strategies are generally posed as a regularized search for the optimal shape $\mathbf{p}$ and texture $\mathbf{c}$ parameters that minimize a *global* measure of misalignment that simultaneously depends on the position of all landmarks i.e.:

$$\mathbf{p}^*, \mathbf{c}^* = \arg\min_{\mathbf{p},\mathbf{c}} \mathcal{R}(\mathbf{p},\mathbf{c}) + \mathcal{D}(\mathbf{s}, I) \tag{1}$$

where $\mathcal{R}$ is a regularization term that penalizes complex shape and texture deformations and $\mathcal{D}$ is a data term that quantifies the global measure of misalignment given the current position of all landmark points $\mathbf{x}_i = (x_i, y_i)^T$ defining the shape $\mathbf{s} = (\mathbf{x}_1^T, \ldots, \mathbf{x}_v^T)^T$ of the object on the image I. HDMs are capable of producing very accurate fitting results [1, 3, 12]. However, the large dimensionality of their parameter space makes them difficult to optimize and likely to converge to undesirable local minima. Additionally, they are also highly sensitive to inaccurate initializations.

On the contrary, PBDMs, such as Constrained Local Models (CLMs) [7, 11], model texture locally as the combination of several independent local texture parts. PBDMs fitting strategies are commonly formulated as a regularized search for the optimal shape $\mathbf{p}$ parameters (local texture parts are usually learned discriminatively) that jointly minimize v *local* measures of misalignment dependent on each landmark $\mathbf{x}_i$ i.e:

$$\mathbf{p}^* = \arg\min_{\mathbf{p}} \mathcal{R}(\mathbf{p}) + \sum_{i=1}^{v} \mathcal{D}_i(\mathbf{x}_i, I) \tag{2}$$

where, in this case, $\mathcal{R}$ is a regularization term that penalizes only complex shape deformations and $\mathcal{D}_i$ are independent data terms quantifying the local misalignment measures given by the current position of each landmark $\mathbf{x}_i$ on the image I. PBDMs are generally easier to optimize than HDMs, less dependent on the initialization and better suited to handle partial occlusions due to their local nature [4, 5, 9, 14]. However, they are unable to match the accuracy of optimally fitted HDMs.

In this paper, we propose to overcome the previous limitations by unifying holistic and parts-based deformable model fitting. To this end, we derived a novel probabilistic formulation of the fitting problem that unifies HDMs and PBDMs. This new probabilistic formulation poses the problem of deformable fitting as a regularized search for the optimal shape $\mathbf{p}$ and texture $\mathbf{c}$ parameters that jointly minimize both a *global* misalignment measure that depends simultaneously on all landmarks and a set of v independent *local* measures of misalignment associated to each landmark, i.e.:

$$\mathbf{p}^*, \mathbf{c}^* = \arg\min_{\mathbf{p},\mathbf{c}} \mathcal{R}(\mathbf{p},\mathbf{c}) + \mathcal{D}(\mathbf{s}, I) + \sum_{i=1}^{v} \mathcal{D}_i(\mathbf{x}_i, I) \tag{3}$$

where $\mathcal{R}(\mathbf{p},\mathbf{c})$ corresponds to a regularization term that penalizes complex shape and texture deformations, $\mathcal{D}(\mathbf{s}, I)$ denotes the global misalignment measure and corresponds to the data term in HDMs fitting and $\sum_{i=1}^{n} \mathcal{D}_i(\mathbf{x}_i, \mathcal{I})$ denote the v local measures of misalignment which correspond to the data term in PBDMs fitting.

Our approach explicitly and optimally combines two of the most successful and well-established frameworks for HDMs and PBDMs fitting, i.e. AAMs and CLMs, using a unified Maximum A Posteriori (MAP) estimation framework. The result is a combined cost function that can be iteratively minimized using a variation of the Gauss-Newton algorithm in which the solution at each iteration is given by an optimal, in a MAP sense, weighted combination of the original AAMs and CLMs iterative solutions.

We show that our unified approach combines the advantages of both HDMs and PBMs and considerably outperforms the accuracy of AAMs and CLMs on the problem of face alignment in-the-wild by a large margin. Furthermore, we show that our unified approach, trained using a relatively small amount of training data, can compete and even surpass the accuracy of two of the most recently proposed state-of-the-art techniques in face alignment in-the-wild [12, 13] potentially trained with thousands of training examples. An open-source implementation of the proposed method will be made available as part of the Menpo Project [2] `http://www.menpo.org/`.

[1] Joan Alabort-i-Medina and Stefanos Zafeiriou. Bayesian active appearance models. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[2] Joan Alabort-i-Medina, Epameinondas Antonakos, James Booth, Patrick Snape, and Stefanos Zafeiriou. Menpo: A comprehensive platform for parametric image alignment and visual deformable models. In *ACM International Conference on Multimedia (ACMM)*, 2014.

[3] Epameinondas Antonakos, Joan Alabort-i-Medina, Georgios Tzimiropoulos, and Stefanos Zafeiriou. HOG active appearance models. In *International Conference on Image Processing (ICIP)*, 2014.

[4] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic. Robust discriminative response map fitting with constrained local models. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.

[5] Tadas BaltruÅąaitis, Peter Robinson, and Louis-Philippe Morency. Continuous conditional neural fields for structured regression. In *European Conference on Computer Vision (ECCV)*, 2014.

[6] Xudong Cao, Yichen Wei, Fang Wen, and Jian Sun. Face alignment by explicit shape regression. In *Computer Vision and Pattern Recognition (CVPR)*, 2012.

[7] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models: Their training and application. *Computer Vision and Image Understanding*, 1995.

[8] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2001.

[9] Pedro Martins, Rui Caseiro, and Jorge Batista. Non-parametric bayesian constrained local models. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[10] Iain Matthews and Simon Baker. Active appearance models revisited. *International Journal of Computer Vision (IJCV)*, 2004.

[11] Jason M. Saragih, Simon Lucey, and Jeffrey F. Cohn. Deformable model fitting by regularized landmark mean-shift. *International Journal of Computer Vision (IJCV)*, 2011.

[12] Georgios Tzimiropoulos and Maja Pantic. Gauss-newton deformable part models for face alignment in-the-wild. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[13] Xuehan-Xiong and Fernando De la Torre. Supervised descent method and its application to face alignment. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.

[14] Xiangxin Zhu and D Ramanan. Face detection, pose estimation, and landmark localization in the wild. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.