Unifying Holistic and Parts-Based Deformable Model Fitting

Joan Alabort-i-Medina Stefanos Zafeiriou
Department of Computing, Imperial College London, United Kingdom
{ja310, s.zafeiriou}@imperial.ac.uk

Abstract

The construction and fitting of deformable models that capture the degrees of freedom of articulated objects is one of the most popular areas of research in computer vision. Two of the most popular approaches are: Holistic Deformable Models (HDMs), which try to represent the object as a whole, and Parts-Based Deformable Models (PBDMs), which model object parts independently. Both models have been shown to have their own advantages. In this paper we try to marry the previous two approaches into a unified one that potentially combines the advantages of both. We do so by merging the well-established frameworks of Active Appearance Models (holistic) and Constrained Local Models (part-based) using a novel probabilistic formulation of the fitting problem. We show that our unified holistic and part-based formulation achieves state-of-the-art results in the problem of face alignment in-the-wild. Finally, in order to encourage open research and facilitate future comparisons with the proposed method, our code will be made publicly available to the research community[1].

1. Introduction

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 [12, 21, 28, 8, 37, 11, 4, 35, 33], 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) [12, 21], 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 $\mathbf{s} = (\mathbf{x}_1^T, \dots, \mathbf{x}_v^T)^T$ of the object on the image I. HDMs are capable of producing very accurate fitting results [2, 33, 3]. 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) [14, 28], 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 [37, 4, 20, 7]. 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

[1] An open-source implementation of the proposed method will be made available as part of the Menpo Project [1] http://www.menpo.org/.

formulation of the fitting problem that unifies Active Appearance Models and Constrained Local Models. Our approach explicitly and optimally combines two different statistically learned appearance models, i.e. holistic (AAMs generative appearance model) and parts-based (CLMs discriminative appearance model) with a single statistically learned 2d shape model (2d point distribution model) using a Maximum A Posteriori (MAP) estimation framework. The result is a unified cost function that can be iteratively solved using a variation of the Gauss-Newton algorithm and 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.

Summarizing, our main contributions are:

- To propose a generic probabilistic interpretation of deformable model fitting that effectively unifies previous holistic and parts-based approaches. In particular, we show that our novel formulation naturally leads to an optimal combination of two of the most successful and well-established frameworks for object alignment, namely AAMs and CLMs.
- To report a comprehensive performance comparison between the most popular algorithms for fitting AAMs and CLMs on the most recently collected datasets for facial alignment.
- To show that our unified approach can match and even surpass the accuracy of state-of-the-art techniques [33, 35], trained on thousands of images, on the problem of face alignment in-the-wild.

The reminder of the paper is structured as follows. Section 2 reviews prior work on face alignment. Section 3 and Section 4 introduce AAMs and CLMs and revisit their respective probabilistic interpretations. Our unified holistic and part-based approach is proposed in Section 5. Implementation details are discussed in Section 6 and experimental results reported in Section 7. Finally, conclusions are drawn in Section 8.

2. Prior Work

In this section, we review prior work on deformable model fitting.

Holistic Deformable Models are global statistical representations of the shape and texture of a particular object. These models were popularized by the seminal works of Cootes *et al*. [12] and Blanz and Vetter [9]. They have a long tradition in computer vision and have been widely used to solve the problems of object alignment, object landmark localization and deformable object tracking.

Cootes *et al*. [12] proposed to fit AAMs by learning a fixed linear relation from texture differences (between the image and the global texture model) to shape parameters. The authors of [27] and [30] extended this approach by using boosted regression considerably improving the accuracy and convergence of the results. Cootes and Taylor [13] showed that the use of non-linear image features for texture representation lead to better fitting performance in AAMs. On the other hand, Blanz and Vetter [9] use stochastic gradient descent to fit 3D Morphable Models (3DMMs) to images. In [21], Matthews and Baker proposed a general and efficient gradient descent framework for fitting AAMs. The authors of [19] extended the previous fitting framework to the Fourier domain increasing the robustness of the original formulation.

HDMs have been criticized due to the limited representational power of their global texture representation. However, recent works in this area [31, 19, 33, 3] suggest that this limitation might have been over-stressed in the literature and that global texture models can produce highly accurate results if appropriate training data [32], image representations [31, 19, 33, 3] and fitting strategies [31, 32] are employed.

Part-Based Deformable Models where introduced by Cootes *et al*. in [14] and re-popularized by the work of Saragih *et al*. [28]. PBDMs define object texture as the combination of several independent parts typically constrained by a global shape representation. Just like HDMs, PBDMs have enjoyed a long-standing popularity in computer vision.

Cootes *et al*. [14] proposed an iterative search procedure for fitting Active Shape Models (ASMs) that approximated local texture responses with isotropic Gaussian estimators. Saragih *et al*. [28] derived a probabilistic interpretation of CLMs fitting and extended the previous approach by assuming a non-parametric representation of the local response maps. This approach was later extended by introducing a sample specific shape prior in [26] and a non-parametric shape prior in [8]. The authors of [17] proposed to used several independent PCA priors to model the object's shape. On the other hand, Asthana *et al*. [4] proposed a robust cascade-regression approach for fitting CLMs. Very recently, Martins *et al*. [20] proposed a CLMs fitting strategy that defines a non-Gaussian posterior distribution and performs inference via efficient Regularized Particle Filters (RPF). The authors of [7] have recently proposed the use of Continuous Conditional Neural Fields (CCNF) to learn local patch experts. Finally, Zhu and Ramanan [37] used a tree-based shape model to derive a parts-based fitting strategy that could be simultaneously used for detection, pose estimation and landmark localization.

Direct Regression Techniques can be used to solve the same type of problems solved by deformable model fitting. These techniques attempt to solve these problems directly without the use of an explicit deformable model. To this end, they generally try to learn a direct function mapping between image appearance and the position of the landmarks in the image.

Cao et. al [11] proposed a two-level cascade boosted regression approach for learning the mapping between shape-indexed image features (computed across the whole image for each landmark) and landmark positions. Similarly, Xiong and De la Torre [35] proposed to learn a cascade of linear regressors from local image descriptors (SIFT extracted around each landmark) to landmark positions. The authors of [10] explicitly incorporated information regarding the visibility of the landmarks into a similar cascade regression framework. Very recently Asthana *et al*. [5], Kazemi *et al*. [15] and Ren *et al*. [24] have augmented the original cascade regression framework of [11] by proposing an incremental algorithm for cascade regression learning, substituting linear regressors by ensembles of regression trees and by learning extremely descriptive binary features using regression forests respectively. On the other hand, in [36] a method to effectively combine multiple landmark hypothesis using structured-SVMs was used to boost the results obtained with the previous cascade-regression frameworks. Finally, the authors of [25] and [29] used Kernel Ridge Regression (KRR) and Deep Convolutional Neural Network (DCNN), respectively, to learn the mapping from an entire object image (defined by the rectangular box obtained from object detection) to landmark positions.

Combined HDMs and PBDMs. The idea of combining HDMs and PBDMs was previously explored in [34]. The authors of [34] proposed to combine a 3d shape model with several view-dependent ASMs for solving the problem of face tracking. The algorithm alternates between: 1) fitting a view-dependent ASMs (selected using the current pose of the 3d shape model) and 2) regularizing the position of the ASMs result by making it consistent with its most likely 3d shape model projection onto the image plane using statistical inference.

On the other hand, our approach explicitly combines a holistic generative appearance model (AAMs apperance model) and a parts-based discriminative appearance model (CLMs apperance model) together with a single 2d shape model (2d point distribution model) to define a novel probabilistic formulation of the face alignment problem. The result is a unified fitting cost function that can be iteratively solved using a variant 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.

3. Active Appearance Models

Active Appearance Models (AAMs) [12, 21] are HDMs that define the shape and texture of a particular object as a linear combination of a set of bases.

Their shape model is built from a collection of (manually) annotated landmarks describing the object's shape. These landmarks are normalized with respect to a global similarity transform (typically using Procrustes Analysis) and Principal Component Analysis (PCA) is applied to obtained a set of linear shape bases. The shape model can be mathematically expressed as:

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{S}\mathbf{p} \tag{3}$$

where $\bar{\mathbf{s}} \in \mathcal{R}^{2v\times 1}$ is the mean shape, and $\mathbf{S} \in \mathcal{R}^{2v\times n}$ and $\mathbf{p} \in \mathcal{R}^{n\times 1}$ denote the shape eigenvectors and shape parameters respectively. In order to allow a particular shape instance $\mathbf{s}$ to be arbitrarily position on the image frame, the previous model is composed with a 2D global similarity transform. This results in the following expression for each landmark point $\mathbf{x}_i = (x_i, y_i)^T \in \mathcal{R}^{2\times 1}$ of the shape model:

$$\mathbf{x}_i = s\mathbf{R}\left(\bar{\mathbf{x}}_i + \mathbf{X}_i\mathbf{p}\right) + \mathbf{t} \tag{4}$$

where s, $\mathbf{R} \in \mathcal{R}^{2\times 2}$, $\mathbf{t} \in \mathcal{R}^{2\times 1}$ denote the scale, rotation and translation parameters of the global similarity transform. Using the orthonormalization procedure described in [21] the final expression for the shape model can be compactly written using 3. Where $\mathbf{S} = (\mathbf{s}_1^*, \ldots, \mathbf{s}_4^*, \mathbf{s}_1, \cdots, \mathbf{s}_n) \in \mathcal{R}^{2v\times(4+n)}$ and $\mathbf{p} = (p_1^*, \ldots, p_4^*, p_1, \ldots, p_n)^T \in \mathcal{R}^{(4+n)\times 1}$ are redefined as the concatenation of the similarity bases $\mathbf{s}_i^*$ and similarity parameters p_i^* with the original $\mathbf{S}$ and $\mathbf{p}$ respectively.

Their texture model is obtained by warping the object's texture onto a common reference frame (typically defined in terms of the previous mean shape $\bar{\mathbf{s}}$) and applying PCA onto the vectorized warped textures. Mathematically, the texture model is defined by the following expression:

$$\mathbf{a} = \bar{\mathbf{a}} + \mathbf{A}\mathbf{c} \tag{5}$$

where $\mathbf{a} \in \mathcal{R}^{F\times 1}$ is the mean texture, and $\mathbf{A} \in \mathcal{R}^{F\times m}$ and $\mathbf{c} \in \mathcal{R}^{m\times 1}$ denote the texture eigenvectors and texture parameters respectively.

Given a particular image I, warping function $\mathcal{W}$ and object shape $\mathbf{s}$, the two main assumptions behind AAMs are: (1) the object shape can be well approximated by the linear shape model and (2) the object texture can be well approximated by the linear texture model after warping the image region defined by the shape model's approximation of $\mathbf{s}$ onto the reference frame it *i.e.*:

$$\begin{aligned} \mathbf{s} &\approx \bar{\mathbf{s}} + \mathbf{S}\mathbf{p} \\ \mathbf{i}[\mathbf{p}] &\approx \bar{\mathbf{a}} + \mathbf{A}\mathbf{c} \end{aligned} \tag{6}$$

where $\mathbf{i}[\mathbf{p}] = \text{vec}(I(\mathcal{W}(\mathbf{p}))$ denotes the vectorized version of the warped image. Note that the warp $\mathcal{W}$, which depends on the shape parameters $\mathbf{p}$, implicitly relates the shape and texture models and it is a central part of the AAMs formulation.

3.1. Fitting Active Appearance Models

AAMs fitting can be defined as a regularized search over the model shape and texture parameters that minimize the Sum of Squared Differences (SSD) between the vectorized warped image and the linear texture model:

$$\begin{aligned}\mathbf{p}^*, \mathbf{c}^* &= \arg\min_{\mathbf{p},\mathbf{c}} \mathcal{R}(\mathbf{p},\mathbf{c}) + \mathcal{D}(I,\mathbf{p},\mathbf{c}) \\ &= \arg\min_{\mathbf{p},\mathbf{c}} \underbrace{||\mathbf{p}||^2_{\mathbf{\Lambda}^{-1}} + ||\mathbf{c}||^2_{\mathbf{\Sigma}^{-1}}}_{\mathcal{R}(\mathbf{p},\mathbf{c})} + \\ &\qquad \underbrace{\frac{1}{\sigma^2}||\mathbf{i}[\mathbf{p}] - \bar{\mathbf{a}} + \mathbf{A}\mathbf{c}||^2}_{\mathcal{D}(I,\mathbf{p},\mathbf{c})}\end{aligned} \tag{7}$$

where $\mathbf{\Lambda}$ and $\mathbf{\Sigma}$ are diagonal matrices containing the eigenvalues associated to shape and texture eigenvectors and σ^2 is the estimated image noise[2].

3.1.1 Probabilistic Interpretation

A probabilistic interpretation of 7 can be obtained by assuming the following probabilistic generative models of shape and texture:

$$\begin{aligned}&\hat{\mathbf{s}} \approx \bar{\mathbf{s}} + \mathbf{S}\mathbf{p} && \mathbf{p} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Lambda}) \\ &\mathbf{i}[\mathbf{p}] = \bar{\mathbf{a}} + \mathbf{A}\mathbf{c} + \boldsymbol{\epsilon} && \mathbf{c} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}) \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})\end{aligned} \tag{8}$$

Given the model parameters $\Theta = \{\bar{\mathbf{s}}, \mathbf{S}, \mathbf{\Lambda}, \bar{\mathbf{a}}, \mathbf{A}, \mathbf{\Sigma}, \sigma^2\}$, one can easily define a Maximum Likelihood (ML) procedure to infer the optimal shape $\mathbf{p}^*$ and texture $\mathbf{c}^*$ parameters:

$$\begin{aligned}\mathbf{p}^*, \mathbf{c}^* &= \arg\max_{\mathbf{p},\mathbf{c}} p(\mathbf{i}[\mathbf{p}]|\mathbf{p},\mathbf{c},\Theta) \\ &= \arg\max_{\mathbf{p},\mathbf{c}} \ln p(\mathbf{i}[\mathbf{p}]|\mathbf{p},\mathbf{c},\Theta) \\ &= \arg\min_{\mathbf{p},\mathbf{c}} \underbrace{-\ln p(\mathbf{i}[\mathbf{p}]|\mathbf{p},\mathbf{c},\Theta)}_{\mathcal{D}(I,\mathbf{p},\mathbf{c})} \\ &= \arg\min_{\mathbf{p},\mathbf{c}} \underbrace{\frac{1}{\sigma^2}||\mathbf{i}[\mathbf{p}] - (\bar{\mathbf{a}} + \mathbf{A}\mathbf{c})||^2}_{\mathcal{D}(I,\mathbf{p},\mathbf{c})}\end{aligned} \tag{9}$$

A Maximum A Posteriori (MAP) procedure can be obtained by taking into account the prior distribution over the shape $\mathbf{p} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Lambda})$ and texture parameters $\mathbf{c} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$:

$$\begin{aligned}\mathbf{p}^*, \mathbf{c}^* &= \arg\max_{\mathbf{p},\mathbf{c}} p(\mathbf{p},\mathbf{c},\mathbf{i}[\mathbf{p}]|\Theta) \\ &= \arg\max_{\mathbf{p},\mathbf{c}} p(\mathbf{p}|\mathbf{\Lambda})p(\mathbf{c}|\mathbf{\Sigma})p(\mathbf{i}[\mathbf{p}]|\mathbf{p},\mathbf{c},\Theta) \\ &= \arg\max_{\mathbf{p},\mathbf{c}} \ln p(\mathbf{p}|\mathbf{\Lambda}) + \ln p(\mathbf{c}|\mathbf{\Sigma}) + \\ &\qquad \ln p(\mathbf{i}[\mathbf{p}]|\mathbf{p},\mathbf{c},\Theta) \\ &= \arg\min_{\mathbf{p},\mathbf{c}} \underbrace{-\ln p(\mathbf{p}|\mathbf{\Lambda}) - \ln p(\mathbf{c}|\mathbf{\Sigma})}_{R(\mathbf{p},\mathbf{c})} + \\ &\qquad \underbrace{-\ln p(\mathbf{i}[\mathbf{p}]|\mathbf{p},\mathbf{c},\Theta)}_{D(\mathcal{I},\mathbf{p},\mathbf{c})} \\ &= \arg\min_{\mathbf{p},\mathbf{c}} \underbrace{||\mathbf{p}||^2_{\mathbf{\Lambda}^{-1}} + ||\mathbf{c}||^2_{\mathbf{\Sigma}^{-1}}}_{\mathcal{R}(\mathbf{p},\mathbf{c})} + \\ &\qquad \underbrace{\frac{1}{\sigma^2}||\mathbf{i}[\mathbf{p}] - (\bar{\mathbf{a}} + \mathbf{A}\mathbf{c})||^2}_{\mathcal{D}(I,\mathbf{p},\mathbf{c})}\end{aligned} \tag{10}$$

Note that the previous MAP formulation is equivalent to the optimization problem defined by 7; the maximization of the prior probability over the shape and texture parameters leads to the minimization of the regularization term $\mathcal{R}(\mathbf{p},\mathbf{c})$ and the maximization of the conditional probability of the vectorized warped image given the shape, texture and model parameters leads to the minimization of the data term $D(\mathcal{I},\mathbf{c},\mathbf{p})$.

Using the well-known Project-out Inverse Compositional (PIC) algorithm [6, 21] the optimal solution for $\mathbf{p}^*$ at each iteration is given by the following expression and update rule:

$$\begin{aligned}\delta\mathbf{p}^* &= -\mathbf{H}^{-1}(\mathbf{\Lambda}^{-1}\mathbf{p} - \frac{1}{\sigma^2}\mathbf{J}^T\mathbf{P}(\mathbf{i}[\mathbf{p}] - \bar{\mathbf{a}})) \\ \mathcal{W}(\mathbf{p}^*) &\leftarrow \mathcal{W}(\mathbf{p}) \circ \mathcal{W}(\delta\mathbf{p}^*)^{-1}\end{aligned} \tag{11}$$

where $\mathbf{H} = \mathbf{\Lambda}^{-1} + \frac{1}{\sigma^2}\mathbf{J}^T\mathbf{P}\mathbf{J}$ is the so-called inverse compositional Hessian and $\mathbf{J} = \nabla\bar{A}\frac{\partial\mathcal{W}}{\partial\mathbf{p}}\big|_{\mathbf{p}=\mathbf{0}}$ and $\mathbf{P} = \mathbf{I} - \mathbf{A}\mathbf{A}^T$ are the inverse compositional Jacobian and the project-out operator[3] respectively.

Alternatively, using an Alternating Inverse Compositional (AIC) algorithm similar to the one proposed in [23] the optimal value for $\mathbf{p}^*$ (and $\mathbf{c}^*$) is given by:

$$\begin{aligned}\mathbf{c}^* &= (\frac{1}{\sigma^2}\mathbf{A}^T\mathbf{A} + \mathbf{\Sigma}^{-1})^{-1}\frac{1}{\sigma^2}\mathbf{A}^T(\mathbf{i}[\mathbf{p}] - \bar{\mathbf{a}}) \\ \delta\mathbf{p}^* &= -\mathbf{H}^{-1}(\mathbf{\Lambda}^{-1}\mathbf{p} - \frac{1}{\sigma^2}\mathbf{J}^T(\mathbf{i}[\mathbf{p}] - \mathbf{a}_\mathbf{c})) \\ \mathcal{W}(\mathbf{p}^*) &\leftarrow \mathcal{W}(\mathbf{p}) \circ \mathcal{W}(\delta\mathbf{p}^*)^{-1}\end{aligned} \tag{12}$$

[2] Theoretically, the optimal value for σ^2 is the average value of the eigenvalues associated to the discarted texture eigenvectors *i.e.* $\sigma^2 = \frac{1}{M-m}\sum_{i=m}^{M}\lambda_{\mathbf{a},i}$ [22].

[3] Alabort-i-Medina and Zafeiriou [2] showed that the project-out operator $\mathbf{P}$ can be naturally derived from 10 by assuming a uniform prior over the texture parameters $\mathbf{c} \sim \mathcal{U}(-\infty, +\infty)$ and marginalizing out.

where $\mathbf{a_c} = \bar{\mathbf{a}} + \mathbf{A}\mathbf{c}^*$ and, in this case, the Hessian and Jacobian are defined as $\mathbf{H} = \mathbf{\Lambda}^{-1} + \frac{1}{\sigma^2}\mathbf{J}^T\mathbf{J}$ and $\mathbf{J} = \nabla(\bar{A} + \sum_{i=1}^{m} c_i A_i)\frac{\partial \mathcal{W}}{\partial \mathbf{p}}\big|_{\mathbf{p}=\mathbf{0}}$ respectively.

For further details on how to compute $\frac{\partial \mathcal{W}}{\partial \mathbf{p}}\big|_{\mathbf{p}=\mathbf{0}}$ and on warp composition and inversion [4] the interested reader is referred to [21], [23] and [33].

4. Constrained Local Models

Constrained Local Models (CLMs) [14, 28] are PBDMs that define the texture of a particular object by independently modelling the local image region around the landmarks defining its shape. They utilize a global PCA-based shape model similar to the one used by AAMs.

Although generative approaches can be used to model local image regions, the usual approach is discriminative. Hence, for each landmark a local patch expert that quantifies the probability of the landmark being correctly aligned $p(l_i = 1|\mathbf{x}_i, I)$ is learned based on the support of its local image region. The previous density can be defined as:

$$p(l_i = 1|\mathbf{x}_i, I) = \frac{1}{1 + \exp\{l_i\, \mathcal{C}_i(I, \mathbf{x}_i)\}} \tag{13}$$

where $\mathcal{C}_i$ is a (typically linear) classifier that discriminates between aligned and misaligned locations *i.e.*:

$$\mathcal{C}_i(I, \mathbf{x}_i) = \mathbf{w}_i\, [I(\mathbf{y}_i), \cdots, I(\mathbf{y}_m)] + b_i \tag{14}$$

and $\{\mathbf{y}_i\}_{i=1}^{k} \in \mathbf{\Omega}_{\mathbf{x}_i}$ denotes the image patch around the current landmark estimate $\mathbf{x}_i$.

Note that authors have used different types of classifiers $\mathcal{C}_i$ to differentiate aligned and misaligned landmarks *e.g.* Logistic Regression (LR) [28], Minimum Output Sum of Squared Errors (MOSSE) filters [20] and Support Vector Machines (SVM) [4] among others.

4.1. Fitting Constrained Local Models

Fitting CLMs involves solving the following optimization problem [28]:

$$\begin{aligned}\mathbf{p}^* &= \arg\min_{\mathbf{p}} \mathcal{R}(\mathbf{p}) + \sum_{i=1}^{v} \mathcal{D}_i(\mathbf{x}_i, I, \mathbf{p}) \\ &= \arg\min_{\mathbf{p}} \underbrace{||\mathbf{p}||^2_{\mathbf{\Lambda}^{-1}}}_{\mathcal{R}(\mathbf{p})} + \underbrace{\sum_{\mathbf{x}_i \in \mathbf{s}} \sum_{j=1}^{k} \frac{w_j}{\rho^2} ||\mathbf{x}_i - \mathbf{y}_j||^2}_{\sum_{i=1}^{v} \mathcal{D}_i(\mathbf{x}_i, I, \mathbf{p})}\end{aligned} \tag{15}$$

where $w_j = \frac{1}{1+\exp\{l_i\, \mathcal{C}_j(\mathcal{I}, \mathbf{x}_j)\}}$, $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues associated to the shape eigenvectors $\mathbf{S}$ and ρ^2 is the estimated shape noise[5].

[4] Depending on the type of warp $\mathcal{W}$ used, the solutions for $\delta\mathbf{p}^*$ might require the computation of the so called parameter Jacobian matrix $\mathbf{J}_{\tilde{p}}$ which converts inverse compositional incremental updates to its forwards additive first-order equivalent [23]

4.1.1 Probabilistic Interpretation

A probabilistic interpretation of 15 can be obtained assuming the following probabilistic generative model of shape:

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{S}\mathbf{p} + \boldsymbol{\varepsilon} \quad \mathbf{p} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Lambda}) \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \rho^2\mathbf{I}) \tag{16}$$

Denoting the previous model parameters by $\Phi = \{\bar{\mathbf{s}}, \mathbf{S}, \mathbf{\Lambda}, \rho^2\}$, CLMs fitting can be defined as a regularized search over the model shape parameters that jointly maximize the probability of all landmark being correctly aligned:

$$\begin{aligned}\mathbf{p}^* &= \arg\max_{\mathbf{p}} p(\mathbf{p}, \{l_i = 1\}_{i=1}^{v} | I, \mathbf{s}, \Phi) \\ &= \arg\max_{\mathbf{p}} p(\mathbf{p}|\mathbf{\Lambda}) p(\{l_i = 1\}_{i=1}^{v} | \mathbf{p}, I, \mathbf{s}, \Phi) \\ &= \arg\max_{\mathbf{p}} p(\mathbf{p}|\mathbf{\Lambda}) \prod_{i=1}^{v} p(l_i = 1|\mathbf{p}, I, \mathbf{x}_i, \Phi) \\ &= \arg\max_{\mathbf{p}} \ln p(\mathbf{p}|\mathbf{\Lambda}) + \\ &\qquad \ln\left(\prod_{i=1}^{v} p(l_i = 1|\mathbf{p}, I, \mathbf{x}_i, \Phi)\right)\end{aligned} \tag{17}$$

Different approaches [28] have been proposed to approximate the true response maps $p(l_i = 1|I, \mathbf{x}_i)$. The most popular one is the non-parametric approach of [28], which approximates the true response maps as:

$$\sum_{\mathbf{y}_j \in \mathbf{\Psi}_{\mathbf{x}_i}} p(l_i = 1|I, \mathbf{y}_i) \mathcal{N}(\mathbf{x}_i, \rho^2\mathbf{I}) \tag{18}$$

where the current landmark positions $\mathbf{x}_i$ are defined in terms of the previous probabilistic generative shape model.

Substituting 18 into 17 leads to the following optimization problem:

$$\begin{aligned}\mathbf{p}^* &= \arg\min_{\mathbf{p}} \underbrace{-\ln p(\mathbf{p}|\mathbf{\Lambda})}_{\mathcal{R}(\mathbf{p})} + \\ &\qquad \underbrace{-\sum_{i=1}^{v} \ln p(l_i = 1|\mathbf{p}, I, \mathbf{x}_i, \Phi)}_{\sum_{i=1}^{v} \mathcal{D}_i(\mathbf{x}_i, I, \mathbf{p})} \\ &= \arg\min_{\mathbf{p}} \underbrace{||\mathbf{p}||^2_{\mathbf{\Lambda}^{-1}}}_{\mathcal{R}(\mathbf{p})} + \\ &\qquad \underbrace{\sum_{\mathbf{x}_i \in \mathbf{s}} \sum_{\mathbf{y}_j \in \mathbf{\Psi}_{\mathbf{x}_i}} \frac{w_{y_j}}{\rho^2} ||\mathbf{x}_i - \mathbf{y}_j||^2}_{\sum_{i=1}^{v} \mathcal{D}_i(\mathbf{x}_i, I, \mathbf{p})}\end{aligned} \tag{19}$$

[5] The theoretically optimal value for ρ^2 can be computed as the average value of the eigenvalues associated to the discarded shape eigenvectors *i.e.* $\rho^2 = \frac{1}{N-n}\sum_{i=n}^{N} \lambda_{\mathbf{s},i}$ [22].

which is equivalent to the optimization problem defined by 15, where the response maps are evaluated at all pixels positions $\{\mathbf{y}_j\}_{j=1}^{k}$ of the local image patches $\mathbf{\Psi}_{\mathbf{x}_i}$.

Treating the true landmark positions $\mathbf{y}_i$ as latent variables, 19 can be solved iteratively using the EM-algorithm [28]. The solution for the optimal $\mathbf{p}^*$ is given by the following expressions at each iteration:

$$\begin{aligned}
\boldsymbol{\mu}_i &= \sum_{\mathbf{y}_j \in \mathbf{\Psi}_{x_i}} \frac{\pi_{y_j} \mathcal{N}(\mathbf{y}_j, \mathbf{x}_i, \rho^2 \mathbf{I})}{\sum_{\mathbf{z}_j \in \mathbf{\Psi}_{x_i}} \pi_{z_j} \mathcal{N}(\mathbf{z}_j, \mathbf{x}_i, \rho^2 \mathbf{I})} \mathbf{y}_j \\
\delta \mathbf{p}^* &= -(\mathbf{\Lambda}^{-1} + \frac{1}{\rho^2} \mathbf{J}^T \mathbf{J})^{-1} \\
&\quad (\mathbf{\Lambda}^{-1} \mathbf{p} - \frac{1}{\rho^2} \mathbf{J}^T (\boldsymbol{\mu} - \mathbf{s})) \\
\mathbf{p}^* &\leftarrow \mathbf{p} + \delta \mathbf{p}^*
\end{aligned} \tag{20}$$

where $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^T, \cdots, \boldsymbol{\mu}_n^T)^T$ and $\mathbf{J}$ is the Jacobian of the shape model. For a detailed derivation of the previous equations the interested reader is referred to [28].

5. Unifying Holistic and Parts-Based Deformable Model Fitting

In order to derive a fitting strategy that unifies the previous two frameworks we propose a novel probabilistic interpretation of the deformable model fitting problem. In particular, we redefine the problem as the maximization of the joint probability distribution of the shape $\mathbf{p}$ and texture parameters $\mathbf{c}$, the image I and the aligned landmarks $\{l_i = 1\}_{i=1}^{v}$ given the model parameters $\Delta = \{\Theta, \Phi\}$:

$$\begin{aligned}
\mathbf{p}^*, \mathbf{c}^* &= \arg\max_{\mathbf{p},\mathbf{c}} p(\mathbf{p}, \mathbf{c}, \mathbf{i}[\mathbf{p}], \{l_i = 1\}_{i=1}^{v} | \Delta) \\
&= \arg\max_{\mathbf{p},\mathbf{c}} p(\mathbf{p}|\mathbf{\Lambda})\, p(\mathbf{c}|\mathbf{\Sigma}) \\
&\qquad p(\mathbf{i}[\mathbf{p}], \{l_i = 1\}_{i=1}^{v} | \mathbf{p}, \mathbf{c}, \Delta) \\
&= \arg\max_{\mathbf{p},\mathbf{c}} p(\mathbf{p}|\mathbf{\Lambda})\, p(\mathbf{c}|\mathbf{\Sigma}) \\
&\qquad p(\mathbf{i}[\mathbf{p}]|\mathbf{p}, \mathbf{c}, \Theta) \\
&\qquad p(\{l_i = 1\}_{i=1}^{v} | I, \mathbf{p}, \Phi) \\
&= \arg\max_{\mathbf{p},\mathbf{c}} \ln p(\mathbf{p}|\mathbf{\Lambda}) + \ln p(\mathbf{c}|\mathbf{\Sigma}) + \\
&\qquad \ln p(\mathbf{i}[\mathbf{p}]|\mathbf{p}, \mathbf{c}, \Theta) + \\
&\qquad \ln p(\{l_i = 1\}_{i=1}^{v} | I, \mathbf{p}, \Phi)
\end{aligned} \tag{21}$$

The previous 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.*:

$$\begin{aligned}
\mathbf{p}^*, \mathbf{c}^* = \arg\min_{\mathbf{p},\mathbf{c}} & \underbrace{-\ln p(\mathbf{p}|\mathbf{\Lambda}) - \ln p(\mathbf{c}|\mathbf{\Sigma})}_{\mathcal{R}(\mathbf{p},\mathbf{c})} + \\
& \underbrace{-\ln p(\mathbf{i}[\mathbf{p}]|\mathbf{p}, \mathbf{c}, \Theta)}_{\mathcal{D}(I,\mathbf{p},\mathbf{c})} + \\
& \underbrace{-\ln p(\{l_i = 1\}_{i=1}^{v} | I, \mathbf{p}, \Theta)}_{\sum_{i=1}^{v} \mathcal{D}_i(\mathbf{x}_i, I, \mathbf{p})}
\end{aligned} \tag{22}$$

where $\mathcal{R}(\mathbf{p}, \mathbf{c})$ corresponds to a regularization term that penalizes complex shape and texture deformations, $\mathcal{D}(I, \mathbf{p}, \mathbf{c})$ denotes the global misalignment measure and corresponds to the data term in AAMs fitting and $\sum_{i=1}^{n} \mathcal{D}_i(\mathbf{x}_i, \mathcal{I}, \mathbf{p})$ denote the v local measures of misalignment which correspond to the data term in CLMs fitting. Substituting the previous terms for their expressions in 10 and 19, our probabilistic formulation can be rewritten as the following optimization problem:

$$\begin{aligned}
\mathbf{p}^*, \mathbf{c}^* = \arg\min_{\mathbf{p},\mathbf{c}} & \underbrace{||\mathbf{p}||^2_{\mathbf{\Lambda}^{-1}} + ||\mathbf{c}||^2_{\mathbf{\Sigma}^{-1}}}_{\mathcal{R}(\mathbf{p},\mathbf{c})} + \\
& \underbrace{\frac{1}{\sigma^2} ||\mathbf{i}[\mathbf{p}] - \bar{\mathbf{a}} + \mathbf{A}\mathbf{c}||^2}_{\mathcal{D}(I,\mathbf{p},\mathbf{c})} + \\
& \underbrace{\sum_{\mathbf{x}_i \in \mathbf{s}} \sum_{\mathbf{y}_j \in \mathbf{\Psi}_{\mathbf{x}_i}} \frac{w_j}{\rho^2} ||\mathbf{x}_i - \mathbf{y}_j||^2}_{\sum_{i=1}^{v} \mathcal{D}_i(\mathbf{x}_i, I, \mathbf{p})}
\end{aligned} \tag{23}$$

Note that our unified formulation naturally accommodates for uncertainty (noise) with respect to both shape and texture by explicitly incorporating both σ^2 and ρ^2 in the cost function. In fact, the ratio $\frac{\sigma^2}{\rho^2}$ determines the relative contribution of each term to the final unified cost function[6].

Equation 23 can be optimized by combining the RLMS algorithm for CLMs fitting described in Section 4 with the gradient descent algorithms for fitting AAMs described in Section 3. Combining PIC and RLMS algorithms, the optimal solution for the incremental shape parameters $\delta \mathbf{p}^*$ is given by:

$$\delta \mathbf{p}^* = -\mathbf{H}^{-1} \mathbf{b} \tag{24}$$

where:

$$\begin{aligned}
\mathbf{H} &= \mathbf{\Lambda}^{-1} + \frac{1}{\sigma^2} \mathbf{J}_{\mathbf{a}}^T \mathbf{P} \mathbf{J}_{\mathbf{a}} + \frac{1}{\rho^2} \mathbf{J}_{\mathbf{s}}^T \mathbf{J}_{\mathbf{s}} \\
\mathbf{b} &= \mathbf{\Lambda}^{-1} \mathbf{p} - \frac{1}{\sigma^2} \mathbf{J}_{\mathbf{a}}^T \mathbf{P} (\mathbf{i}[\mathbf{p}] - \bar{\mathbf{a}}) - \frac{1}{\rho^2} \mathbf{J}_{\mathbf{s}}^T (\boldsymbol{\mu} - \mathbf{s})
\end{aligned} \tag{25}$$

[6] Although theoretically optimal values for σ^2 and ρ^2 can be used, in practice, a better perfomant algorithm can be obtained by computing the value of the ratio $\frac{\sigma^2}{\rho^2}$ experimentally by using cross-validation on a small validation set.

and $\mathbf{J_a}$ and $\mathbf{J_s}$ are the Jacobians defined in 11 and 20 respectively.

Alternatively, by combining AIC with RLMS the optimal value for $\delta\mathbf{p}^*$ is again given by 24 where, in this case:

$$\begin{aligned}\mathbf{H} &= \mathbf{\Lambda}^{-1} + \frac{1}{\sigma^2}\mathbf{J_a}^T\mathbf{J_a} + \frac{1}{\rho^2}\mathbf{J_s}^T\mathbf{J_s} \\ \mathbf{b} &= \mathbf{\Lambda}^{-1}\mathbf{p} - \frac{1}{\sigma^2}\mathbf{J_a}^T(\mathbf{i}[\mathbf{p}] - \mathbf{a_a}) - \frac{1}{\rho^2}\mathbf{J_s}^T(\boldsymbol{\mu} - \mathbf{s})\end{aligned} \tag{26}$$

and $\mathbf{J_a}$ is defined in 12. Note that the solution for the optimal texture parameters $\mathbf{c}^*$ is again given by 12 and that both algorithm still utilize the exact same update rules defined in 11, 12 and 20.

6. Implementation Details

Code. We developed our own open-source implementations[7] of the previously described deformable model fitting algorithms *i.e.* PIC and AIC for AAMs, RLMS for CLMs and PIC+RLMS and AIC+RLMS for the proposed Unified model.

Training data. Reported results for the previous algorithms are obtained by training our implementations on the same 813 images of the Labelled Face Parts in the Wild (LFPW) [8] training dataset.

Shape and texture models. All our methods are implemented using a 2 level multi-resolution pyramidal scheme (face images are normalized to have a face size of roughly 100 pixels on the top resolution level). Similar to [33] we use a reduced version of the Dense Scale Invariant Feature Transform (DSIFT) [18] to define the image representation of both holistic and parts-based texture models. For holistic generative appearance models the number of texture components is set to 50 and kept constant throughout the optimization. Multi-Channel Correlation Filters [16] are used to learn parts-based discriminative appearance models. The size of each local patches is set to 17x17 and kept constant throughout the optimization. Finally, the dimensionality of the 2d shape model is set to 7 (4 similarity parameters + 3 nonrigid shape components) at the low resolution level and to 16 (4 + 12) at the top resolution one.

Run time. The average run time for each method (20 iteration per image) using our unoptimized single threaded Python implementations on a laptop equipped with a 2.3GHz quad-core Intel Core i7 processor are: PIC $\sim$ 80 ms, AIC $\sim$ 130 ms, RLMS $\sim$ 100 ms, PIC-RLMS $\sim$ 110 ms, and AIC-RLMS $\sim$ 140 ms.

[7] All implementations will be made publicly available as part of the Menpo Project [1] http://www.menpo.org/.

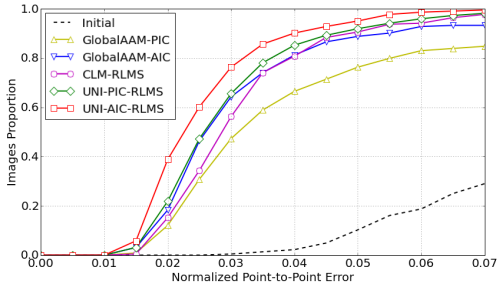

Figure 1: Cumulative Error Distributions over 66 landmarks on the LFPW dataset

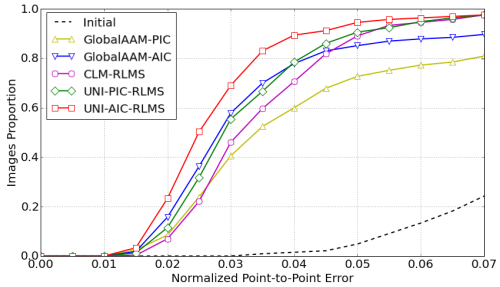

Figure 2: Cumulative Error Distributions over 66 landmarks on the Helen dataset.

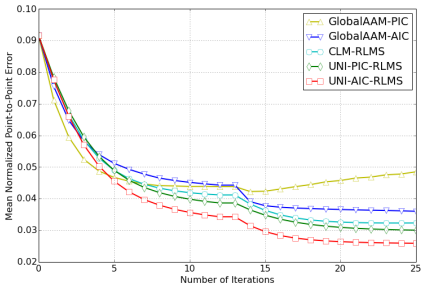

Figure 3: Mean normalized point to point error as a function of the number of iterations on the LFPW dataset

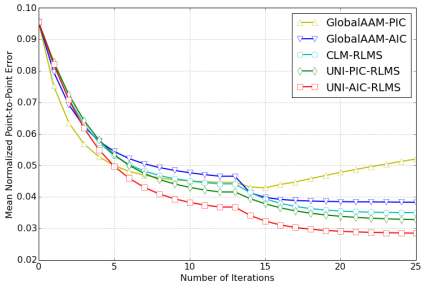

Figure 4: Mean normalized point to point error as a function of the number of iterations on the Helen dataset

7. Experiments

This section reports the performance of the proposed method on the problem of face alignment in-the-wild. Results for two different experiments are reported. The first experiment compares the accuracy and convergence properties of the proposed Unified PIC-RLMS and AIC-RLMS algorithms with respect to those of PIC [6], AIC [23] and and RLMS [28] on the popular LFPW [8] and Helen [17] datasets. The second experiment compares the performance of the previous algorithms against recently proposed state-of-the-art methods for face alignment in-the-wild, *i.e.* the Supervised Descent Method of Xiong and De La Torre [35] and the Gauss-Newton Deformable Parts Model of Tzimiropoulos and Pantic [33], on the very challenging Annotated Faces in the Wild (AFW) dataset.

7.1. Comparison with AAMs and CLMs

Results for this experiment are reported over the 224 and 330 test images of the LFPW [8] and Helen[17] datasets. 66 points ground truth landmark annotations were provided by the iBUG group[8]. All methods were initialized by perturbing the ground truth scale and translation parameters with Gaussian noise (rotations were not considered) and applying the resulting transformation to mean of the shape model. (Notice that this procedure produces initializations that are considerably more challenging than those reported in the recent AAM literature [32, 33]). The Cumulative Error Distributions (CED) for this experiment is shown in Figures 1 and 2. Figures 3 and 4 shows the evolution of the mean normalized point-to-point error as a function of the number of iterations run by each algorithm. This experiment shows that our Unified AIC-RLMS approach considerably outperforms all other methods by a large margin on both datasets. More specifically, AIC-RLMS achieves a constant improvement of between 10% to 20% over PIC-RLMS and AIC at the significant region $0.020 < err > 0.040$ (at which the results are generally considered adequate by visual inspection). Note that the fast Unified PIC-RLMS algorithm is also the second most performant algorithm, surpassing both AIC and RLMS, on this particular experiment.

7.2. Comparison with state of the art

Results for this experiment are reported over the 337 images of the AFW [37] dataset. In this case, 49 points ground truth landmark annotations for this dataset were again provided by the iBUG group[8]. Results for [35] and [33] were directly obtained using the publicly available models and fitting code kindly provided by the authors[9,10]. Note that, the provided models have been potentially trained using thousands of images in contrast to the only 813 images used to trained our method. In this experiment, all algorithms were initialized using the bounding box provided by our own in-house implementation of the face detector of [37]. The CED for this experiment are reported in Figure 5. The results show that our Unified AIC-RLMS algorithm achieves state-of-the-art results on the AFW dataset, considerably outperforming both the Gauss-Newton Deformable Parts-Model of Tzimiropoulos and Pantic [33] (which can be extremely accurate but sensitive to inaccurate initializations) and the SDM method of Xiong and De la Torre [35] (which can deal with very noisy initializations but is significantly less accurate than our method).

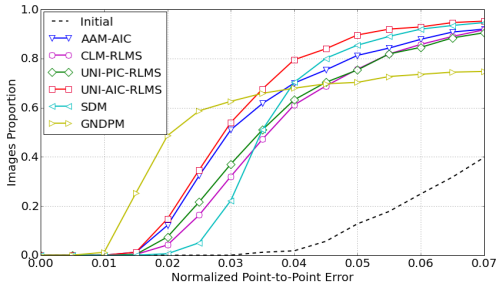

Figure 5: Cumulative Error Distributions over 49 landmarks for the AFW dataset.

8. Conclusion

In this paper we present a novel approach to deformable model fitting that unifies previous holistic and part-based deformable model fitting approaches. We show that our approach naturally arises by defining a novel probabilistic formulation of the fitting process. We use such formulation to derive two novel deformable model fitting algorithms that unify the well-established frameworks of Active Appearance Models (AAMs) and Constrained Local Models (CLMs) fitting. Finally, we show that our unified approach, trained on a relatively small amount data, can compete and even surpass the accuracy of two of the most recently proposed state-of-the-art techniques, trained using thousands of training examples, on the challenging problem of face alignment in-the-wild.

Acknowledgements The work of Joan Alabort-i-Medina is funded by a DTA studentship from Imperial College London and by the Qualcomm Innovation Fellowship. The work of Stefanos Zafeiriou is partially supported by the EPSRC project EP/L026813/1 Adaptive Facial Deformable Models for Tracking (ADAManT)

[8] http://ibug.doc.ic.ac.uk/resources/300-W/
[9] http://www.humansensing.cs.cmu.edu/intraface/
[10] http://ibug.doc.ic.ac.uk/resources/gauss-newton-deformable-part-models-face-alignment/

References

[1] J. Alabort-i-Medina, E. Antonakos, J. Booth, P. Snape, and S. Zafeiriou. Menpo: A comprehensive platform for parametric image alignment and visual deformable models. In *ACM International Conference on Multimedia (ACMM)*, 2014. 1, 7

[2] J. Alabort-i-Medina and S. Zafeiriou. Bayesian active appearance models. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 1, 4

[3] E. Antonakos, J. Alabort-i-Medina, G. Tzimiropoulos, and S. Zafeiriou. HOG active appearance models. In *International Conference on Image Processing (ICIP)*, 2014. 1, 2

[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. 1, 2, 5

[5] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic. Incremental face alignment in the wild. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 3

[6] S. Baker and I. Matthews. Lucas-kanade 20 years on: A unifying framework. *International Journal of Computer Vision (IJCV)*, 2004. 4, 8

[7] T. Baltruaitis, P. Robinson, and L.-P. Morency. Continuous conditional neural fields for structured regression. In *European Conference on Computer Vision (ECCV)*, 2014. 1, 2

[8] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar. Localizing parts of faces using a consensus of exemplars. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011. 1, 2, 7, 8

[9] V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. In *SIGGRAPH*, 1999. 2

[10] X. P. Burgos-Artizzu, P. Perona, and P. Dollr. Robust face landmark estimation under occlusion. In *International Conference on Computer Vision (ICCV)*, 2013. 3

[11] X. Cao, Y. Wei, F. Wen, and J. Sun. Face alignment by explicit shape regression. In *Computer Vision and Pattern Recognition (CVPR)*, 2012. 1, 3

[12] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2001. 1, 2, 3

[13] T. F. Cootes and C. J. Taylor. On representing edge structure for model matching. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2001. 2

[14] 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. 1, 2, 5

[15] V. Kazemi and J. Sullivan. One millisecond face alignment with an ensemble of regression trees. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 3

[16] H. Kiani Galoogahi, T. Sim, and S. Lucey. Multi-channel correlation filters. In *The IEEE International Conference on Computer Vision (ICCV)*, 2013. 7

[17] V. Le, B. Jonathan, Z. Lin, L. Boudev, and T. S. Huang. Interactive facial feature localization. In *European Conference on Computer Vision (ECCV)*, 2012. 2, 8

[18] D. G. Lowe. Object recognition from local scale-invariant features. In *International Conference on Computer Vision (ICCV)*, 1999. 7

[19] S. Lucey, R. Navarathna, A. B. Ashraf, and S. Sridharan. Fourier lucas-kanade algorithm. *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2013. 2

[20] P. Martins, R. Caseiro, and J. Batista. Non-parametric bayesian constrained local models. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 1, 2, 5

[21] I. Matthews and S. Baker. Active appearance models revisited. *International Journal of Computer Vision (IJCV)*, 2004. 1, 2, 3, 4, 5

[22] B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 1997. 4, 5

[23] G. Papandreou and P. Maragos. Adaptive and constrained algorithms for inverse compositional active appearance model fitting. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008. 4, 5, 8

[24] S. Ren, X. Cao, Y. Wei, and J. Sun. Face alignment at 3000 fps via regressing local binary features. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 3

[25] S. Rivera and A. Martinez. Learning deformable shape manifolds. *Pattern Recognition (PR)*, 2012. 3

[26] J. Saragih. Principal regression analysis. In *Computer Vision and Pattern Recognition (CVPR)*, 2011. 2

[27] J. Saragih and R. Goecke. A nonlinear discriminative approach to aam fitting. In *International Conference on Computer Vision (ICCV)*, 2007. 2

[28] J. M. Saragih, S. Lucey, and J. F. Cohn. Deformable model fitting by regularized landmark mean-shift. *International Journal of Computer Vision (IJCV)*, 2011. 1, 2, 5, 6, 8

[29] Y. Sun, X. Wang, and X. Tang. Deep convolutional network cascade for facial point detection. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013. 3

[30] P. Tresadern, P. Sauer, and T. Cootes. Additive update predictors in active appearance models. In *British Machine Vision Conference (BMVC)*, 2010. 2

[31] G. Tzimiropoulos, J. Alabort-i-Medina, S. Zafeiriou, and M. Pantic. Generic active appearance models revisited. In *Asian Conference on Computer Vision (ACCV)*, 2012. 2

[32] G. Tzimiropoulos and M. Pantic. Optimization problems for fast aam fitting in-the-wild. In *International Conference on Computer Vision (ICCV)*, 2013. 2, 8

[33] G. Tzimiropoulos and M. Pantic. Gauss-newton deformable part models for face alignment in-the-wild. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 1, 2, 5, 7, 8

[34] C. Vogler, Z. Li, A. Kanaujia, S. Goldenstein, and D. Metaxas. The best of both worlds: Combining 3d deformable models with active shape models. In *International Conference on Computer Vision (ICCV)*, 2007. 3

[35] Xuehan-Xiong and F. De la Torre. Supervised descent method and its application to face alignment. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013. 1, 2, 3, 8

[36] J. Yan, Z. Lei, D. Yi, and S. Li. Learn to combine multiple hypotheses for accurate face alignment. In *International Conference on Computer Vision Workshops (ICCV-W)*, 2013. 3

[37] X. Zhu and D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012. 1, 2, 8