Active Orientation Models

Georgios Tzimiropoulos, Joan Alabort-i-Medina, *Student Member, IEEE*, Stefanos Zafeiriou *Member, IEEE*, and Maja Pantic *Fellow, IEEE*

Abstract—We present Active Orientation Models, generative models of facial shape and appearance, which extend the well-known paradigm of Active Appearance Models (AAMs) for the case of generic face alignment under unconstrained conditions. Robustness stems from the fact that the proposed AOMs employ a statistically robust appearance model based on the principal components of image gradient orientations. We show that when incorporated within standard optimization frameworks for AAM learning and fitting, this kernel PCA results in robust algorithms for model fitting. At the same time, the resulting optimization problems maintain the same computational cost. As a result the main similarity of AOMs with AAMs is computational complexity. In particular, the project-out version of AOMs is as computationally efficient as the standard project-out inverse compositional algorithm which is admittedly one of the fastest algorithms for fitting AAMs. We verify experimentally that (i) AOMs generalize well to unseen variations and (ii) outperform all other state-of-the-art AAM methods considered by a large margin. This performance improvement brings AOMs at least in par with other contemporary methods for face alignment. Finally, we provide Matlab code at http://ibug.doc.ic.ac.uk/resources.

Index Terms—Active Orientation Models, Active Appearance Models, Face alignment.

I. INTRODUCTION

BECAUSE of their numerous applications in HCI, face analysis/recognition and medical imaging, the problems of learning and fitting deformable models have been the focus of cutting edge research in computer vision and machine learning for more than two decades. Put in simple terms, these problems can be summarized as follows: Learning a deformable model consists of (a) annotating (typically manually) a set of points (or landmarks) over a set of training images capturing an object of interest (e.g. faces), (b) learning a shape model (or point distribution model) which effectively represents the structure and variations among the annotated points and (c) learning appearance models from the image texture associated with the learned shape. Fitting a deformable model utilizes the learned shape and appearance models to detect the location of landmarks in new images; this can be done using regression, classification or could be formulated as a non-linear optimization problem.

Depending on the application and/or approach many terms have been used to coin this research: deformable model fitting, Active Shape Models (ASMs) [1], Constrained Local Models (CLMs) [2], [3] landmark localization, point detection, and Active Appearance Models (AAMs) [4], [5] to name a few. The latter approach and the problem of deformable face alignment are of particular interest to this work along with the seminal work [5] for fitting AAMs to face images. AAMs are generative models of shape and appearance typically learned by applying Principal Component Analysis to both shape and
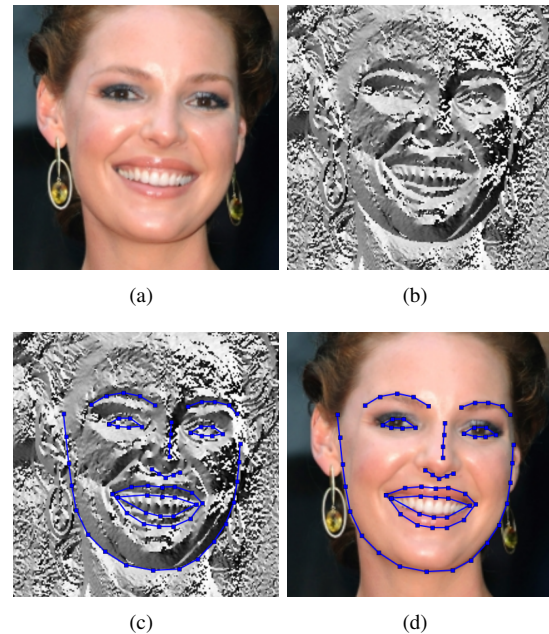

Fig. 1. (a) The aim of this work is to detect a set of facial features like the corners of the mouth and the tip of the nose in facial images captured in unconstrained conditions. This problem is known as face alignment in-the-wild. (b) To this end, we propose Active Orientation Models (AOMs), generative deformable models that utilize a statistically robust appearance model based on image gradient orientations. (c) The localization of the facial features is performed in the space of image gradient orientations via a robust algorithm for model fitting and parameter estimation. The obtained fitting result is shown in (d).

texture. In [5], fitting was formulated as a non-linear minimization problem which consists of minimizing the error between the model instance and the given image with respect to the model parameters which control the shape and appearance variation of faces. This problem was solved using the project-out inverse compositional algorithm, which decouples shape from appearance and results in a computationally efficient algorithm. Owing to its efficiency and accuracy, the algorithm for fitting AAMs proposed in [5] has become the de facto choice for building and fitting person-specific AAMs (i.e. AAMs trained to fit face images of a specific subject which is known in advance).

Despite their efficiency and accuracy, AAMs in general, and the project-out algorithm of [5] in particular, have been criticized for their inability to generalize well to unseen variations of illumination, expression and identity. Like AAMs, the proposed AOMs are generative models which however better address all the aforementioned limitations. Our motivation to pursue a generative AAM-based approach to generic face alignment is two fold. (a) For specific scenarios of interest [6], generative models have been shown to model sufficiently

accurately unseen variations, i.e. it is the fitting algorithm which fails to fit the model to unseen images. (b) New tools and insights on how to solve (a) have been recently suggested in [7], [8].

Main results. (a) **Models**: We propose Active Orientation Model, a generative deformable model that uses a statistically robust appearance model based on the principal components of image gradient orientations (see Fig. 1). We show how to use gradient decent in order to minimize the distance between this non-linear appearance model and a new image with respect to the model parameters. (b) **Complexity**: Similarly to the AAM formulation of [5], we show that AOMs can be optimized using the project-out inverse compositional algorithm which is admittedly one of the fastest algorithms for fitting deformable models in images. (c) **Robustness/Accuracy**: To the best of our knowledge, we demonstrate for the first time that this algorithm can be used to fit the learned models to faces not seen in the training set. (d) **Comparison to previous work**: We conducted a large number of experiments on many popular in-the-wild databases, in which all algorithms were trained, initialized and tested in the same way. We show that the proposed AOM largely outperforms all other state-of-the-art AAM methods considered as well as an enhanced version of a state-of-the-art CLM method [9], and our version of the SDM method of [10]. In the context of our experiments, we also propose a new method: we developed an AAM variant based on least-squares optimization and the appearance model of [11] which is based on similar features. One of the main contributions in our work is to show that the kernel PCA implied by the appearance model of AOMs also outperforms this AAM variant. (d) **Code**: Similarly to [12], [9], [13], we provide Matlab code at http://ibug.doc.ic.ac.uk/resources.

II. RELATED WORK

A. State-of-the-art.

Because of the inability of AAMs to generalize well to unseen variations, recent research has suggested the use of simpler (often local) and thus easier to optimize models and the application of discriminative methods for model fitting. The family of methods termed ASM-CLMs combine patch-based image representations, discriminatively trained point detectors and global shape constraints to localize landmarks in unseen images [2], [14], [3], [15]. A recent notable example of ASMs is the component-based model of [16] which produces good results on the difficult Helen data set introduced in the same work. Another recent approach that combines the output of SVM-based local detectors with a non-parametric set of global models has been shown to produce excellent results on unconstrained images in [17]. Finally, because sliding-window landmark detectors may be slow, regression-based techniques have been proposed to learn a mapping between local patches and landmarks [17], [12], [18], [19]. Not only do these methods enjoy a high degree of computational efficiency, but also they have been shown to achieve state-of-the-art performance for difficult experiments with unconstrained images. In particular, the results presented in the shape regression method of [18] and the supervised decent method of [10] are

impressive and considered state-of-the-art for the difficult data set of LPFW [17]. Finally, a radically different approach to face alignment is the globally optimized part-based model of [13]. Our experiments have shown that [13] is better able to detect faces rather than accurately locating facial landmarks [20].

B. AAMs.

Discriminative methods have been shown to improve the ability of AAMs to fit new faces in [21], [22], [23], [24]. One of the first attempts is the AAM of Cootes and Taylor [11] which is based on regression and features similar to the ones employed by AOMs. Please see [25] for a review on the effect of different texture representations on the performance of regression-based approaches to AAM fitting. Note that no matter the texture representation used, regression methods as proposed in [11] have been reported to produce limited fitting accuracy and robustness (please see [10] for a recent example). One of the main contributions of the proposed work is to illustrate that the kernel PCA implied by the appearance model of AOMs when combined with analytic gradient descent (as opposed to regression) and trained in-the-wild produce results comparable (if not superior) to state-of-the-art methods. Other generic AAMs learn a fitting function through maximizing the score of a two-class classifier (aligned or not aligned) or ranking [21], [22]. Boosting a huge number of Haar features is very inefficient, and results are reported only for low resolution images. This immediately rules out the possibility of accurate landmark localization in high resolution images and it is clearly unsatisfactory. Other discriminative approaches include learning non-linear regressors from features to model parameters through boosting and simulation [23], [24]. However, all these approaches seem to produce inferior results compared to the family of methods coined CLMs [2], [14], [3], which build upon the Active Shape Model [1]. Finally, a notable example of generative AAMs is the Fourier-Gabor AAMs of [26]. However, as it is mentioned by the authors, Fourier-Gabor AAMs appear to be more suitable for person-specific face alignment for the case of unseen illumination rather than generic face alignment.

III. ACTIVE APPEARANCE MODELS

A. Shape and appearance models

An AAM is defined by the shape, appearance and motion models. The shape model is typically learned by annotating N fiducial points on the object (e.g. a face) of training image $\mathbf{I}_i$. These points are said to define the shape of each object. Next, Procrustes analysis is applied to remove similarity transformations from the original shapes. Finally, PCA is applied on the similarity-free shapes $\mathbf{s}_i = [x_1, y_1, x_2, y_2, \ldots x_N, y_N]$. The resulting model $\{\mathbf{\Phi}_{S,0}, \mathbf{\Phi}_S \in \mathcal{R}^{2N \times p}\}$ can be used to represent a test shape $\mathbf{s}_y$ as

$$\hat{\mathbf{s}}_y = \mathbf{\Phi}_{S,0} + \mathbf{\Phi}_S \mathbf{p}, \qquad \mathbf{p} = \mathbf{\Phi}_S^T (\mathbf{s}_y - \mathbf{\Phi}_{S,0}). \qquad (1)$$

The eigenvectors of $\mathbf{\Phi}_S$ represent pose, expression and identity variation. The appearance model of an AAM is learned by first warping each of the training images $\mathbf{I}_i(\mathbf{x})$ to the canonical

reference frame defined by the mean shape using the motion model $\mathbf{W}(\mathbf{x}; \mathbf{p})$ and then applying PCA on the shape-free textures. We choose piecewise affine warps as the motion model in this work. The resulting model $\{\boldsymbol{\Phi}_{A,0}, \boldsymbol{\Phi}_A \in \mathcal{R}^{K \times q}\}$ can be used to represent a shape-free test texture $\mathbf{a}_y$ as

$$\hat{\mathbf{a}}_y = \boldsymbol{\Phi}_{A,0} + \boldsymbol{\Phi}_A \mathbf{c}, \quad \mathbf{c} = \boldsymbol{\Phi}_A^T(\mathbf{a}_y - \boldsymbol{\Phi}_{A,0}). \quad (2)$$

Finally, a model instance is synthesized to represent test object y by warping $\mathbf{a}_y$ from the mean shape $\bar{\mathbf{s}}$ to $\hat{\mathbf{s}}_y$ using piecewise affine warp.

B. Model fitting

Given a new image $\mathbf{I}$, inference in AAMs entails estimating $\mathbf{p}$ and $\mathbf{c}$ assuming "reasonable" initialization of the fitting process. This initialization is typically performed by placing the mean shape according to the output of an object (in this work, face) detector. Note that only $\mathbf{p}$ needs to be estimated for deformable model fitting. Estimating $\mathbf{c}$ is a by-product of the fitting algorithm. Various algorithms and cost functions have been proposed to estimate $\mathbf{p}$ and $\mathbf{c}$ including regression, classification and non-linear optimization methods. The latter approach is of particular interest in this work. It minimizes the ℓ_2-norm of the error between the model instance and the given image with respect to the model parameters as follows

$$\{\mathbf{p}_o, \mathbf{c}_o\} = \arg \min_{\{\mathbf{p},\mathbf{c}\}} ||\mathbf{I}[\mathbf{p}] - \boldsymbol{\Phi}_{A,0} - \boldsymbol{\Phi}_A \mathbf{c}||^2, \quad (3)$$

where for notational convenience we write $\mathbf{I}[\mathbf{p}](k)$ to denote the pixel intensity $\mathbf{I}(\mathbf{W}(\mathbf{x}_k; \mathbf{p}))$, and $\mathbf{I}[\mathbf{p}]$ to denote image $\mathbf{I}(\mathbf{W}(\mathbf{x}; \mathbf{p}))$ re-arranged as a $K \times 1$ vector. In a series of seminal papers [27], [5], Baker and Matthews illustrated that problem (3) can be solved using an optimization framework based on the Lukas-Kanade (LK) algorithm [28]. LK algorithm is an image alignment method. Suppose for the time being that an oracle provides $\mathbf{c}_o$ so that $\mathbf{c}$ is fixed. Then, (3) becomes an image alignment problem. In this case, and according to [28], the resulting optimization problem can be solved using Gauss-Newton: First, $\mathbf{I}[\mathbf{p}]$ is linearized with respect to $\mathbf{p}$ and then a solution for a $\Delta \mathbf{p}$ is computed from the set of the derived normal equations obtained by setting the derivative of (3) with respect to $\Delta \mathbf{p}$ to $\mathbf{0}$. Finally, $\mathbf{p}$ is updated in an additive fashion $\mathbf{p} \leftarrow \mathbf{p} + \Delta \mathbf{p}$. We refer the interested reader to [27] for an excellent coverage of the LK algorithm.

Inverse Compositional. As illustrated in [27], [5], problem (3) can be solved in two coordinate frames. The *forward* case is the standard LK algorithm, as summarized above. In general, forward algorithms are slow because the Jacobian and its inverse must be re-evaluated at each iteration. Fortunately, computationally efficient algorithms can be derived by solving (3) using the *inverse* compositional framework. Let us drop $_A$ from $\boldsymbol{\Phi}_A$ for notational convenience and let $\boldsymbol{\Phi}_i$ represent the i-th column (eigenvector) of $\boldsymbol{\Phi}$. In inverse algorithms, each eigenvector $\boldsymbol{\Phi}_i$ is linearized around $\mathbf{p} = \mathbf{0}$. By additionally linearizing with respect to $\mathbf{c}$ (3) becomes

$$\arg \min_{\Delta \mathbf{p}, \Delta \mathbf{c}} ||\mathbf{I}[\mathbf{p}] - \boldsymbol{\Phi}_0 + \mathbf{J}_0 \Delta \mathbf{p} - \sum_{i=1}^{m} (c_i + \Delta c_i)(\boldsymbol{\Phi}_i + \mathbf{J}_i \Delta \mathbf{p})||^2,$$
$$(4)$$

where $\mathbf{J}_i$ is the $K \times p$ matrix each row of which contains the $1 \times p$ vector $[\boldsymbol{\Phi}_{i,x}[\mathbf{p}](k) \quad \boldsymbol{\Phi}_{i,y}[\mathbf{p}](k)]\frac{\partial \mathbf{W}(\mathbf{x}_k; \mathbf{p})}{\partial \mathbf{p}}$. $\boldsymbol{\Phi}_{i,x}[\mathbf{p}](k)$ and $\boldsymbol{\Phi}_{i,y}[\mathbf{p}](k)$ are the x and y gradients of $\boldsymbol{\Phi}_i$ for the k-th pixel and $\frac{\partial \mathbf{W}(\mathbf{x}_k; \mathbf{p})}{\partial \mathbf{p}} \in \mathcal{R}^{2 \times p}$ is the Jacobian of the piecewise affine warp. Please see [5] for calculating and implementing $\frac{\partial \mathbf{W}}{\partial \mathbf{p}}$. All these terms are defined in the model coordinate frame for $\mathbf{p} = \mathbf{0}$ and can be pre-computed. An update for $\Delta \mathbf{c}$ and $\Delta \mathbf{p}$ can be obtained in closed form only after second order terms are omitted as follows

$$\arg \min_{\Delta \mathbf{p}, \Delta \mathbf{c}} ||\mathbf{I}[\mathbf{p}] - \boldsymbol{\Phi}_0 - \boldsymbol{\Phi} \mathbf{c} - \boldsymbol{\Phi} \Delta \mathbf{c} - \mathbf{J} \Delta \mathbf{p}||^2, \quad (5)$$

where $\mathbf{J} = \mathbf{J}_0 + \sum_{i=1}^{q} c_i \mathbf{J}_i$. In [27], the update was derived as

$$[\Delta \mathbf{p}; \Delta \mathbf{c}] = \mathbf{H}_s^{-1} \mathbf{J}_s^T(\mathbf{I}[\mathbf{p}] - \boldsymbol{\Phi}_0 - \boldsymbol{\Phi} \mathbf{c}), \quad (6)$$

where $\mathbf{J}_s = [\boldsymbol{\Phi}; \mathbf{J}] \in \mathcal{R}^{K \times (p+q)}$ and $\mathbf{H}_s = \mathbf{J}_s^T \mathbf{J}_s$. Once $\Delta \mathbf{p}$ is computed, $\mathbf{p}$ is updated in a compositional fashion $\mathbf{p} \leftarrow \mathbf{p} \circ \Delta \mathbf{p}^{-1}$, where $\circ$ denotes the composition of two warps. (Please see [5] for a principled way of applying the inverse composition to AAMs). This is the well-known simultaneous algorithm.

The simultaneous algorithm is slow because the Jacobian $\mathbf{J}_s$, the Hessian $\mathbf{H}_s$ and its inverse must be re-computed at each iteration. One can easily show that the cost for the Hessian computation is $O((p + q)^2 K)$. Nonetheless, more efficient ways to optimize (5) exist. Let us define the projection operator $\mathbf{P} = \mathbf{E} - \boldsymbol{\Phi} \boldsymbol{\Phi}^T$, where $\mathbf{E}$ is the identity matrix. Then, a number of works [29], [30], [31] have shown that one can update the appearance and shape parameters in an alternating fashion from

$$\Delta \mathbf{c} = \boldsymbol{\Phi}^T(\mathbf{I}[\mathbf{p}] - \boldsymbol{\Phi}_0 - \boldsymbol{\Phi} \mathbf{c} - \mathbf{J} \Delta \mathbf{p}) \quad (7)$$

$$\Delta \mathbf{p} = \mathbf{H}_a^{-1} \mathbf{J}_a^T(\mathbf{I}[\mathbf{p}] - \boldsymbol{\Phi}_0), \quad (8)$$

where the projected-out Jacobian and Hessian are given by $\mathbf{J}_a = \mathbf{P} \mathbf{J}$ and $\mathbf{H}_a = \mathbf{J}_a^T \mathbf{J}_a$, respectively. As shown in [30], the above update rules result in an algorithm with complexity per iteration $O(pqK + p^2K + p^3)$ which can be readily handled by current systems.

By far the most efficient algorithm for fitting AAMs is the so-called project-out inverse compositional algorithm, which in essence is a LK algorithm. This algorithm decouples shape and appearance by solving (5) in the subspace orthogonal to $\boldsymbol{\Phi}$. Observe that $||\mathbf{I}[\mathbf{p}] - \boldsymbol{\Phi}_0 - \boldsymbol{\Phi} \mathbf{c}||_{\mathbf{P}}^2 = ||\mathbf{I}[\mathbf{p}] - \boldsymbol{\Phi}_0||_{\mathbf{P}}^2$ [1]. Hence an update for $\Delta \mathbf{p}$ can be computed by optimizing

$$\arg \min_{\Delta \mathbf{p}} ||\mathbf{I}[\mathbf{p}] - \boldsymbol{\Phi}_0 - \mathbf{J}_0 \Delta \mathbf{p}||_{\mathbf{P}}^2. \quad (9)$$

The solution to the above problem is given by

$$\Delta \mathbf{p} = \mathbf{H}_p^{-1} \mathbf{J}_p^T(\mathbf{I}[\mathbf{p}] - \boldsymbol{\Phi}_0), \quad (10)$$

where the projected-out Jacobian $\mathbf{J}_p = \mathbf{P} \mathbf{J}_0$ and Hessian $\mathbf{H}_p = \mathbf{J}_p^T \mathbf{J}_p$, can be both pre-computed. This reduces the cost per iteration to $O(pK)$, only [5], which is the cost of the inverse compositional LK algorithm [27]. This algorithm has been shown to track faces at 300 fps [6].

---

[1]For a vector $\mathbf{x}$, we use the notation $||\mathbf{x}||_{\mathbf{P}}^2$ to denote the weighted norm $\mathbf{x}^T \mathbf{P} \mathbf{x}$.

IV. ACTIVE ORIENTATION MODELS

The deformable model fitting framework of the previous section and especially the project-out inverse compositional algorithm has been highly criticized as difficult to optimize mainly due to the high-dimensional parameter space and the existence of numerous undesirable local minima in the derived cost functions. Therefore, the problem in hand is how to avoid these local minima during optimization. We propose to address this problem by using a kernel PCA based on a similarity criterion robust to outliers. In particular, AOMs employ a statistically robust appearance model based on the principal components of images gradient orientations. As we show below both shape and texture models can be optimized in a standard least-squares framework which results in computationally efficient algorithms.

A. Appearance Model

At the heart of the appearance model of AOMs there exists a robust kernel for measuring similarity. We define outliers to be anything that the learned appearance model cannot reconstruct because (a) it was not seen in the training set (e.g. appearance variation due to different identity, expression or illumination) (b) it does not belong to the face space at all (e.g. glasses) and (c) it was excluded from $\mathbf{\Phi}_A$ as noise because in any case the number of principal components in $\mathbf{\Phi}_A$ should be kept as small as possible so that the model is easier to optimize and cannot generate appearance which is unrelated to faces. Note that as it was shown in [6] for some cases of interest (e.g. appearance variation in frontal views), a very compact appearance space, learned from a training set with a few persons only, in general, results in relatively small reconstruction errors of unseen faces. This illustrates that a generative model is not an unreasonable choice for generic deformable model fitting. All that is needed is a robust cost function to fit this model.

A general framework for robust estimation is weighted least squares [32]. Let us define $\mathbf{e} = \mathbf{I}[\mathbf{p}] - \bar{\mathbf{a}} - \mathbf{\Phi}_A \mathbf{c}$. Then, weighted least squares methods optimize

$$\{\mathbf{p}_o, \mathbf{c}_o\} = \arg\min_{\{\mathbf{p}, \mathbf{c}\}} \mathbf{e}^T \mathbf{Q} \mathbf{e}, \tag{11}$$

where $\mathbf{Q} \in R^{\{K, K\}}$ is a diagonal weighting matrix which down-weighs pixels corrupted by outliers. An ideal case would be $\mathbf{Q}_k = 0$ if pixel k is an outlier and $\mathbf{Q}_k = 1$ otherwise. The estimation of $\mathbf{Q}$ along with the optimal model parameters have been extensively studied in the literature of robust statistics (please see [32] for a review). However, none of these out-of-the-box approaches has been proven successful so far in AAM fitting because (a) the noise model for outliers in our case is very hard to define and (b) the estimation process is also very prone to local minima.

We propose to address this problem in AAMs by using a robust similarity criterion based on image gradient orientations [7], [8]. Suppose that we wish to measure the similarity between two images $\mathbf{I}_i$, $i = 1, 2$. For each image, we extract image gradients $\mathbf{g}_{i,x}, \mathbf{g}_{i,y}$ and the corresponding estimates of gradient orientation ϕ_i. Let us denote by $\mathbf{z}_i$ the the so-called

normalized gradients

$$\mathbf{z}_i = \frac{1}{\sqrt{K}} [\cos(\phi_i)^T, \sin(\phi_i)^T]^T, \tag{12}$$

where $\cos(\phi_i) = [\cos(\phi_i(1)), \dots, \cos(\phi_i(K))]^T$ and $\sin(\phi_i)$ is similarly defined. Then, the following kernel can be used to measure image similarity

$$\begin{aligned} s &= \mathbf{z}_1^T \mathbf{z}_2 \\ &= \frac{1}{K} \sum_{k \in \Omega} \cos(\phi_1(k) - \phi_2(k)), \end{aligned} \tag{13}$$

where Ω denotes the image support.

Let us also denote by Ω_1 the image support that is outlier-free and Ω_2 the image support that is corrupted by outliers ($\Omega = \Omega_1 \cup \Omega_2$). Then, as it was shown in [8], under some assumptions, it holds

$$\sum_{k \in \Omega_2} \cos(\phi_1(k) - \phi_2(k)) \approx 0. \tag{14}$$

Note that (a) in contrary to [27], no assumption about the structure of outliers is made and (b) no actual knowledge of Ω is required. Based on (14), we can re-write (13) as follows

$$\begin{aligned} s &= \sum_{k \in \Omega_1} \cos(\phi_1(k) - \phi_2(k)) + \sum_{k \in \Omega_2} \cos(\phi_1(k) - \phi_2(k)) \\ &= \sum_{k \in \Omega_1} 1 \cdot \cos(\phi_1(k) - \phi_2(k)) + \sum_{k \in \Omega_2} \epsilon \cdot \cos(\phi_1(k) - \phi_2(k)) \\ &\approx \mathbf{z}_1^T \mathbf{Q}_{\text{ideal}} \mathbf{z}_2, \end{aligned} \tag{15}$$

where $\epsilon \to 0$ and $\mathbf{Q}_{\text{ideal}}$ is the "ideal" weighting matrix defined above. Note that $\mathbf{Q}_{\text{ideal}}$ in (15) is never calculated explicitly. We can write (15) only because outliers are approximately "canceled out" when the above kernel is used to measure image similarity.

The robust kernel of (13) can be used to define a kernel PCA [8]. The appearance model in AOMs is learned using this robust PCA. Note that the kernel can be written using the explicit mapping of (12) and therefore no pre-image computation is required. Suppose that for each training image $\mathbf{I}_i$, $i = 1, \dots h$ we extract the shape-free normalized gradients of each training face $\mathbf{z}_i$, and then we form the data matrix $\mathbf{Z} \in \mathcal{R}^{2K \times h}$ the columns of which contain $\mathbf{z}_i$. Then we apply PCA on $\mathbf{Z}$. We denote by

$$\mathbf{\Phi}_Z = [\mathbf{\Phi}_x; \mathbf{\Phi}_y] \in \Re^{2K \times q} \tag{16}$$

the learned appearance model, where $\mathbf{\Phi}_x$ and $\mathbf{\Phi}_y \in \Re^{K \times q}$ are the parts corresponding to the cosine and sine terms, respectively. Note that to preserve the kernel properties no subtraction of the "mean" normalized gradient is needed and the first eigenvector is treated as the mean where it is required.

B. Shape Model

The shape model of AOMs is the same as the one used in AAMs. This implies that all calculations required for the Jacobian of the piecewise affine warp and the update of the parameters in the inverse compositional fashion remain the same.

C. Inference

We perform inference in AOMs by minimizing the error between a test image and a model instance

$$\{\mathbf{p}_o, \mathbf{c}_o\} = \arg\max_{\{\mathbf{p},\mathbf{c}\}} ||\mathbf{z}[\mathbf{p}] - \mathbf{\Phi}_Z \mathbf{c}||^2, \qquad (17)$$

where $\mathbf{z}[\mathbf{p}](k)$ denotes the normalized gradient of $\mathbf{I}(\mathbf{W}(\mathbf{x}_k; \mathbf{p}))$. Our optimization strategies are summarized below:

Simultaneous. The simultaneous AOM minimizes (17) with respect to both $\{\mathbf{p}, \mathbf{c}\}$. This requires the computation of the Hessian ($O((p+q)^2 2K)$) and its inverse ($O((p+q)^3)$) and is very slow. For this reason, we did not look into this algorithm further.

Alternating. The minimization of (17) can be also performed in an alternating fashion. Because the "mean" normalized gradient is the first eigenvector of our kernel PCA, we write

$$\mathbf{J} = \sum_{i=0}^{q} c_i \mathbf{J}_i \mathbf{J}_i = [\mathbf{J}_{i,x}; \mathbf{J}_{i,y}], \qquad (18)$$

where $\mathbf{J}_{i,x}$ and $\mathbf{J}_{i,y}$ are the Jacobians computed from $\mathbf{\Phi}_{i,x}$ and $\mathbf{\Phi}_{i,y}$ respectively. As before let us drop the dependency of notation from $_Z$ and define the projection operator $\mathbf{P} = \mathbf{E} - \mathbf{\Phi}\mathbf{\Phi}^T$. Let us also define the projected-out Jacobian and Hessian

$$\mathbf{J}_a = \mathbf{PJH}_a = \mathbf{J}_a^T \mathbf{J}_a. \qquad (19)$$

Then, the updates for the appearance and shape parameters are given by

$$\Delta\mathbf{c} = \mathbf{\Phi}^T(\mathbf{z}[\mathbf{p}] - \mathbf{\Phi}\mathbf{c} - \mathbf{J}\Delta\mathbf{p}) \qquad (20)$$

$$\Delta\mathbf{p} = \mathbf{H}_a^{-1}\mathbf{J}_a^T\mathbf{z}[\mathbf{p}]. \qquad (21)$$

For completeness, we derive the update rules for $\Delta\mathbf{c}$ and $\Delta\mathbf{p}$ in the Appendix. Notice that in the above update for $\Delta\mathbf{p}$, the "error image" is implicitly calculated because

$$\mathbf{J}_a^T\mathbf{z}[\mathbf{p}] = \mathbf{J}^T\mathbf{P}^T\mathbf{z}[\mathbf{p}] = \mathbf{J}^T(\mathbf{z}[\mathbf{p}] - \mathbf{\Phi}\mathbf{c}), \qquad (22)$$

which shows that the error is simply the difference between the normalized gradient and its reconstruction from the appearance subspace. The steps of the algorithm are summarized in Algorithm 1.

Algorithm 1 AOM - Alternating Optimization

Given: Models $\{\mathbf{\Phi}_{S,0}, \mathbf{\Phi}_S\}$ and $\mathbf{\Phi}_Z$, and current estimates for $\mathbf{c}$ and $\mathbf{p}$

Pre-compute: Jacobian $\mathbf{J}_i$ for each appearance eigenvector

1: **repeat**
2:　　Compute warped image $\mathbf{I}[\mathbf{p}]$ and extract normalized gradients $\mathbf{z}[\mathbf{p}]$
3:　　Calculate $\mathbf{J}$ and $\mathbf{J}_a$ from (18) and (19)
4:　　Calculate $\mathbf{H}_a$ and $\mathbf{H}_a^{-1}$ from (19)
5:　　Calculate $\Delta\mathbf{c}$ and $\Delta\mathbf{p}$ from (20) and (21)
6:　　Update from $\mathbf{c} \leftarrow \mathbf{c} + \Delta\mathbf{c}$ and $\mathbf{p} \leftarrow \mathbf{p} \circ \Delta\mathbf{p}^{-1}$
7: **until** convergence

The main cost of the above algorithm is in steps 3-5. In particular, the complexity of calculating $\mathbf{J}$ and $\mathbf{J}_a$ is $O(q2K)$ and $O(pq2K)$, respectively. The complexity of calculating $\mathbf{H}_a$ and $\mathbf{H}_a^{-1}$ is $O(p^2 2K)$ and $O(p^3)$. Finally, the complexity of calculating $\Delta\mathbf{c}$ and $\Delta\mathbf{p}$ is $O(q2K)$ and $O(p2K)$.

Project-out. To derive the project-out AOM fitting algorithm we treat the first eigenvector $\mathbf{\Phi}_0$ as the "mean normalized gradient", and then perform a first order Taylor on it $\mathbf{\Phi}_0 = \mathbf{\Phi}_0[0] + \mathbf{J}_0\Delta\mathbf{p}$. Then, we write

$$||\mathbf{z}[\mathbf{p}] - \mathbf{\Phi}_0 - \mathbf{J}_0\Delta\mathbf{p}||_{\mathbf{P}}^2 = ||\mathbf{z}[\mathbf{p}] - \mathbf{J}_0\Delta\mathbf{p}||_{\mathbf{P}}^2, \qquad (23)$$

and hence the update for $\Delta\mathbf{p}$ can be readily derived as

$$\Delta\mathbf{p} = \mathbf{H}_p^{-1}\mathbf{J}_p^T\mathbf{z}[\mathbf{p}], \qquad (24)$$

where ,

$$\mathbf{J}_p = \mathbf{PJ}_0, \mathbf{H}_p = \mathbf{J}_p^T\mathbf{J}_p \qquad (25)$$

can be both pre-computed. Overall, the steps of the algorithm are summarized in Algorithm 2.

Algorithm 2 AOM - Project-Out

Given: Models $\{\mathbf{\Phi}_{S,0}, \mathbf{\Phi}_S\}$ and $\mathbf{\Phi}_Z$, and current estimates for $\mathbf{c}$ and $\mathbf{p}$

Pre-compute: $\mathbf{R} = \mathbf{H}_p^{-1}\mathbf{J}_p^T$ from (25)

1: **repeat**
2:　　Compute warped image $\mathbf{I}[\mathbf{p}]$ and extract normalized gradients $\mathbf{z}[\mathbf{p}]$
3:　　Calculate $\Delta\mathbf{p} = \mathbf{Rz}[\mathbf{p}]$
4:　　Update from $\mathbf{p} \leftarrow \mathbf{p} \circ \Delta\mathbf{p}^{-1}$
5: **until** convergence

The main cost of the above algorithm is in step 2. In particular, calculating $\Delta\mathbf{p}$ takes $O(q2K)$ only, and hence this algorithm is extremely fast.

V. RESULTS

Evaluating and comparing different methods for deformable object and face alignment is difficult, because when training and fitting codes are not publicly available, various implementation aspects including training and initialization can be very different, and this in turn possibly results in very different performance. Hence, for the sake of a fair comparison, in our experiments we did not attempt to directly compare our results with the ones reported in previously published methods. Instead, the main focus of our experiments was to compare the proposed AOMs with methods for which we have in-house implementations and hence comparison is fairer. In particular, we compare AOMs with: (a) an AAM based on pixel intensities (PI) [30], (b) a state-of-the-art Gabor-Fourier (GF) AAM [26], (c) an AAM using the edge structure gradient features of [11] (ES), (d) the CLM method of [9] using discriminatively trained HOG features for local detectors [2] and (e) our version of the SDM method of [10]. Both CLM and SDM are trained on the same data as our AAMs. All variants of AAMs were fitted using both alternating and project-out

---

[2]Our version performs significantly better than the pre-trained model provided by the authors of [9].

inverse compositional algorithms implemented in a multi-scale (pyramid) fashion with 15 shape parameters at the highest level. Note that the the edge structure gradient features of [11] are similar to the normalized gradients of AOMs, with the main difference being that [11] uses the gradient magnitude to normalize the gradients which is less robust to outliers [7]. We further note that these features have been employed only in regression-based approaches to AAM fitting, and hence their combination with non-linear optimization (alternating or project-out) is a method that is also proposed for the first time (to the best of our knowledge) in this work. Similarly to our previous work [7], we evaluate these features because they are similar to the normalized gradients of the appearance model of AOMs.

For our experiments we used the training set of LFPW [17] to train all aforementioned methods. For testing, we used the test set of LFPW, as well as the test set of Helen [16] and AFW [13]. Although all databases are in-the-wild, LPFW mainly contain images of celebrities with posed smiles while the faces of Helen and AFW seem to be much more natural, with much more shape and appearance variation, and hence are even more challenging to fit. For all databases, we used the landmark annotations of the 300-W challenge [33], [34]. In order to assess performance, we used the same average (computed over all 68 points) point-to-point Euclidean error normalized by the face size as the one used in [13]. Similarly to [13], for this error measure, we produced the cumulative curve corresponding to the percentage of test images for which the error was less than a specific value. In all cases, fitting was initialized by the bounding box of the detector proposed in [13].

Our comparisons/results are summarized in Fig. 2 and Tables 1-6. The quality of fittings produced by AOM alternating and project-out optimization can be visually compared in Fig. 3. As we may observe, AOM alternating optimization produces better fitting results especially with regards to the face boundary. Additional fitting results obtained with AOM alternating optimization and SDM on selected challenging images from the AFW database can be seen in Fig. 4. Each row of Fig. 4 highlights a different source of facial variation: age, pose, illumination, expression, occlusion and low resolution or blur. Overall, AOMs are able to fit some difficult cases of faces from Helen and AFW data sets with shape and appearance variation quite different from that seen during training on LPFW. From the presented results we conclude the following: **Alternating optimization.**(a) AOM alternating optimization algorithm performs the best among all algorithms, with ES-AAMs achieving similar performance. (b) All AAMs when trained in-the-wild and optimized using alternating optimization perform notably well. These results are not so surprising and should be attributed to the appearance model of the AAM which for all variants was trained in-the-wild. We refer the reader to [30] for a detailed discussion. (c) All AAMs when trained in-the-wild and optimized using alternating optimization largely outperform the state-of-the-art CLM method. (d) For errors greater than 0.05, SDM is the most robust method. **Project-out.**(a) AOM project-out is by far the most robust project-out algorithm, outperforming both ES-AAMs and GF-AAMs. Notably, its fitting accuracy is similar to that of PI-

Pt-Pt error	< 0.02	< 0.03	< 0.05
AOMs - A	0.33	**0.78**	0.93
PI-AAMs - A	0.2	0.62	0.91
ES-AAMs - A	**0.35**	0.74	0.92
GF-AAMs - A	0.10	0.56	0.84
CLMs	0.00	0.05	0.80
SDM	0.14	0.67	**0.98**

TABLE I
FITTING PERFORMANCE ON LPFW USING ALTERNATING OPTIMIZATION: PROPORTION OF IMAGES THAT WERE FITTED WITH PT-PT ERROR < 0.02, < 0.03 AND < 0.05.

Pt-Pt error	< 0.02	< 0.03	< 0.05
AOMs - A	**0.21**	**0.66**	0.88
PI-AAMs - A	0.10	0.50	0.82
ES-AAMs - A	0.20	0.66	0.87
GF-AAMs - A	0.02	0.37	0.84
CLMs	0.00	0.04	0.74
SDM	0.05	0.49	**0.94**

TABLE II
FITTING PERFORMANCE ON HELEN USING ALTERNATING OPTIMIZATION: PROPORTION OF IMAGES THAT WERE FITTED WITH PT-PT ERROR < 0.02, < 0.03 AND < 0.05.

AAM when optimized using alternating optimization. This is a notable result given that AOM project-out is significantly faster. These results also clearly illustrate the robustness of the proposed AOM: the project-out algorithm is literally a LK algorithm the goal of which is to align the test image with the mean image of the appearance subspace. Because the appearance model is not employed to compensate for large discrepancies in appearance between the test image and the mean (as opposed to alternating optimization), a robust algorithm is required to perform the alignment. As the presented results clearly illustrate, AOM project-out is the most robust alignment algorithm. We note that these results are in accordance with the ones presented in our previous work [7] which have shown that the kernel of AOMs performs by far the best for LK-based image-to-image alignment. (b) AOM project-out largely outperforms the state-of-the-art CLM method. This is an important result because AOM project-out is also faster than the CLM method which relies on filtering the image with HOG filters.
Cross-database experiments. As expected, for these experiments performance drops for all methods but the relative difference in performance is the similar. Again, AOM outperforms all other algorithms considered. The drop in performance should be also attributed to the less accurate initialization provided by the face detector.

VI. CONCLUSIONS

We introduced Active Orientation Models, generative models for deformable face alignment that generalize well to unseen faces and variations. AOMs employ a statistically robust appearance model based on the principal components of image
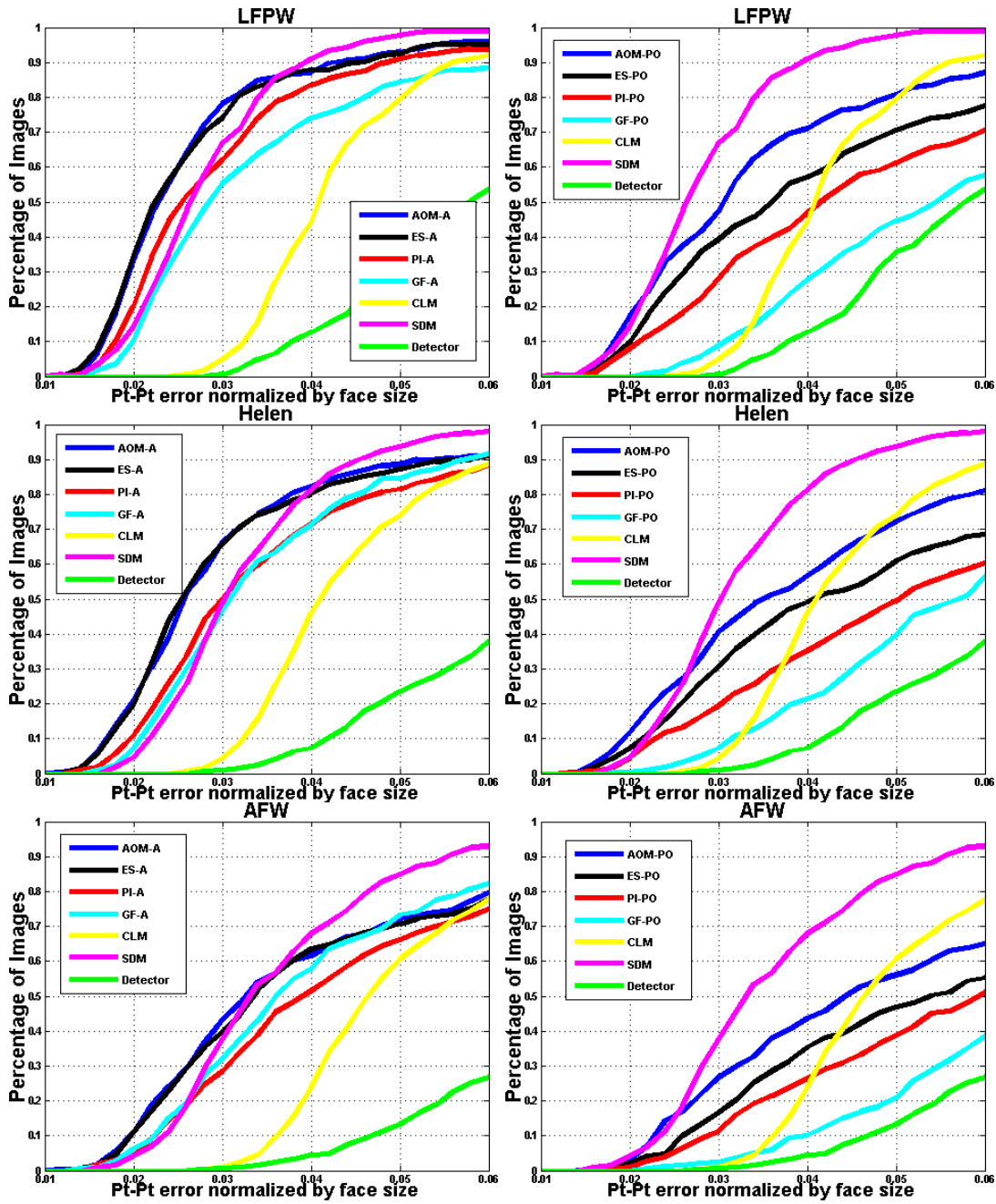
Fig. 2. Fitting performance on LPFW (first row), HELEN (second row) and AFW (third row). Left column: Alternating optimization. Right column: Project-out optimization. In all figures the point-to-point error normalized by the face size vs the percentage of test images is plotted.

Pt-Pt error	< 0.02	< 0.03	< 0.05
AOMs - A	**0.11**	**0.43**	0.72
PI-AAMs - A	0.05	0.29	0.66
ES-AAMs - A	0.11	0.40	0.71
GF-AAMs - A	0.00	0.18	0.66
CLMs	0.01	0.01	0.61
SDM	0.04	0.38	**0.85**

TABLE III
FITTING PERFORMANCE ON AFW USING ALTERNATING OPTIMIZATION:
PROPORTION OF IMAGES THAT WERE FITTED WITH PT-PT ERROR < 0.02,
< 0.03 AND < 0.05.

Pt-Pt error	< 0.02	< 0.03	< 0.05
AOMs - PO	**0.17**	0.47	0.81
PI-AAMs - PO	0.08	0.28	0.61
ES-AAMs - PO	0.10	0.39	0.71
GF-AAMs - PO	0.00	0.09	0.45
CLMs	0.00	0.05	0.79
SDM	0.14	**0.67**	**0.98**

TABLE IV
FITTING PERFORMANCE ON LPFW USING PROJECT-OUT ALGORITHM:
PROPORTION OF IMAGES THAT WERE FITTED WITH PT-PT ERROR < 0.02,
< 0.03 AND < 0.05.

Fig. 3. Comparison between AOM Alternating and Project-Out Optimization on HELEN. Odd rows: Alternating optimization. Even rows: project-out optimization. Although our system was trained on LPFW which primarily contains celebrity faces with posed smiles, it is able to fit some difficult faces from Helen with shape and appearance variation quite different from the faces of LPFW. Alternating optimization fits the face boundary better.

Pt-Pt error	< 0.02	< 0.03	< 0.05
AOMs - PO	**0.12**	0.41	0.72
PI-AAMs - PO	0.05	0.20	0.49
ES-AAMs - PO	0.07	0.31	0.61
GF-AAMs - PO	0.00	0.07	0.40
CLMs	0.00	0.04	0.74
SDM	0.05	**0.49**	**0.94**

TABLE V
FITTING PERFORMANCE ON HELEN USING PROJECT-OUT ALGORITHM: PROPORTION OF IMAGES THAT WERE FITTED WITH PT-PT ERROR < 0.02, < 0.03 AND < 0.05.

Pt-Pt error	< 0.02	< 0.03	< 0.05
AOMs - PO	0.03	0.27	0.56
PI-AAMs - PO	0.01	0.11	0.38
ES-AAMs - PO	0.02	0.17	0.47
GF-AAMs - PO	0.00	0.02	0.20
CLMs	0.00	0.01	0.60
SDM	**0.04**	**0.38**	**0.85**

TABLE VI
FITTING PERFORMANCE ON AFW USING PROJECT-OUT ALGORITHM: PROPORTION OF IMAGES THAT WERE FITTED WITH PT-PT ERROR < 0.02, < 0.03 AND < 0.05.

gradient orientations. We demonstrated that when incorporated within standard optimization frameworks for AAM learning and fitting, this kernel PCA results in robust and efficient algorithms for model fitting. Finally, we showed that the proposed AOM largely outperforms all other state-of-the-art AAM methods considered as well as a state-of-the-art CLM method, when all methods are trained on the same training set and initialized in the same way. Future work includes combining AOMs with the Gauss-Newton Deformable Part

Model which by-passes the complicated motion model of AAMs [35].

APPENDIX

In this section we will derive (20) and (21). After linearization of (17), we obtain the following optimization problem

$$\arg \min_{\Delta\mathbf{p},\Delta\mathbf{c}} ||\mathbf{z}[\mathbf{p}] - \mathbf{\Phi}_Z\mathbf{c} - \mathbf{\Phi}\Delta\mathbf{c} - \mathbf{J}\Delta\mathbf{p}||^2. \quad (26)$$

To solve the above problem our strategy is to optimize (26) first with respect to $\Delta\mathbf{c}$, and then plug in the solution back to (26). Then, we can optimize (26) with respect to $\Delta\mathbf{p}$. In particular, let us denote by $\mathbf{C}_1 = \mathbf{z}[\mathbf{p}] - \mathbf{\Phi}\Delta\mathbf{c} - \mathbf{J}\Delta\mathbf{p}$. Then, we have

$$\arg \min_{\Delta\mathbf{c}}(\mathbf{C}_1 - \mathbf{\Phi}_Z\mathbf{c})^T(\mathbf{C}_1 - \mathbf{\Phi}\mathbf{c}). \quad (27)$$

By setting the derivative of the above with respect to $\Delta\mathbf{c}$ to $\mathbf{0}$, we readily obtain (20)

$$\Delta\mathbf{c} = \mathbf{\Phi}_Z^T\mathbf{C}_1 = \mathbf{\Phi}_Z^T(\mathbf{z}[\mathbf{p}] - \mathbf{\Phi}\Delta\mathbf{c} - \mathbf{J}\Delta\mathbf{p}). \quad (28)$$

Plugging the above to (26) and, after some straightforward mathematical manipulations, we get the following optimization problem for $\Delta\mathbf{p}$

$$\arg \min_{\Delta\mathbf{p}}(\mathbf{z}[\mathbf{p}] - \mathbf{J}\Delta\mathbf{p})^T\mathbf{P}(\mathbf{z}[\mathbf{p}] - \mathbf{J}\Delta\mathbf{p}), \quad (29)$$

where $\mathbf{P} = \mathbf{E} - \mathbf{\Phi}_Z\mathbf{\Phi}_Z^T$. By setting the derivative of the above with respect to $\Delta\mathbf{p}$ to $\mathbf{0}$, we readily obtain (21).
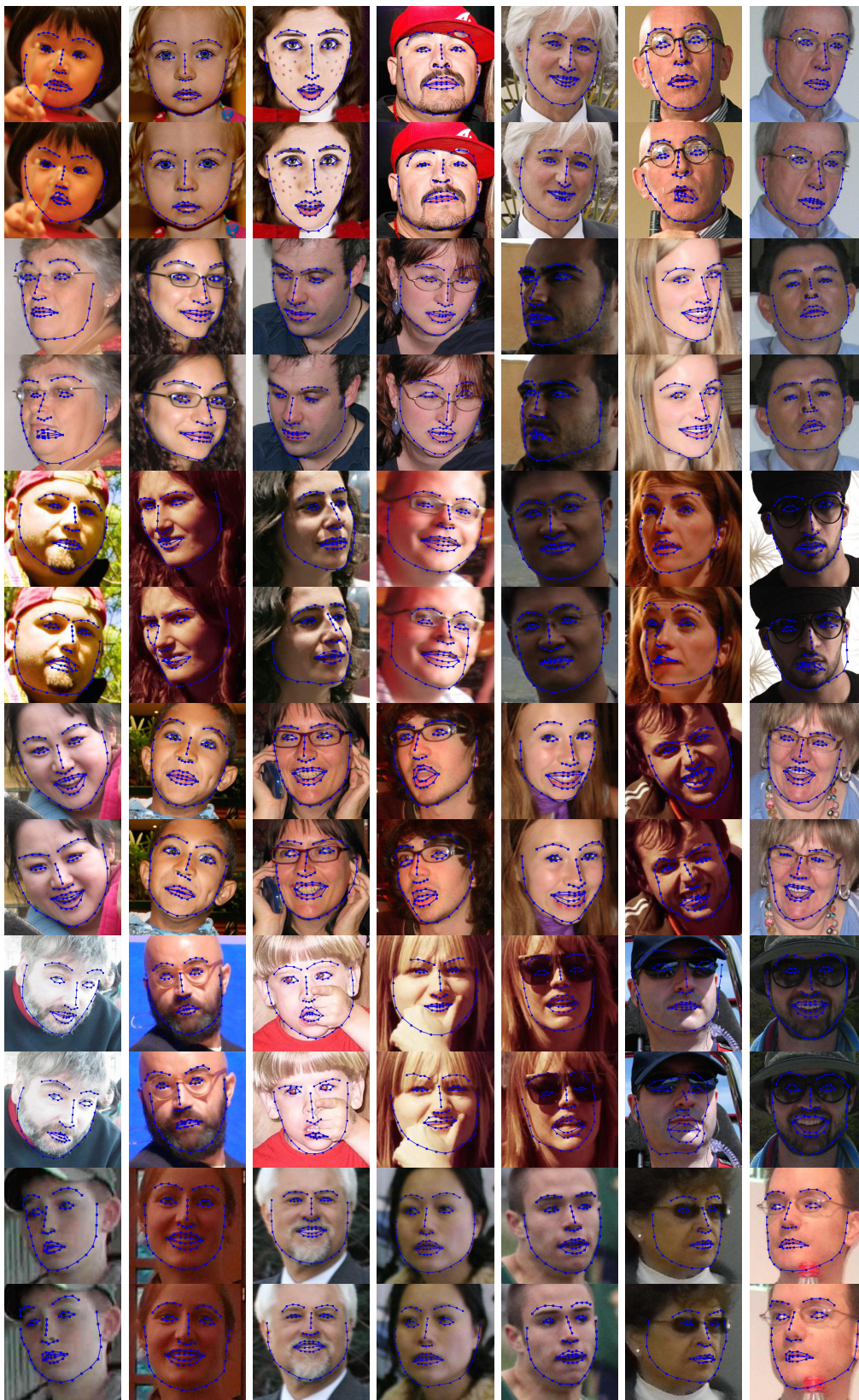
Fig. 4. Fitting results obtained with our method and SDM on selected images from the AFW database. Different rows highlight a different source of facial variation: age, pose, illumination, expression, occlusion and low resolution or blur. Odd rows: AOMs. Even rows: SDM

REFERENCES

[1] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, "Active shape models-their training and application," *CVIU*, vol. 61, no. 1, pp. 38–59, 1995.

[2] D. Cristinacce and T. Cootes, "Automatic feature localisation with constrained local models," *Pattern Recognition*, vol. 41, no. 10, pp.

3054–3067, 2008.

[3] J.M. Saragih, S. Lucey, and J.F. Cohn, "Face alignment through subspace constrained mean-shifts," in *ICCV*, 2009.

[4] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active appearance models," *TPAMI*, vol. 23, no. 6, pp. 681–685, 2001.

[5] I. Matthews and S. Baker, "Active appearance models revisited," *IJCV*, vol. 60, no. 2, pp. 135–164, 2004.

[6] R. Gross, I. Matthews, and S. Baker, "Generic vs. person specific active appearance models," *Image and Vision Computing*, vol. 23, no. 12, pp. 1080–1093, 2005.

[7] G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "Robust and efficient parametric face alignment," in *ICCV*, 2011.

[8] G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "Subspace learning from image gradient orientations," *IEEE TPAMI*, vol. 34, no. 12, pp. 2454–2466, 2012.

[9] J.M. Saragih, S. Lucey, and J.F. Cohn, "Deformable model fitting by regularized landmark mean-shift," *IJCV*, vol. 91, no. 2, pp. 200–215, 2011.

[10] Xuehan Xiong and Fernando De la Torre, "Supervised descent method and its applications to face alignment," 2013.

[11] T.F. Cootes and C.J. Taylor, "On representing edge structure for model matching," in *CVPR*, 2001.

[12] M. Valstar, B. Martinez, X. Binefa, and M. Pantic, "Facial point detection using boosted regression and graph models," in *CVPR*, 2010.

[13] X. Zhu, , and D. Ramanan, "Face detection, pose estimation, and landmark estimation in the wild.," in *CVPR*, 2012.

[14] S. Lucey, Y. Wang, M. Cox, S. Sridharan, and J.F. Cohn, "Efficient constrained local model fitting for non-rigid face alignment," *Image and Vision Computing*, vol. 27, no. 12, pp. 1804–1813, 2009.

[15] Pedro Martins, Rui Caseiro, João F Henriques, and Jorge Batista, "Discriminative bayesian active shape models," in *ECCV*. 2012.

[16] Vuong Le, Jonathan Brandt, Zhe Lin, Lubomir Bourdev, and Thomas S Huang, "Interactive facial feature localization," in *ECCV*. 2012.

[17] P.N. Belhumeur, D.W. Jacobs, D.J. Kriegman, and N. Kumar, "Localizing parts of faces using a consensus of exemplars," in *CVPR*, 2011.

[18] Xudong Cao, Yichen Wei, Fang Wen, and Jian Sun, "Face alignment by explicit shape regression," in *CVPR*, 2012.

[19] Tim F Cootes, Mircea C Ionita, Claudia Lindner, and Patrick Sauer, "Robust and accurate shape model fitting using random forest regression voting," in *ECCV*. 2012.

[20] Georgios Tzimiropoulos, Joan Alabort-i Medina, Stefanos Zafeiriou, and Maja Pantic, "Generic active appearance models revisited," in *ACCV 2012*, 2013.

[21] X. Liu, "Generic face alignment using boosted appearance model," in *CVPR*, 2007.

[22] H. Wu, X. Liu, and G. Doretto, "Face alignment via boosted ranking model," in *CVPR*, 2008.

[23] J. Saragih and R. Gocke, "Learning aam fitting through simulation," *Pattern Recognition*, vol. 42, no. 11, pp. 2628–2636, 2009.

[24] J. Saragih and R. Goecke, "A nonlinear discriminative approach to aam fitting," in *ICCV*, 2007.

[25] P Kittipanya-ngam and TF Cootes, "The effect of texture representations on aam performance," in *ICPR*, 2006.

[26] R. Navarathna, S. Sridharan, and S. Lucey, "Fourier active appearance models," in *ICCV*, 2011.

[27] S. Baker, R. Gross, and I. Matthews, "Lucas-kanade 20 years on: Part 3," *Robotics Institute, Carnegie Mellon University, Tech. Rep. CMU-RI-TR-03-35*, 2003.

[28] Bruce D Lucas, Takeo Kanade, et al., "An iterative image registration technique with an application to stereo vision," in *Proceedings of the 7th international joint conference on Artificial intelligence*, 1981.

[29] Gregory D Hager and Peter N Belhumeur, "Efficient region tracking with parametric models of geometry and illumination," *IEEE TPAMI*, vol. 20, no. 10, pp. 1025–1039, 1998.

[30] Georgios Tzimiropoulos and Maja Pantic, "Optimization problems for fast aam fitting in-the-wild," in *ICCV*, 2013.

[31] George Papandreou and Petros Maragos, "Adaptive and constrained algorithms for inverse compositional active appearance model fitting," in *CVPR*. IEEE, 2008.

[32] F. De La Torre and M.J. Black, "A framework for robust subspace learning," *IJCV*, vol. 54, no. 1, pp. 117–142, 2003.

[33] Christos Sagonas, Georgios Tzimiropoulos, Stefanos Zafeiriou, and Maja Pantic, "A semi-automatic methodology for facial landmark annotation," in *CVPR-W*, 2013.

[34] Christos Sagonas, Georgios Tzimiropoulos, Stefanos Zafeiriou, and Maja Pantic, "300 faces in-the-wild challenge: The first facial landmark localization challenge," in *Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on*. IEEE, 2013.

[35] Georgios Tzimiropoulos and Maja Pantic, "Gauss-newton deformable part models for face alignment in-the-wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1851–1858.
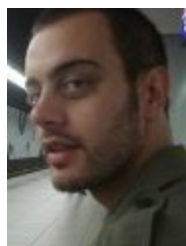
Georgios Tzimiropoulos Georgios Tzimiropoulos received the diploma degree in Electrical and Computer Engineering from Aristotle University of Thessaloniki, Greece, and the M.Sc. and Ph.D. degrees in Signal Processing and Computer Vision both from Imperial College London, UK. He is currently a Senior Lecturer (Assistant Professor) with the School of Computer Science at the University of Lincoln, UK. After the completion of his Ph.D., he worked as a Research and Development engineer with Imperial College/Selex Galileo. From 2010 to 2012, he was a Research Associate in the Department of Computing at Imperial College working on face analysis. He is currently an Associate Editor of the Image and Vision Computing Journal. His main research interests are in the areas of face and object recognition, alignment and tracking and facial expression analysis.

Joan Alabort-i-Medina received a BSc degree in Computer Science and Engineering from Universitat Autnoma de Barcelona in 2008 and a MSc degree in Visual Information Processing from Imperial College London in 2011. He is currently working towards a PhD degree in Computer Vision at Imperial College London. His research interests lie in the fields of Computer Vision, Machine Learning and Human-Computer Interaction.

Stefanos Zafeiriou (M'09), is a Lecturer (equivalent to Assistant Professor) in Pattern Recognition/Statistical Machine Learning for Computer Vision in the Department of Computing, Imperial College London. He has been awarded one of the prestigious Junior Research Fellowships (JRF) from Imperial College London in 2011 to start his own independent research group. He has received various awards during his doctoral and postdoctoral studies. Dr. Zafeiriou currently serves as an Associate Editor in IEEE Transactions on Cybernetics and Image and Vision Computing journal. He has been guest editor in more than five journal special issues and co-organized more than five workshops/ special sessions in top venues such as CVPR/FG/ICCV/ECCV. He has co-authored more than 40 journal papers mainly on novel statistical machine learning methodologies applied to computer vision problems such as 2D/3D face analysis, deformable object fitting and tracking, shape from shading, human behaviour analysis etc published in the most prestigious journals in his field of research (such as IEEE T-PAMI, IJCV, IEEE T-IP, IEEE T-NNLS, IEEE T-VCG, IEEE T-IFS etc), as well many papers in top conferences (such as CVPR, ICCV, ECCV, ICML etc). His students are frequent recipients of very prestigious and highly competitive fellowships such as Google Fellowship, Intel Fellowship and the Qualcomm fellowship.

Maja Pantic (M98, SM06, F12) is Professor in Affective and Behavioural Computing at Imperial College London, Department of Computing, UK, and at the University of Twente, Department of Computer Science, the Netherlands. She received various awards for her work on automatic analysis of human behaviour including the European Research Council Starting Grant Fellowship 2008 and the Roger Needham Award 2011. She currently serves as the Editor in Chief of Image and Vision Computing Journal and as an Associate Editor for both the IEEE Trans. Pattern Analysis and Machine Intelligence and the IEEE Trans. Cybernetics.