FAST AND EXACT BI-DIRECTIONAL FITTING OF ACTIVE APPEARANCE MODELS

Jean Kossaifi[⋆] *Georgios Tzimiropoulos*[†] *Maja Pantic*[⋆1]

[⋆] Imperial College London, UK, Department of Computing
[†]University of Nottingham, UK, School of Computer Science
[1] University of Twente, The Netherlands

ABSTRACT

Finding landmarks on objects like faces is a challenging computer vision problem, especially in real life conditions (or in-the-wild) and Active Appearance Models have been widely used to solve it. State-of-the-art algorithms for fitting an AAM to a new image are based on Gauss-Newton (GN) optimization. Recently fast GN algorithms have been proposed for both forward additive and inverse compositional fitting frameworks. In this paper, we propose a fast and exact bi-directional (Fast-Bd) approach to AAM fitting by combining both approaches. Although such a method might appear to increase computational burden, we show that by capitalizing on results from optimization theory, an exact solution, as computationally efficient as the original forward or inverse formulation, can be derived. Our proposed bi-directional approach achieves state-of-the-art performance and superior convergence properties. These findings are validated on two challenging, in-the-wild data sets, LFPW and Helen, and comparison is provided to the state-of-the art methods for Active Appearance Models fitting.

Index Terms— Active Appearance Models, Gauss-Newton, forward additive, inverse compositional, bi-directional fitting.

1. INTRODUCTION

Active Appearance Models are generative deformable models of shape and appearance widely used in computer vision and in particular for face and medical image analysis [1]. Fitting an AAM to a new image is usually formulated as a non-linear least-squares problem which is typically solved using iterative methods. State-of-the-art methods for AAM fitting are based on analytic gradient descent and, in particular, on Gauss-Newton (GN) optimization [2]. The problem can also be solved efficiently using fast full-Newton method as presented in [3].

GN optimization in computer vision goes back to the classical Lukas-Kanade image alignment algorithm [4] and the appearance-based tracking framework of Hager and Belhumeur [5]. In the context of AAM fitting, GN optimization was introduced in the seminal work of Matthews and Baker [2]. In this work, the authors proposed a very efficient GN algorithm for AAM fitting which was coined project-out inverse compositional algorithm (POIC). POIC has two main features: (a) it decouples shape from appearance by projecting out appearance variations and (b) applies the so-called inverse composition by computing a warp update in the model coordinate frame which is then composed to the current warp estimate. This is in contrast to the standard LK algorithm in which the warp parameters are updated in a forward additive fashion. Although being an approximate algorithm, owing to its efficiency, POIC has become the standard approach for fitting person specific AAMs.

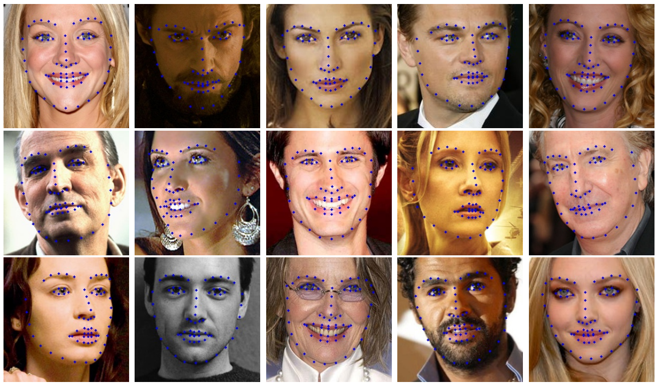


Fig. 1. Examples of images taken from the LFPW dataset and fitted with our proposed bi-directional method.

Following the seminal work of [6], inverse algorithms have gained increased popularity. Note however that not all inverse algorithms are computationally efficient. This is particularly true for the simultaneous inverse compositional (SIC) algorithm which, albeit exact and very robust, has a computational cost which is almost prohibitive for most current systems [7, 8].

Recently, the authors of [9] proposed Fast-SIC, an efficient algorithm for solving the original SIC problem without resorting to any approximations at all. In the same work, the authors have shown that one can actually devise a GN forward additive algorithm, called Fast-Forward, which is also very computationally efficient. In this work, we build upon

[9] to propose an algorithm which simultaneously solves the forward and inverse problems, and hence is called "bi-directional". Although such an approach might appear to increase computational burden, we show that one can come up with an exact solution which is as computationally efficient as the original forward or inverse formulation. At the same time, the proposed Fast Bi-Directional approach achieves state-of-the art performance and superior convergence properties. We verify these findings on two in-the-wild data sets, namely LFPW [10] and Helen [11]. Finally, we emphasize that although a somewhat similar Bi-Directional approach was proposed in [12], our method capitalizes on optimization theory to provide a both exact and computationally efficient solution. In contrast, the method in [12] is both inexact and slower.

2. ACTIVE APPEARANCE MODELS

Models. Active Appearance Models are generative models of shape and appearance. The shape model is obtained by first annotating the location of u landmarks across a training set of objects belonging to the same class (e.g. faces in our case) before normalizing the resulting annotated shapes using Procrustes Analysis. This step removes variations due to translation, scaling and rotation. PCA is then performed on these normalized shapes and the first n shape eigenvectors are kept as the column of $\mathbf{S} \in \mathbb{R}^{2u,n}$ to define the shape model, along with the mean shape $\mathbf{s}_0 \in \mathbb{R}^{2u}$. A shape can then be generated from $\hat{\mathbf{s}} = \mathbf{s}_0 + \mathbf{S}\mathbf{p}$, where $\mathbf{p} \in \mathbb{R}^n$ is a vector representing the shape parameters. Similarly, the appearance model is obtained from the texture of the training images, after appearance variation due to shape deformation has been removed by warping each texture into the mean shape $\mathbf{s}_0$ using the motion model $\mathbf{W}$, which in this work is assumed to be a piecewise affine warp. After PCA has been applied to all training shape-free textures, the resulting texture eigenvectors are stacked as columns of $\mathbf{A} \in \mathbb{R}^{N,m}$ and the mean texture is noted $\mathbf{A}_0 \in \mathbb{R}^N$. This constitutes the appearance model which can be used to generate a texture from $\hat{\mathbf{I}} = \mathbf{A}_0 + \mathbf{A}\mathbf{c}$, where $\mathbf{c} \in \mathbb{R}^m$ is a vector representing the texture parameters. Finally, a model instance is synthesized to represent a test object by warping a texture instance from the mean shape $\mathbf{s}_0$ to a shape instance $\mathbf{s}$ using the piecewise affine warp $\mathbf{W}$ defined by $\mathbf{s}_0$ and $\mathbf{s}$. Please see [2] for more details on AAMs.

Objective function. Given the shape and appearance models, the problem of finding facial landmarks in a new image can be formulated as finding the shape and appearance parameters such that a model instance is "close" to the given image usually in a least-squares sense. This is equivalent to solving the following non-linear least-squares problem:

$$\arg\min_{\mathbf{p},\mathbf{c}} \|\mathbf{I}[\mathbf{p}] - \mathbf{A}_0 - \mathbf{A}\mathbf{c}\|^2. \tag{1}$$

We vectorise the computation over all N pixels $\mathbf{x}$ of the image and denote $\mathbf{I}[\mathbf{p}]$ the warped image $\mathbf{I}(\mathbf{W}(\mathbf{x},\mathbf{p}))$ rearranged as a vector of size N. The above cost function can be optimized iteratively using Gauss-Newton in two coordinate frames. In the forward algorithm, the image $\mathbf{I}$ is linearized around $\mathbf{p}$, an update $\Delta\mathbf{p}$ is found using least-squares, and $\mathbf{p}$ is updated from $\mathbf{p} \leftarrow \mathbf{p} + \Delta\mathbf{p}$. In the inverse algorithm, the model $\{\mathbf{A}_0, \mathbf{A}\}$ is linearized around $\mathbf{p} = \mathbf{0}$ using the fact that $\mathbf{W}(\mathbf{x};\mathbf{p})$ is the identity for $\mathbf{p} = \mathbf{0}$. An update $\Delta\mathbf{p}$ is then found using least-squares and $\mathbf{p}$ is updated in a compositional fashion $\mathbf{p} \leftarrow \mathbf{p} \circ \Delta\mathbf{p}^{-1}$. Please see [6] for more details on AAMs.

SIC. At each iteration SIC (Simultaneous Inverse Compositional) linearizes (1) with respect to both $\mathbf{c}$ and $\mathbf{p} = \mathbf{0}$. This is equivalent to solving, at each iteration, the following optimization problem:

$$\arg\min_{\Delta\mathbf{p},\Delta\mathbf{c}} \|\mathbf{I} - \mathbf{A}_0 - \mathbf{A}\mathbf{c} - \mathbf{A}\Delta\mathbf{c} - \mathbf{J}_T\Delta\mathbf{p}\|^2, \tag{2}$$

where $\mathbf{J}_T \in \mathbb{R}^{N,n}$ is the Jacobian matrix of the template, $\mathbf{J}_T = \mathbf{J}_0 + \sum_{i=1}^m c_i\mathbf{J}_i$, with $\mathbf{J}_i = [\mathbf{A}_{i,x} \quad \mathbf{A}_{i,y}]\frac{\partial\mathbf{W}(\mathbf{x};\mathbf{p})}{\partial\mathbf{p}}$. $\mathbf{A}_{i,x}$ and $\mathbf{A}_{i,y} \in \mathbb{R}^{1,N}$ are the x and y gradients of $\mathbf{A}_i$ and $\frac{\partial\mathbf{W}(\mathbf{x};\mathbf{p})}{\partial\mathbf{p}} \in \mathbb{R}^{2,n}$ is the Jacobian matrix of the piecewise affine warp. All of these are defined in the model coordinate frame for $\mathbf{p} = \mathbf{0}$ and can be pre-computed. One can show that the cost per iteration for SIC is $O((n+m)^2N)$ and hence this algorithm is very slow [7]. For more detail refer to [9].

POIC. POIC (Project Out Inverse Compositional) reduces the cost of SIC by solving (2) in the subspace orthogonal to $\mathbf{A}$. Let us define the projection operator $\mathbf{P} = \mathbf{E} - \mathbf{A}\mathbf{A}^T$, where $\mathbf{E}$ is the identity matrix. Then, $\|\mathbf{I} - \mathbf{A}_0 - \mathbf{A}\mathbf{c}\|^2_{\mathbf{P}} = \|\mathbf{I} - \mathbf{A}_0\|^2_{\mathbf{P}}$, where we write $\|\mathbf{x}\|^2_{\mathbf{P}}$ to denote the weighted ℓ_2-norm $\mathbf{x}^T\mathbf{P}\mathbf{x}$. Based on this, POIC computes an update for $\Delta\mathbf{p}$ by optimizing

$$\arg\min_{\Delta\mathbf{p}} \|\mathbf{I} - \mathbf{A}_0 - \mathbf{J}_0\Delta\mathbf{p}\|^2_{\mathbf{P}}. \tag{3}$$

One can show that solving the above optimization problem has a cost of $O(nN)$ only [6].

Fast-SIC. Fast-SIC capitalizes on optimization theory [13]

$$\min_{x,y} f(x,y) = \min_x [\min_y f(x,y)] \tag{4}$$

to solve (2) in a computationally efficient way. Using (4), we can firstly optimize (2) with respect to $\Delta\mathbf{c}$:

$$\Delta\mathbf{c} = \mathbf{A}^T(\mathbf{I} - \mathbf{A}_0 - \mathbf{A}\mathbf{c} - \mathbf{J}_T\Delta\mathbf{p}). \tag{5}$$

Plugging the above into (2), we get

$$\arg\min_{\Delta\mathbf{p}} \|\mathbf{I} - \mathbf{A}_0 - \mathbf{J}_T\Delta\mathbf{p}\|^2_{\mathbf{P}}. \tag{6}$$

One can show that solving the above optimization problem has a cost $O(nmN + n^2N)$ which is much less than

$O((n+m)^2N)$ for the original SIC algorithm [9].

Fast-Forward. Fast-Forward capitalizes on (4) to solve problem (1) efficiently by linearizing the test image rather than the model:

$$\arg \min_{\{\Delta\mathbf{p},\mathbf{c}\}} \|\mathbf{I} + \mathbf{J}_I \Delta\mathbf{p} - \mathbf{A}_0 - \mathbf{A}\mathbf{c}\|^2, \tag{7}$$

where $\mathbf{p} \in \mathbb{R}^n$ and $\mathbf{J}_I$ is the Jacobian matrix of the image $\mathbf{I}$,

$$\mathbf{J}_I = \frac{\partial \mathbf{I}[\mathbf{p}]}{\partial \mathbf{p}} \in \mathbb{R}^{N,n}.$$

At each iteration, the optimal $\mathbf{c}$ is given by

$$\mathbf{c} = \mathbf{A}^T(\mathbf{I} + \mathbf{J}_I \Delta\mathbf{p} - \mathbf{A}_0). \tag{8}$$

Plugging the above into (7), we get

$$\arg\min_{\Delta\mathbf{p}} \|\mathbf{I} + \mathbf{J}_I \Delta\mathbf{p} - \mathbf{A}_0\|^2_{\mathbf{P}}. \tag{9}$$

Similarly, one can show that solving the above optimization problem has a cost $O(nmN + n^2N)$ [9].

Bi-directional. In [12], an approximate bi-bidirectional approach is presented where the similarity parameters are updated in a forward additive fashion while the appearance and shape parameters are optimised jointly in an inverse compositional fashion. However, the solution proposed does not use the structure of the problem resulting in a computationally complex algorithm in $O(N(m+n)^2)$. In addition, the solution presented is approximate as second order terms are neglected.

3. FAST AND EXACT BI-DIRECTIONAL FITTING OF ACTIVE APPEARANCE MODELS

In this paper, we propose a fast and exact bi-directional Gauss-Newton algorithm for AAM fitting by deforming at each iteration both the image and the template while also optimising the appearance parameters. To achieve this, we linearize both the image as in (7) and the template as in (2) and optimize jointly over all three parameters $\Delta\mathbf{q}, \Delta\mathbf{p}$ and $\Delta\mathbf{c}$:

$$\arg \min_{\Delta\mathbf{q},\Delta\mathbf{p},\Delta\mathbf{c}} \|\mathbf{I} + \mathbf{J}_I \Delta\mathbf{q} - \mathbf{A}_0 - \mathbf{A}\mathbf{c} - \mathbf{A}\Delta\mathbf{c} - \mathbf{J}_T \Delta\mathbf{p}\|^2. \tag{10}$$

To solve (10) in a computationally efficient way, we additionally propose to capitalize on

$$\min_{x,y,z} f(x,y,z) = \min_x[\min_y[\min_z f(x,y,z)]]. \tag{11}$$

In particular, we can firstly optimize (10) with respect to $\Delta\mathbf{c}$ which yields

$$\Delta\mathbf{c} = \mathbf{A}^T(\mathbf{I} + \mathbf{J}_I \Delta\mathbf{q} - \mathbf{A}_0 - \mathbf{A}\mathbf{c} - \mathbf{J}_T \Delta\mathbf{p}). \tag{12}$$

Plugging the result back into (10) gives the following optimization problem:

$$\arg \min_{\Delta\mathbf{q},\Delta\mathbf{p}} \|\mathbf{I} + \mathbf{J}_I \Delta\mathbf{q} - \mathbf{A}_0 - \mathbf{J}_T \Delta\mathbf{p}\|^2_{\mathbf{P}}, \tag{13}$$

using the projection operator $\mathbf{P} = \mathbf{E} - \mathbf{A}\mathbf{A}^T$, where $\mathbf{E}$ is the identity matrix (as specified in the introduction, we write $\|\mathbf{x}\|^2_{\mathbf{P}}$ to denote the weighted ℓ_2-norm $\mathbf{x}^T\mathbf{P}\mathbf{x}$). We go on by optimizing (13) with respect to $\Delta\mathbf{q}$. This gives

$$\Delta\mathbf{q} = -\mathbf{H}_q^{-1}\mathbf{J}_q^T(\mathbf{I} - \mathbf{A}_0 - \mathbf{J}_T \Delta\mathbf{p}), \tag{14}$$

where the projected-out Jacobian and Hessian matrices are given by $\mathbf{J}_q = \mathbf{P}\mathbf{J}_I \in \mathbb{R}^{N,n}$ and $\mathbf{H}_q = \mathbf{J}_q^T\mathbf{J}_q \in \mathbb{R}^{n,n}$, respectively. Next, we plug (14) into (13), to get the following optimization problem

$$\arg\min_{\Delta\mathbf{p}} \|\mathbf{I} - \mathbf{A}_0 - \mathbf{J}_T \Delta\mathbf{p}\|^2_{\mathbf{R}}, \tag{15}$$

where $\mathbf{R} = \mathbf{P}(\mathbf{E} - \mathbf{Q})$ and $\mathbf{Q} = \mathbf{J}_q\mathbf{H}_q^{-1}\mathbf{J}_q^T$. The final step is to optimize (15) with respect to $\Delta\mathbf{p}$. This gives:

$$\Delta\mathbf{p} = \mathbf{H}_p^{-1}\mathbf{J}_p^T(\mathbf{I} - \mathbf{A}_0), \tag{16}$$

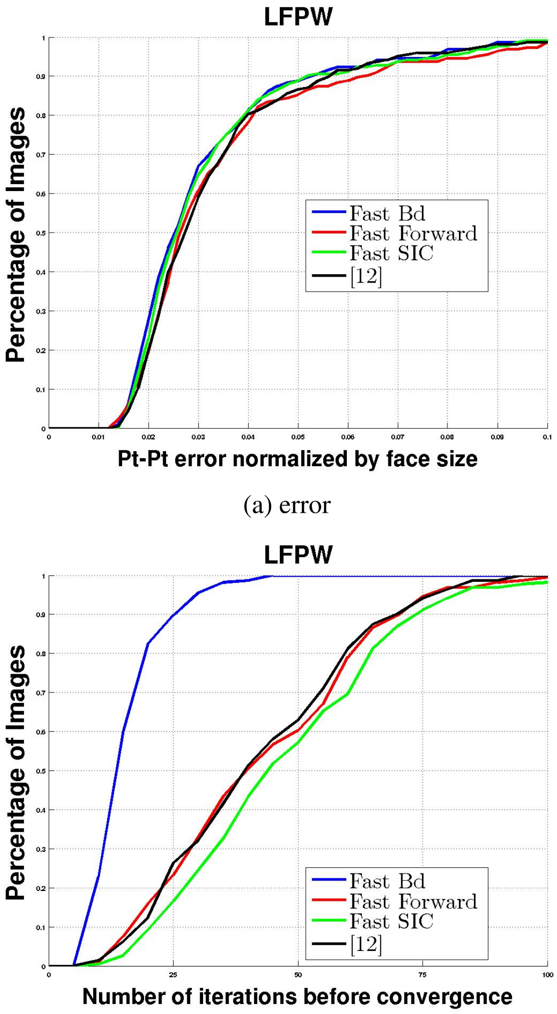
where the projected-out Jacobian and Hessian matrices are given by $\mathbf{J}_p = \mathbf{R}\mathbf{J}_T \in \mathbb{R}^{N,n}$ and $\mathbf{H}_p = \mathbf{J}_p^T\mathbf{J}_p \in \mathbb{R}^{n,n}$, respectively. Finally the shape and appearance parameters are updated as $\mathbf{q} \leftarrow \mathbf{q} \circ \Delta\mathbf{p}^{-1} + \Delta\mathbf{q}$ and $\mathbf{c} \leftarrow \mathbf{c} + \Delta\mathbf{c}$.

The complexity of computing the above updates per iteration is readily given by $O(nmN + n^2N)$.

4. EXPERIMENTS

We tested the proposed Fast-Bd algorithm on two very challenging data sets and compared it to the state of the art for AAM fitting (Fast SIC and Fast Forward), as well as to [12] which we implemented. For training, we used the training set of LFPW data set [10]. For testing, we used the test set of LFPW and also verified our findings on Helen [11]. For both data sets, we used the 68-point landmark annotations provided in [14, 15]. In all cases, fitting was initialized by the face detector recently proposed in [16]. Finally, we fitted AAMs in two scales with 7 and 14 shape eigenvectors and 50 and 400 texture eigenvectors, respectively.

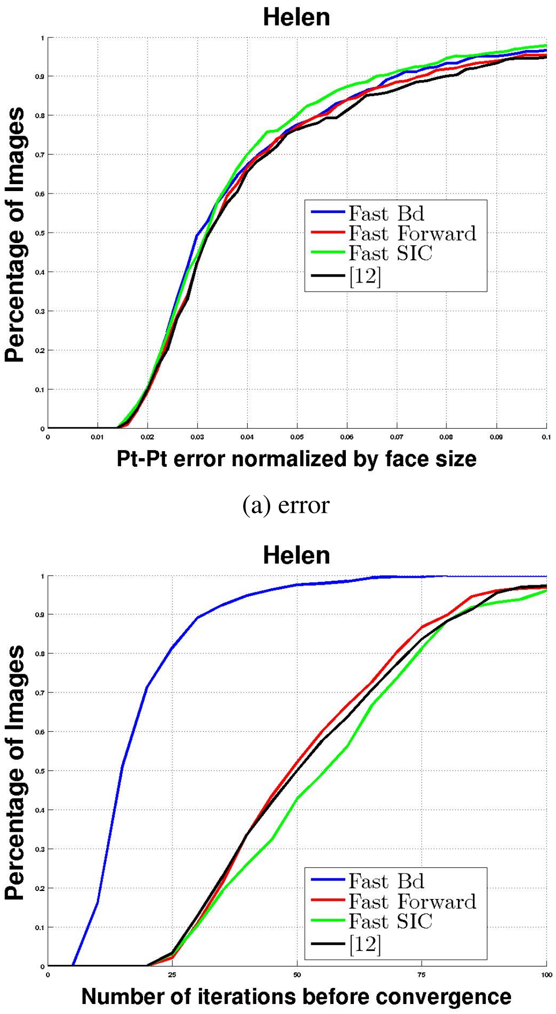
We measured fitting accuracy by producing the familiar cumulative curve corresponding to the percentage of test images for which the error between the ground truth landmarks and the fitted shape was less than a specific value. As error metric, we used the point-to-point error normalized by the face size [16]. To measure speed of convergence, we considered that an algorithm converged when $\left|\frac{\mathbf{e}_k - \mathbf{e}_{k+1}}{\mathbf{e}_k}\right| < \epsilon$, with $\epsilon \in \mathbb{R}^*_+$ a convergence threshold and $\mathbf{e}_k$ the value of the objective function ($\|\mathbf{I} - \mathbf{A}_0 - \mathbf{A}\mathbf{c}\|^2$) at iteration k.



(a) error

(b) convergence

Fig. 2. Results on the LFPW dataset.



(a) error

(b) convergence

Fig. 3. Results on the Helen dataset.

Fig.2 shows the obtained results on LFPW. Our bidirectional version (Fast Bd) performs the same as the Fast-SIC and better than the Fast-Forward and [12] while converging much faster. The same can be observed on Helen (Fig.3), although this time our method performs slightly worse than the Fast-SIC, but still better than the Fast-Forward and [12]. Again, our method has the fastest convergence rate by far.

5. CONCLUSION

We introduced a new fast and exact way of solving bidirectionally the AAM problem. We tested our method on two challenging datasets, compared it to state of the art algorithms for AAM fitting and provided the derivation of the update rule, as well as the algorithmic complexity. Our method yields state-of-the-art results while converging much faster and offering the same computational complexity as the Fast-SIC. In the future we aim to combine our Fast-Bd fitting approach with the generative deformable part model of [17], explore the use of robust features [18, 19, 20], and try to apply bidirectional fitting to regression-based methods [21].

6. ACKNOWLEDGEMENTS

This work has been funded by the European Community 7th Framework Programme [FP7/2007-2013] under grant agreement no. 611153 (TERESA). The work of Maja Pantic is also funded in part by the European Community Horizon 2020 [H2020/2014-2020] under grant agreement no. 645094 (SEWA).

7. REFERENCES

[1] Gareth J. Edwards, Christopher J. Taylor, and Timothy F. Cootes, "Interpreting face images using active appearance models," in *FG*. 1998, pp. 300–305, IEEE Computer Society. 1

[2] Iain Matthews and Simon Baker, "Active appearance models revisited," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 135 – 164, November 2004. 1, 2

[3] J. Kossaifi, G. Tzimiropoulos, and M. Pantic, "Fast newton active appearance models," in *IEEE International Conference on Image Processing (ICIP)*, 2014. 1

[4] 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. 1

[5] 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. 1

[6] I. Matthews and S. Baker, "Active appearance models revisited," *IJCV*, vol. 60, no. 2, pp. 135–164, 2004. 1, 2

[7] 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. 1, 2

[8] 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. 1

[9] G. Tzimiropoulos and M. Pantic, "Optimization problems for fast aam fitting in-the-wild," in *Proceedings of IEEE Intl Conf. on Computer Vision (ICCV 2013)*. 1, 2, 3

[10] Peter N. Belhumeur, David W. Jacobs, David J. Kriegman, and Neeraj Kumar, "Localizing parts of faces using a consensus of exemplars," in *The 24th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2011. 2, 3

[11] J. Brandt F. Zhou and Z. Lin, "Exemplar-based graph matching for robust facial landmark localization," in *IEEE International Conference on Computer Vision (ICCV)*, 2013. 2, 3

[12] A. Mollahosseini and M.H. Mahoor, "Bidirectional warping of active appearance model," in *Computer Vision and Pattern Recognition Workshops (CVPRW)*, June 2013, pp. 875–880. 2, 3, 4

[13] Stephen Boyd and Lieven Vandenberghe, *Convex optimization*, Cambridge university press, 2004. 2

[14] Christos Sagonas, Georgios Tzimiropoulos, Stefanos Zafeiriou, and Maja Pantic, "A semi-automatic methodology for facial landmark annotation," in *CVPR Workshops*, 2013. 3

[15] Christos Sagonas, Georgios Tzimiropoulos, Stefanos Zafeiriou, and Maja Pantic, "300 faces in-the-wild challenge: The first facial landmark localization challenge," in *The IEEE International Conference on Computer Vision (ICCV) Workshops*, December 2013. 3

[16] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark estimation in the wild.," in *CVPR*, 2012. 3

[17] Georgios Tzimiropoulos and Maja Pantic, "Gauss-newton deformable part models for face alignment in-the-wild," in *CVPR*, 2014. 4

[18] Georgios Tzimiropoulos, Joan Alabort-i-Medina, Stefanos Zafeiriou, and Maja Pantic, "Generic active appearance models revisited," in *Computer Vision–ACCV 2012*, pp. 650–663. Springer, 2013. 4

[19] A. Asthana, S. Zafeiriou, G. Tzimiropoulos, S. Cheng, and M. Pantic, "From pixels to response maps: Discriminative image filtering for face alignment in the wild," *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI). In Press.*, 2015. 4

[20] E. Antonakos, J. Alabort-i-Medina, G. Tzimiropoulos, and S. Zafeiriou, "Feature-based lucas-kanade and active appearance models," *IEEE Transactions on Image Processing*, Accepted for publication. 4

[21] Georgios Tzimiropoulos, "Project-out cascaded regression with an application to face alignment," in *CVPR*, 2015. 4