Project-Out Cascaded Regression with an application to Face Alignment

Georgios Tzimiropoulos
School of Computer Science
University of Nottingham, U.K.
yorgos.tzimiropoulos@nottingham.ac.uk

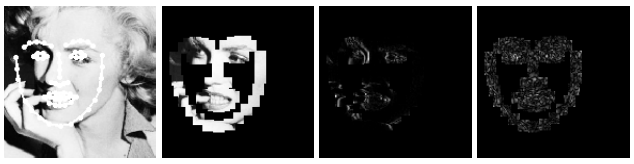
Abstract

Cascaded regression approaches have been recently shown to achieve state-of-the-art performance for many computer vision tasks. Beyond its connection to boosting, cascaded regression has been interpreted as a learning-based approach to iterative optimization methods like the Newton's method. However, in prior work, the connection to optimization theory is limited only in learning a mapping from image features to problem parameters.

In this paper, we consider the problem of facial deformable model fitting using cascaded regression and make the following contributions: (a) We propose regression to learn a sequence of averaged Jacobian and Hessian matrices from data, and from them descent directions in a fashion inspired by Gauss-Newton optimization. (b) We show that the optimization problem in hand has structure and devise a learning strategy for a cascaded regression approach that takes the problem structure into account. By doing so, the proposed method learns and employs a sequence of averaged Jacobians and descent directions in a subspace orthogonal to the facial appearance variation; hence, we call it Project-Out Cascaded Regression (PO-CR). (c) Based on the principles of PO-CR, we built a face alignment system that produces remarkably accurate results on the challenging iBUG data set outperforming previously proposed systems by a large margin. Code for our system is available from `http://www.cs.nott.ac.uk/~yzt/`.

1. Introduction

Regression is a standard tool for approaching various computer vision problems like human and head pose estimation [30, 12], deformable model fitting [7, 37], object localization and tracking [33], and face and behaviour analysis [24] to name a few. Typically, regression-based methods wish to learn a function that maps object appearance to the desired target output variables. Being discriminative in nature and by capitalizing on the very large annotated data sets that are now readily available, they have been shown to produce state-of-the-art performance for many of the aforementioned tasks. At the same time, regression-based methods enjoy a high degree of computational efficiency in both training and testing. In this work, the focus is on regression-based fitting of facial deformable models to unconstrained images, also known as face alignment in-the-wild. Arguably, for this problem, regression-based approaches have recently emerged as the state-of-the-art.



(a) (b) (c) (d)

Figure 1. Project-Out Cascaded Regression vs Gauss-Newton optimization. In prior work in face alignment, given the current estimate of the landmarks' location (a)-(b), image specific Jacobians are calculated to be used in analytic gradient descent. In (c), the image Jacobian with respect to the 3rd shape parameter is shown. In this work, we propose regression to learn a sequence of averaged Jacobians from data, and from them descent directions. In (d), the learned averaged Jacobian with respect to the 3rd shape parameter for the first level of the cascade is shown. Notably, PO-CR learns averaged Jacobians from which facial appearance variation is *projected-out*.

A plethora of regression methods have been employed to tackle the above mentioned problems including linear and ridge [4], Support Vector [31], Boosted [13], Gaussian process [26], and more recently, Deep Neural Nets [18]. A recent notable approach that is of particular interest in this work is the so-called Cascaded Pose Regression (CPR) [11]. CPR is an iterative (cascaded) regression method that is related to boosting with the main difference being that it uses pose-indexed features i.e. features that are sampled from the image based on the current pose estimate. This idea has been shown to produce excellent results on a variety of tasks and, owing to its efficiency and accuracy, it has been recently extensively explored by a number of authors for the problem of face alignment [6, 38, 32, 40, 39, 27, 1, 17].

Regression, as a learning-based solution to optimization, dates back to the seminal work of [7]. More recently, the Supervised Descent Method (SDM) [38] considers the problem of fitting deformable models to facial images using non-linear least squares optimization and derives CPR as a supervised (learning-based) solution to that problem. As we show hereafter, in prior work (including [7] and [38]), (a) the connection to optimization theory is limited only in learning a mapping from image features to problem parameters, (b) there is no attempt to estimate the Jacobian and Hessian matrices (key concepts in optimization) and (c) the structure of the optimization problem in hand is not taken into account.

1.1. Contributions and main results

In this paper, we consider the problem of facial deformable model fitting using cascaded regression and make the following contributions:

- We propose regression to learn a sequence of averaged Jacobian and Hessian matrices from data, and from them descent directions. Our method is inspired by prior work on Gauss-Newton optimization for fitting facial deformable models but rather than calculating image specific Jacobians to be used in analytic gradient descent, we propose cascaded regression to learn a sequence of averaged Jacobians from data, one per iteration.

- We show that the optimization problem in hand has structure and devise a learning strategy for a cascaded regression approach that takes the problem structure into account. In particular, we propose Project-Out Cascaded Regression (PO-CR), a cascaded regression approach for fitting facial deformable models to unconstrained images that learns and employs regressors in a subspace orthogonal to the appearance variation. In particular, the key idea in the proposed learning strategy for PO-CR is to compute a sequence of averaged Jacobians from which facial appearance variation is *projected-out*.

- Based on the principles of PO-CR, we built a face alignment system and tested it on the most popular facial databases, namely LFPW [3], Helen [19], AFW [41] and iBUG [28]. Notably, our system produces remarkably accurate results on the challenging iBUG data set outperforming previously proposed systems by a large margin. Code for our system is available from `http://www.cs.nott.ac.uk/~yzt/`.

We note that there are many examples of computer vision problems including bundle adjustment [34, 20], parameterized model fitting [15, 23] and detection/tracking [14, 16], in which the resulting optimization problems have structure; for example, the underlying normal equations might exhibit a sparse block or circulant structure [5]. Within the proposed formulation, our results show that this structure must be exploited during learning to produce accurate and robust solutions during testing.

1.2. Related work

The proposed Project-Out Cascaded Regression (PO-CR) is a cascaded regression approach and hence the starting point for our work is the CPR of [11]. CPR is an iterative regression method in which the output of regression at iteration $k-1$ is used as input for iteration k, and each regressor uses image features that depend on the current pose estimate. This idea was explored for the problem of face alignment in [6] where the authors demonstrated excellent results on the LFPW data set [3]. The proposed PO-CR is a cascaded regression approach that is derived as a solution to a non-linear least squares optimization problem for fitting generative deformable models to facial images and as such is related to the recently proposed SDM of [38]. Interestingly, the connection between regression and non-linear least squares optimization dates back to the original Active Appearance Model (AAM) formulation of [7]. None of these approaches however proposes to learn a sequence of averaged Jacobian and Hessian matrices from data nor takes into account the problem structure in the formulated optimization problem as suggested by PO-CR. This structure has been occasionally explored by a number of authors in the context of fitting facial deformable models to images using analytic gradient descent (Gauss-Newton optimization) [15, 23, 25, 35], with well-known examples being the Project-Out Inverse Compositional algorithm of [23] and, more recently, the Gauss-Newton generative deformable part model of [36]. Notably, in these methods, the update of the shape parameters at each iteration is found by *projecting-out* the facial appearance variation from the image specific Jacobian. A similar idea is explored for learning in the proposed PO-CR. See also Fig. 1.

2. State-of-the-art in face alignment

The problem of face alignment has a long history in computer vision and a large number of approaches have been proposed to tackle it. Typically, faces are modelled as deformable objects which can vary in terms of shape and appearance. Much of early work revolved around the Active Shape Models (ASMs) and the Active Appearance Models (AAMs) [8, 7, 23]. In ASMs, facial shape is expressed as a linear combination of shape bases learned via Principal Component Analysis (PCA), while appearance is modelled locally using (most commonly) discriminatively learned templates. In AAMs, shape is modelled as in ASMs but appearance is modelled globally using PCA in a canonical coordinate frame where shape variation has

been removed. More recently, the focus has been shifted to the family of methods coined Constrained Local Models (CLMs) [9, 22, 29] which build upon the ASMs. Besides new methodologies, another notable development in the field has been the collection and annotation of large facial data sets captured in unconstrained conditions (in-the-wild) [3, 41, 19, 28]. Being able to capitalize on large amounts of data, a number of (cascaded) regression-based techniques have been recently proposed which achieve impressive performance [37, 6, 38, 32, 27, 1, 17]. The approaches described in [38, 27, 1, 17] along with the part-based generative deformable model of [36] are considered to be the state-of-the-art in face alignment.

3. Project-Out Cascaded Regression

The proposed Project-Out Cascaded Regression (PO-CR) uses generative models of facial shape and appearance fitted via cascaded regression in a subspace orthogonal to the learned appearance variation. In the following sections, we describe (a) the facial shape and appearance models employed by PO-CR (section 3.1), (b) the optimization problem which provides the basis for learning in PO-CR (section 3.2), (c) the learning and fitting process in PO-CR (section 3.3), and finally (d) the differences between PO-CR and related prior work (section 3.4).

3.1. Shape and appearance models

In this section, we describe the shape and appearance models employed by the proposed PO-CR. In particular, we use a parametric global shape model and a parametric part-based appearance model akin to the ones originally proposed in [10] and more recently employed in [36]. A notable difference from recent work on cascaded regression is that we use parametric generative models for shape and appearance both learned via PCA from an annotated training set as explained below. Although recent regression approaches advocate the use of non-parametric shape models [6], the parametric one employed here is more compact having far less number of parameters to optimize. Additionally, learning a generative appearance model is a key idea in PO-CR. In contrast to recently proposed cascaded regression methods, PO-CR learns and employs averaged Jacobians and descent directions in a subspace orthogonal to the learned appearance model.

As in most works in face alignment, we assume a supervised setting where a set of training facial images $\mathbf{I}_i$ are annotated with u fiducial points. For each image, the set of all points is a vector $\in \mathcal{R}^{2u\times 1}$ that is said to define the shape of each face. To learn the shape model used in PO-CR, the annotated shapes are firstly normalized using Procrustes Analysis. This step removes variations due to similarity transformations (translation, rotation and scaling). Then, PCA is applied on the normalized shapes to obtain the shape model. The model is defined by the mean shape $\mathbf{s}_0$ and n shape eigenvectors $\mathbf{s}_i$ compactly represented as columns of matrix $\mathbf{S} \in \mathcal{R}^{2u\times n}$. Finally, to model similarity transforms, $\mathbf{S}$ is appended with 4 additional bases as described in [23]. An instance of the shape model is given by

$$\mathbf{s}(\mathbf{p}) = \mathbf{s}_0 + \mathbf{S}\mathbf{p}, \tag{1}$$

where $\mathbf{p} \in \mathcal{R}^{n\times 1}$ is the vector of the shape parameters.

To learn the appearance model used in PO-CR, each training image $\mathbf{I}_i$ is warped to a reference frame so that similarity transformations are removed. Then, a descriptor (e.g. image patch or SIFT [21]) describing the local appearance around each landmark is computed and all descriptors are stacked in a vector $\in \mathcal{R}^{N\times 1}$ which defines the part-based appearance of $\mathbf{I}_i$. Then, PCA is applied on the part-based representations of all training images to obtain the appearance model. The model is defined by the mean appearance $\mathbf{A}_0$ and m appearance eigenvectors $\mathbf{A}_i$ compactly represented as columns of matrix $\mathbf{A} \in \mathcal{R}^{N\times m}$. An instance of the appearance model is given by

$$\mathbf{A}(\mathbf{c}) = \mathbf{A}_0 + \mathbf{A}\mathbf{c}, \tag{2}$$

where $\mathbf{c} \in \mathcal{R}^{m\times 1}$ is the vector of the appearance parameters.

3.2. Optimization problem for PO-CR

In this section, we formulate and solve the non-linear least squares optimization problem which provides the basis for learning and fitting in PO-CR. Similarly to [38], we will proceed by employing analytic gradient descent [23, 35] which will give rise to Eqs. (7) and (8). Then, in the next section, we will use Eqs. (7) and (8) to devise the learning and fitting process for the proposed PO-CR.

The derived optimization problem below is akin to the one described in [36] with one difference being that here we consider forward rather than inverse fitting algorithms. Note that the fundamental difference between PO-CR and all the aforementioned works (including the method described below) is that PO-CR proposes a *regression-based* solution as opposed to analytic gradient descent.

Let us denote by $\mathbf{I}(\mathbf{s}(\mathbf{p})) \in \mathcal{R}^{N\times 1}$ the vector obtained by generating u landmarks from a shape instance $\mathbf{s}(\mathbf{p})$ and concatenating the computed descriptors for all landmarks. To localize the landmarks in a new image, we would like to find $\mathbf{p}$ and $\mathbf{c}$ such that

$$\arg\min_{\mathbf{p},\mathbf{c}} ||\mathbf{I}(\mathbf{s}(\mathbf{p})) - \mathbf{A}(\mathbf{c})||^2. \tag{3}$$

To find a locally optimal solution to the above problem, we iterate the following procedure: given a current estimate of $\mathbf{p}$ and $\mathbf{c}$ at iteration k, we perform a first-order Taylor approximation in a similar fashion to the Lucas-Kanade algorithm [2]. Then, an update for $\mathbf{p}$ and $\mathbf{c}$ can be found by

solving the following optimization problem

$$\arg\min_{\Delta\mathbf{p},\Delta\mathbf{c}} ||\mathbf{I}(\mathbf{s}(\mathbf{p})) + \mathbf{J}_I\Delta\mathbf{p} - \mathbf{A}_0 - \mathbf{A}\mathbf{c} - \mathbf{A}\Delta\mathbf{c}||^2, \quad (4)$$

where $\mathbf{J}_I \in \mathcal{R}^{N\times n}$ is the image specific Jacobian with respect to the shape parameters.

Let us define $\Delta\mathbf{q} = [\Delta\mathbf{p}; \Delta\mathbf{c}] \in \mathcal{R}^{(n+m)\times 1}$, $\mathbf{J}_q = [\mathbf{J}_I \ -\mathbf{A}] \in \mathcal{R}^{N\times(n+m)}$ and $\mathbf{H}_q = \mathbf{J}_q^T\mathbf{J}_q$. Then, a solution for $\Delta\mathbf{q}$ at iteration k can be found from

$$\Delta\mathbf{q} = -\mathbf{H}_q^{-1}\mathbf{J}_q^T(\mathbf{I}(\mathbf{s}(\mathbf{p})) - \mathbf{A}(\mathbf{c})). \quad (5)$$

As we may observe at each iteration one needs to solve for *both* $\Delta\mathbf{p}$ and $\Delta\mathbf{c}$. Fortunately, there is an alternative way that by-passes the computation for the optimal $\Delta\mathbf{c}$ at each iteration and guarantees an exact update for $\Delta\mathbf{p}$ by taking into account the problem structure. This structure can be readily seen by writing

$$\mathbf{H}_q = \begin{bmatrix}\mathbf{H}_{pp} & \mathbf{H}_{pc}\\ \mathbf{H}_{cp} & \mathbf{H}_{cc}\end{bmatrix} = \begin{bmatrix}\mathbf{J}_I^T\mathbf{J}_I & -\mathbf{J}_I^T\mathbf{A}\\ -\mathbf{A}^T\mathbf{J}_I & \mathbf{E}_m\end{bmatrix},$$

where $\mathbf{E}_m = \mathbf{A}^T\mathbf{A}$ is the $m \times m$ identity matrix.

To take advantage of the problem structure, we *firstly* optimize the problem of Eq. (4) with respect to $\Delta\mathbf{c}$. The optimal $\Delta\mathbf{c}$ is readily given by

$$\Delta\mathbf{c} = \mathbf{A}^T(\mathbf{I}(\mathbf{s}(\mathbf{p})) + \mathbf{J}_I\Delta\mathbf{p} - \mathbf{A}(\mathbf{c})), \quad (6)$$

which as we may observe is a function of $\Delta\mathbf{p}$. Then, we plug in the solution back to Eq. (4) [5, 15, 35]. By doing so, we end up with the following optimization problem

$$\arg\min_{\Delta\mathbf{p}} ||\mathbf{I}(\mathbf{s}(\mathbf{p})) + \mathbf{J}_I\Delta\mathbf{p} - \mathbf{A}_0||^2_{\mathbf{P}}, \quad (7)$$

where we have used the notation $||\mathbf{x}||^2_{\mathbf{W}} = \mathbf{x}^T\mathbf{W}\mathbf{x}$ to denote the weighted ℓ_2-norm of a vector $\mathbf{x}$. The solution to the above problem is readily given by [1]

$$\Delta\mathbf{p} = -\mathbf{H}_P^{-1}\mathbf{J}_P^T(\mathbf{I}(\mathbf{s}(\mathbf{p})) - \mathbf{A}_0), \quad (8)$$

where $\mathbf{J}_P = \mathbf{P}\mathbf{J}_I$ and $\mathbf{H}_P = \mathbf{J}_P^T\mathbf{J}_P$, $\mathbf{P} = \mathbf{E} - \mathbf{A}\mathbf{A}^T$ is a projection operator that *projects out the facial appearance variation* from the image Jacobian $\mathbf{J}_I$, and $\mathbf{E}$ is the identity matrix. Note that the Jacobian, the Hessian and its inverse need to be re-computed per iteration giving rise to an algorithm with complexity $O(nmN + n^2N)$ per iteration.

To summarize, we have derived Eqs. (7) and (8) from an analytic gradient descent perspective. In the next section, we will describe the learning and fitting process for the proposed PO-CR as a *regression-based* solution to Eqs. (7) and (8).

[1] Alternatively, we could use Schur's complement to derive $\Delta\mathbf{p}$, but this way does not allow us to derive (7) which is used in PO-CR for learning averaged Jacobians from data. See also section 3.3.

3.3. Learning and fitting in PO-CR

Learning in PO-CR is based on Eqs. (7) and (8). In particular, as we may observe from Eq. (8), at each iteration calculating $\Delta\mathbf{p}$ requires (a) computing the image Jacobian, (b) projecting-out the facial appearance variation from it and (c) computing the Hessian and its inverse. Based on this procedure, we propose to adopt a similar idea for our learning strategy in PO-CR.

In particular, for notational clarity let us first make the dependency of variables on iteration k explicit. Then, the key idea in PO-CR is to compute from a set of training examples an averaged Jacobian $\widehat{\mathbf{J}}(k)$ from which the facial appearance variation is *projected-out*. The averaged projected-out Jacobian, denoted as $\widehat{\mathbf{J}}_P(k)$, is then used to compute an averaged projected-out Hessian and descent directions. In detail, our learning strategy for PO-CR is as follows:

Step I. Starting from the ground truth shape parameters $\mathbf{p}_i^*$ for each training image $\mathbf{I}_i$, $i = 1, \ldots, H$, we generate a set of K perturbed shape parameters for iteration 1 $\mathbf{p}_{i,j}(1)$, $j = 1, \ldots, K$ that capture the statistics of the face detection initialization process. Using the set $\Delta\mathbf{p}_{i,j}(1) = \mathbf{p}_i^* - \mathbf{p}_{i,j}(1)$, PO-CR learns the averaged projected-out Jacobian $\widehat{\mathbf{J}}_P(1) = \mathbf{P}\widehat{\mathbf{J}}(1)$ for iteration 1 by solving the following weighted least squares problem

$$\arg\min_{\widehat{\mathbf{J}}_P(1)} \sum_{i=1}^{H}\sum_{j=1}^{K} ||\mathbf{I}(\mathbf{s}(\mathbf{p}_{i,j}(1))) + \mathbf{J}(1)\Delta\mathbf{p}_{i,j}(1) - \mathbf{A}_0||^2_{\mathbf{P}}, \quad (9)$$

where the solution for $\widehat{\mathbf{J}}_P(1)$ is obtained using ridge-regression [2]. Notice that the above optimization problem is formulated in $\mathbf{P}$. As our experiments have shown working in this subspace is necessary for achieving good performance. See also section 4.

Step II. Having computed $\widehat{\mathbf{J}}_P(1)$, we further compute the averaged projected-out Hessian $\widehat{\mathbf{H}}_P(1) = \widehat{\mathbf{J}}_P(1)^T\widehat{\mathbf{J}}_P(1)$ and its inverse.

Step III. Given $\widehat{\mathbf{J}}_P(1)$ and $\widehat{\mathbf{H}}_P(1)^{-1}$, the descent directions $\mathbf{R}(1) \in \mathcal{R}^{n\times N}$ for iteration 1 are given by

$$\mathbf{R}(1) = \widehat{\mathbf{H}}_P(1)^{-1}\widehat{\mathbf{J}}_P(1)^T. \quad (10)$$

Step IV. For each training sample, a new estimate for its shape parameters (to be used at the next iteration) is obtained from

$$\mathbf{p}_{i,j}(2) = \mathbf{p}_{i,j}(1) + \mathbf{R}(1)(\mathbf{I}(\mathbf{s}(\mathbf{p}_{i,j}(1))) - \mathbf{A}_0). \quad (11)$$

Finally, Steps I-IV are sequentially repeated until convergence and the whole process produces a set of L regressor matrices $\mathbf{R}(l)$, $l = 1, \ldots, L$.

[2] Notice that by simple mathematical manipulation, the ℓ_2-norm in Eq. (9) becomes a function of $\widehat{\mathbf{J}}_P(1)$.

During testing, and in a similar fashion to cascaded regression techniques, given a current estimate of the shape parameters at iteration k, $\mathbf{p}(k)$, we extract image features $\mathbf{I}(\mathbf{s}(\mathbf{p}(k)))$ and then compute an update for the shape parameters from

$$\Delta\mathbf{p}(k) = \mathbf{R}(k)(\mathbf{I}(\mathbf{s}(\mathbf{p}(k))) - \mathbf{A}_0). \tag{12}$$

Finally, after L iterations we obtain the fitted shape. Notice that the complexity per iteration is $O(nN)$ only, and hence at testing time PO-CR maintains the high degree of computational efficiency typically characterizing cascaded regression techniques. We note that optimized implementations of such methods have been shown to operate in tens of frames per second (e.g. [38, 1]).

3.4. Comparison with prior work

In this section, we highlight similarities and differences between the proposed PO-CR and related prior work in analytic gradient descent and cascaded regression.

Against AAMs. The proposed project-out formulation is reminiscent of the well-known Project-Out Inverse Compositional (PO-IC) algorithm used in AAM fitting [23]. Both algorithms work in a subspace orthogonal to the appearance variation and have the same computational complexity per iteration ($O(nN)$). However, PO-IC precomputes and employs an image Jacobian from the mean appearance $\mathbf{A}_0$ which remains fixed in all iterations. In contrast, PO-CR proposes Eq. (9) and regression to precompute a sequence of averaged Jacobians from data, one per iteration. PO-IC is an approximate algorithm for solving the problem of Eq. (4) [35]. In contrast, PO-CR uses Eqs. (7) and (8) as a basis for regression, i.e. the exact method for solving the problem of Eq. (4).

Against SDM. Both PO-CR and SDM learn a sequence of regression matrices (one per iteration) and during fitting the update of the shape parameters is computed in a very similar fashion. Both methods have similar computational complexity. However, SDM uses non-parametric shape and appearance models as opposed to the parametric ones employed by PO-CR. More importantly, learning in PO-CR and SDM is very different. SDM learns directly a mapping from image features to problem parameters. In contrast, PO-CR learns a set of averaged Jacobian and Hessian matrices from data, and from them descent directions in a subspace orthogonal to the appearance variation.

4. Experiments

4.1. Performance evaluation

In this section, we evaluate the performance of PO-CR for the problem of face alignment in-the-wild. To this end, we conducted a large number of experiments on the most popular facial databases, namely LFPW [3], Helen [19], AFW [41] and iBUG [28]. We compare the performance of PO-CR with that of a variant of our method as well as with that of two publicly available systems.

In-house. As in [36], we used the SIFT implementation of [38]. For training, we used the training sets of LFPW and Helen and the available landmark annotations of the 300-W challenge [28]. In addition to PO-CR, we implemented a version of our method in which the project-out component was intentionally omitted. This version simply replaces the *projected-out* $\widehat{\mathbf{J}}_P(k)$ with $\widehat{\mathbf{J}}(k)$, i.e. the solution to Eq. (9) but after dropping the projection operator $\mathbf{P}$. We simply denote this method as "No projection". We included this version in order to illustrate the importance of working in a subspace orthogonal to the learned appearance variation.

Publicly available systems. We compared the performance of PO-CR with that of two publicly available systems: SDM [38] and Chehra [1]. SDM was trained on internal CMU data that are not publicly available, and Chehra on the whole LFPW, Helen, AFW and iBUG data sets including data that is not publicly available. We note that the training data for Chehra included the test sets of LFPW, Helen, AFW and iBUG on which we report performance below, and hence Chehra has an inherent advantage over all other methods.

For initialization, we used the ground truth points to compute the ground truth bounding box for each image (rotation angle was removed). This bounding box was then scaled and translated according to a noise distribution, defined by standard deviation σ. In this way, we could identify the range of initializations that SDM [38] and Chehra [1] can handle. We used a noise level of $\sigma_{\text{noise}} = 3.5$ for which both methods performed very well on LFPW. We found that Chehra works satisfactorily for noise level up to $\sigma_{\text{noise}} = 5$. For the same noise level (i.e. $\sigma_{\text{noise}} = 5$) our systems operates with literally no loss in performance. To measure performance, we used the point-to-point (pt-pt) error normalized by the face size defined in [41]. We report the cumulative curve corresponding to the percentage of images for which the error was less than a specific value. To facilitate comparison with [38] and [1], we report performance on the 49 interior points.

Fig. 2 shows our results on LFPW, Helen, AFW and iBUG. From these results, we can draw a number of interesting conclusions: (a) LFPW and Helen are the "easiest to fit" data sets, followed by AFW and iBUG. It seems that iBUG is by far the most challenging data set. (b) By removing the project-out component from our method ("No projection"- cyan), fitting performance drops dramatically. In fact, this method performs the worst compared to all other methods. This shows the importance of the proposed project-out formulation. (c) Our system consistently produces the most accurate results on all data sets. (d) Our system is the most robust among all other methods producing

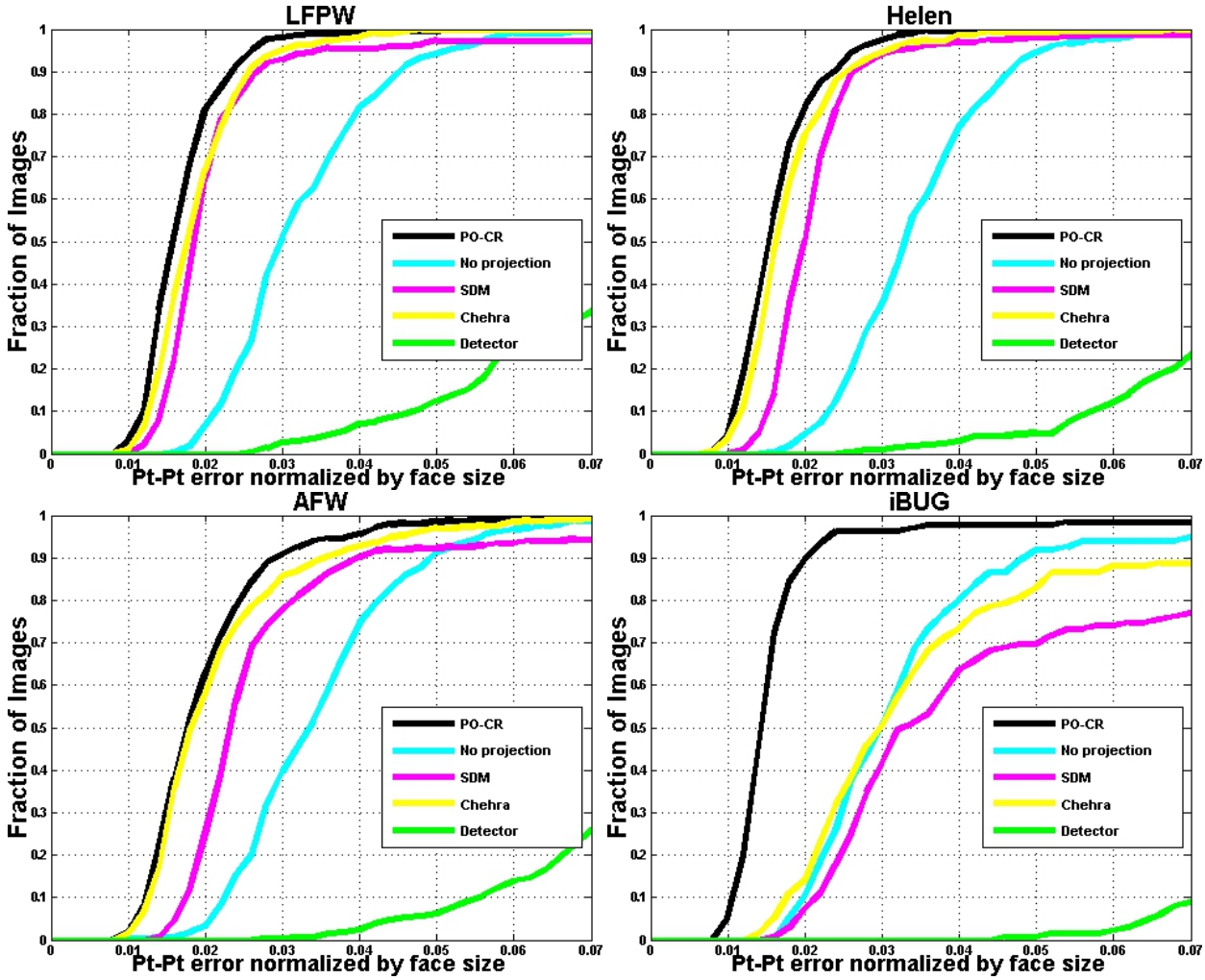


Figure 2. Average pt-pt Euclidean error (normalized by the face size) vs fraction of images for LFPW, Helen, AFW and iBUG. We compare the performance of Project-Out Cascaded Regression (black), our approach without projecting-out (cyan), SDM [38] (magenta) and Chehra [1] (yellow). The average error is computed over 49 points.

literally the same fitting accuracy on all data sets, including iBUG.

4.2. Fitting results from the iBUG data set

As the iBUG data set is the most challenging among all data sets but contains only 135 images, in Figs. 3 and 4, we present the fittings produced by PO-CR for all 135 images of this data set as well as the bounding box initializations used (produced by noise level $\sigma_{\text{noise}} = 5$). As it can be observed, our system is able to fit images with very large shape and appearance variation even for the case of challenging initializations.

5. Conclusions

We proposed Project-Out Cascaded Regression, a cascaded regression approach derived from a Gauss-Newton solution to a non-linear least squares problem that has structure. The learning strategy in PO-CR capitalizes on this structure to compute averaged Jacobians from which the facial appearance variation is projected-out and then employs the projected-out Jacobians to compute descent directions. The fitting process in PO-CR is similar to that of other cascaded regression techniques and hence our method maintains a high degree of computational efficiency. We conducted a large number of experiments on the most popular facial databases, namely LFPW, Helen, AFW and iBUG that show that our system outperforms state-of-the-art systems sometimes by a large margin.

Figure 3. Application of Project-Out Cascaded Regression to the alignment of the iBUG data set. For each image, the black bounding box shows the face detection initialization. Our algorithm is able to produce highly accurate fittings for images with very large shape and appearance variation even with challenging initializations. The first 70 images of the iBUG data set are shown.

Figure 4. Application of Project-Out Cascaded Regression to the alignment of the iBUG data set. The fittings and the initializations for the remaining 65 images of the iBUG data set are shown.

References

[1] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic. Incremental face alignment in the wild. In *CVPR*, 2014.

[2] S. Baker and I. Matthews. Lucas-kanade 20 years on: A unifying framework. *IJCV*, 56(3):221–255, 2004.

[3] P. Belhumeur, D. Jacobs, D. Kriegman, and N. Kumar. Localizing parts of faces using a consensus of exemplars. In *CVPR*, 2011.

[4] C. M. Bishop. *Pattern recognition and machine learning*. springer New York, 2006.

[5] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[6] X. Cao, Y. Wei, F. Wen, and J. Sun. Face alignment by explicit shape regression. In *CVPR*, 2012.

[7] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *TPAMI*, 23(6):681–685, 2001.

[8] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models-their training and application. *CVIU*, 61(1):38–59, 1995.

[9] D. Cristinacce and T. Cootes. Feature detection and tracking with constrained local models. In *BMVC*, 2006.

[10] D. Cristinacce and T. Cootes. Automatic feature localisation with constrained local models. *Pattern Recognition*, 41(10):3054–3067, 2008.

[11] P. Dollár, P. Welinder, and P. Perona. Cascaded pose regression. In *CVPR*, 2010.

[12] G. Fanelli, J. Gall, and L. Van Gool. Real time head pose estimation with random regression forests. In *CVPR*, 2011.

[13] J. H. Friedman. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, pages 1189–1232, 2001.

[14] H. K. Galoogahi, T. Sim, and S. Lucey. Multi-channel correlation filters. In *ICCV*, 2013.

[15] G. D. Hager and P. N. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *IEEE TPAMI*, 20(10):1025–1039, 1998.

[16] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista. Exploiting the circulant structure of tracking-by-detection with kernels. In *ECCV*, 2012.

[17] V. Kazemi and S. Josephine. One millisecond face alignment with an ensemble of regression trees. In *CVPR*, 2014.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012.

[19] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang. Interactive facial feature localization. In *ECCV*, 2012.

[20] M. A. Lourakis and A. Argyros. Sba: A software package for generic sparse bundle adjustment. *ACM TOMS*, 36(1):1–30, 2009.

[21] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[22] S. Lucey, Y. Wang, M. Cox, S. Sridharan, and J. Cohn. Efficient constrained local model fitting for non-rigid face alignment. *Image and Vision Computing*, 27(12):1804–1813, 2009.

[23] I. Matthews and S. Baker. Active appearance models revisited. *IJCV*, 60(2):135–164, 2004.

[24] M. A. Nicolaou, H. Gunes, and M. Pantic. Output-associative rvm regression for dimensional and continuous emotion prediction. *IVCJ*, 30(3):186–196, 2012.

[25] G. Papandreou and P. Maragos. Adaptive and constrained algorithms for inverse compositional active appearance model fitting. In *CVPR 2008*, 2008.

[26] C. E. Rasmussen. *Gaussian processes for machine learning*. MIT Press, 2006.

[27] S. Ren, X. Cao, Y. Wei, and J. Sun. Face alignment at 3000 fps via regressing local binary features. In *CVPR*, 2014.

[28] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic. A semi-automatic methodology for facial landmark annotation. In *CVPR-W*, 2013.

[29] J. Saragih, S. Lucey, and J. Cohn. Deformable model fitting by regularized landmark mean-shift. *IJCV*, 91(2):200–215, 2011.

[30] J. Shotton, R. Girshick, A. Fitzgibbon, T. Sharp, M. Cook, M. Finocchio, R. Moore, P. Kohli, A. Criminisi, A. Kipman, et al. Efficient human pose estimation from single depth images. *IEEE TPAMI*, 35(12):2821–2840, 2013.

[31] A. J. Smola and B. Schölkopf. A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222, 2004.

[32] Y. Sun, X. Wang, and X. Tang. Deep convolutional network cascade for facial point detection. In *CVPR*, 2013.

[33] D. J. Tan, S. Holzer, N. Navab, and S. Ilic. Deformable template tracking in 1ms. In *BMVC*, 2014.

[34] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon. Bundle adjustment: a modern synthesis. In *Vision algorithms: theory and practice*, pages 298–372. Springer, 2000.

[35] G. Tzimiropoulos and M. Pantic. Optimization problems for fast aam fitting in-the-wild. In *ICCV*, 2013.

[36] G. Tzimiropoulos and M. Pantic. Gauss-newton deformable part models for face alignment in-the-wild. In *CVPR*, 2014.

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

[38] X. Xiong and F. De la Torre. Supervised descent method and its applications to face alignment. In *CVPR*, 2013.

[39] H. Yang and I. Patras. Face parts localization using structured-output regression forests. In *ACCV 2012*. 2013.

[40] H. Yang and I. Patras. Sieving regression forest votes for facial feature detection in the wild. In *ICCV*, 2013.

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