Robust Statistical Face Frontalization

Christos Sagonas* Yannis Panagakis* Stefanos Zafeiriou* Maja Pantic*†

*Department of Computing,
Imperial College London,
180 Queen's Gate,
London SW7 2AZ, U.K.

†EEMCS,
University of Twente,
Drienerlolaan 5,
7522 NB Enschede, The Netherlands

{c.sagonas, i.panagakis, s.zafeiriou, m.pantic}@imperial.ac.uk

Abstract

Recently, it has been shown that excellent results can be achieved in both facial landmark localization and pose-invariant face recognition. These breakthroughs are attributed to the efforts of the community to manually annotate facial images in many different poses and to collect 3D facial data. In this paper, we propose a novel method for joint frontal view reconstruction and landmark localization using a small set of ***frontal images only****. By observing that the frontal facial image is the one having the minimum rank of all different poses, an appropriate model which is able to jointly recover the frontalized version of the face as well as the facial landmarks is devised. To this end, a suitable optimization problem, involving the minimization of the nuclear norm and the matrix ℓ_1 norm is solved. The proposed method is assessed in frontal face reconstruction, face landmark localization, pose-invariant face recognition, and face verification in unconstrained conditions. The relevant experiments have been conducted on 8 databases. The experimental results demonstrate the effectiveness of the proposed method in comparison to the state-of-the-art methods for the target problems.*

1. Introduction

Face frontalization refers to the recovery of the frontal view of faces from single unconstrained images. Accurate face frontalization is a cornerstone for many face analysis problems. For example the recent success in face recognition in unconstrained conditions would not be possible without a meticulously designed face frontalization procedure [35].

An essential step towards face frontalization is facial landmark localization. State-of-the-art landmark localization methods [5, 17, 28, 32, 36, 40] model the problem discriminatively by capitalizing on the availability of annotated data in terms of facial landmarks [30, 31]. Unfortunately, the annotation of facial landmarks is laborious, expensive, and time consuming process. This is even more the case for faces that are not in frontal pose[1].

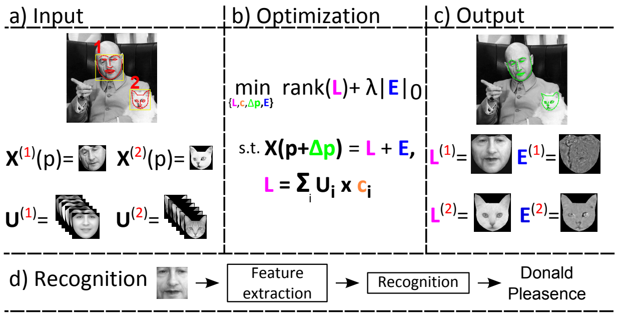


Figure 1. Flowchart of the proposed method: a) Given an input image, the results from a detector, and a statistical model $\mathbf{U}$, built on frontal images only, b) a constrained low-rank minimization problem is solved. c) Face alignment and frontal view reconstruction are performed simultaneously. Finally, d) face recognition is performed using the frontalized image.

In many cases, accurate 2D landmark localization is not enough for successful face frontalization. That is, in practice, the frontalization step is very elaborate requiring both landmark localization and pose correction by usually resorting to 3D face models [33–35, 41]. In general, 3D model-based methods achieve high recognition accuracy [35, 41]. However, such methods cannot be widely applied since they require: (a) a method for accurate landmark localization in various poses, (b) fitting learned 3D model of faces, which is expensive to built, and (c) a robust image warping algorithm for frontal view image reconstruction [35]. Recently, a simple method for the reconstruction of frontal views using a 3D reference mesh has been proposed in literature [12]. The main difference between the frontalization system in [35] and the one proposed in [12] is that the latter

[1] From experience we know that annotation of facial image with poses take in many cases twice the time compared with frontal poses.

method uses the same 3D reference mesh as the approximation of different subject's face shape. The main drawback of this method is that it relies on the perfect localization of facial landmarks. In addition, the frontalized view is affected by the existence of noise in the non-frontal view. An approach that does not require a 3D model but only a small set of landmarks is presented in [13]. This method aims to recover the frontal view of a non-frontal image by employing Markov Random Field (MRF). The main drawback of the latter is that for each non-frontal image, an exhaustively batch-based alignment algorithm is applied (trained on frontal patches). Clearly, such a procedure is time consuming.

In this paper, we propose a simple yet extremely powerful statistical frontalization of faces. The key motivational observation is that, for the facial images lying in a linear space, the rank of a frontal facial image, due to the approximately structure of human face, is much smaller than the rank of facial images in other poses. To demonstrate the above observation 'Neutral' images of twenty objects from Multi-PIE database [11] under poses -30° to 30° were warped into a reference frontal-pose frame and the nuclear norm (convex surrogate of the rank) of each shape-free texture was computed. In Table 1 the average value of the nuclear norm for different poses is reported. Clearly, the frontal pose has the smallest nuclear norm value compared to the corresponding values computed for other poses. However, severe deviations from the above linear facial model occur in the presence of pose, occlusions, expressions, and illumination changes.

To remedy the aforementioned challenges, we propose a unified method for joint face frontalization (pose correction), landmark localization, and pose-invariant face recognition, using a small set of **frontal images only**. In particular, we show that if: (a) deformations due to pose and expressions are approximately removed, (b) occlusion/specular highlights and warping errors due to pose are modelled as sparse errors and, (c) illumination is modelled by using *in-the-wild* training facial images, then the linear space assumption is valid. Inspired by recent advances in learning using low-rank and sparsity e.g., [22,24,25,29,38], a suitable optimization problem, involving the minimization of the nuclear norm and the matrix ℓ_1 norm is solved to achieve the above mentioned goals. The flowchart of the proposed method (coined as RSF–*Robust Statistical Face Frontalization*) is depicted in Fig. 1.

Pose	-30°	-15°	$\mathbf{0^\circ}$	15°	30°
Average value of nuclear norm	0.72	0.71	**0.65**	0.70	0.73

Table 1. The average value of nuclear norm computed based on neutral images of twenty subjects from Multi-PIE database under poses $-30^0 : 30^0$.

The most closely related work to the proposed method is the Transform Invariant Low-rank Textures (TILT) method [47]. In TILT, texture rectification is obtained by applying a global affine transformation onto a low-rank term, modelling the texture. By blindly imposing low-rank constraints without regularization, for non-rigid alignment opposite effects may occur. Recently, it was demonstrated [9,29], that non-rigid deformable models cannot be straightforward combined with optimization problems [25] that involve low-rank terms without a proper regularization. To overcome this and ensure that unnatural faces will be not created, a model of frontal images is employed in this work. In that sense, our method can be seen as a deformable TILT model regularized within a frontal face subspace.

Summarizing, the contributions of the paper can be summarised as follows:

- Technical contributions:
  1. A novel RSF method for joint landmark localization and face frontalization is proposed that uses a statistical model of frontal images, low-rank, and sparsity in order to adequately model pose, occlusions, expressions, and illumination variations.
  2. An effective algorithm for the RSF is developed.
- Applications in computer vision:
  1. To the best of our knowledge this is the first generic landmark localization method which achieves state-of-the-art results using a model of **frontal images only**.
  2. It is possible to improve the state-of-the-art in pose-invariant face recognition and unconstrained face verification using only frontal faces and simple features for classification unlike other complex feature extraction procedures e.g., [33, 35].[2]
  3. Furthermore, we demonstrate the performance of RSF in handling all human faces, cat faces, and face sketches.

The most important and surprising contribution of our paper is that we show that when phenomena are properly modelled simple statistical linear models, even pixel intensities, could produce state-of-the-art results.

Notations. Throughout the paper, scalars are denoted by lower-case letters, vectors (matrices) are denoted by lower-case (upper-case) boldface letters i.e., $\mathbf{x}$, $(\mathbf{X})$. $\mathbf{I}$ denotes the identity matrix. The ith column of $\mathbf{X}$ is denoted by

[2] We note that we refer to the restricted protocol of the LFW [15] and not to the unrestricted which unfortunately we cannot compete since we do not have access to millions of annotated faces.

$\mathbf{x}_i$. A vector $\mathbf{x} \in \mathbb{R}^{m\cdot n}$ (matrix $\mathbf{X} \in \mathbb{R}^{m\times n}$) is reshaped into a matrix (vector) via the reshape operator : $\mathcal{R}(\mathbf{x}) = \mathbf{X} \in \mathbb{R}^{m\times n}$, $\big(\text{vec}(\mathbf{X}) = \mathbf{x} \in \mathbb{R}^{m\cdot n\times 1}\big)$. The ℓ_1 and the ℓ_2 norms of $\mathbf{x}$ are defined as $\|\mathbf{x}\|_1 = \sum_i |x_i|$ and $\|\mathbf{x}\|_2 = \sqrt{\sum_i x_i^2}$, respectively. The matrix ℓ_1 norm is defined as $\|\mathbf{X}\|_1 = \sum_i \sum_j |x_{ij}|$, where $|\cdot|$ denotes the absolute value operator. The Frobenius norm is defined as $\|\mathbf{X}\|_F = \sqrt{\sum_i \sum_j x_{ij}^2}$, and the nuclear norm of $\mathbf{X}$ (i.e., the sum of singular values of a matrix) is denoted by $\|\mathbf{X}\|_*$. Given a Point Distribution Model (PDM) [10], denoted as $\mathcal{S} = \{\bar{\mathbf{s}}, \mathbf{U}_S \in \mathbb{R}^{2N\times N_S}\}$, a new shape instance is generated as $\mathbf{s} = \bar{\mathbf{s}} + \mathbf{U}_S\mathbf{p}$, where $\mathbf{p} \in \mathbb{R}^{N_S\times 1}$ is the vector of shape parameters. The warp function $\mathbf{x}(\mathcal{W}(\mathbf{z};\mathbf{p}))$ $\big(\mathbf{X}(\mathcal{W}(\mathbf{z};\mathbf{p}))\big)$ denotes the warping of each 2D point $\mathbf{z} = [x, y]$ within a shape instance to its corresponding location in a reference (common) frame. To simplify the notation $\mathbf{x}(\mathbf{p})$ $\big(\mathbf{X}(\mathbf{p})\big)$ will be used throughout the paper instead of $\mathbf{x}(\mathcal{W}(\mathbf{z};\mathbf{p}))$ $\big(\mathbf{X}(\mathcal{W}(\mathbf{z};\mathbf{p}))\big)$.

2. Robust Face Frontalization

Let $\mathbf{X} \in \mathbb{R}^{h\times r}$ be an image depicting a non-frontal view of a face and $\mathbf{s} = [x_1, y_1, \cdots, x_N, y_N]^T$ an initial estimation of N landmark points, describing the shape. To create a shape-free texture, the input image is warped into a common frame by employing a warp function $\mathcal{W}(\cdot)$. In many cases, the warped image $\mathbf{X}(\mathbf{p}) \in \mathbb{R}^{m\times n}$ can be corrupted by sparse errors of large magnitude. Such sparse errors indicate that only a small fraction of the image pixels may be corrupted by non-Gaussian noise and occlusions. In this paper, the goal is to recover the clean low-rank frontal view (i.e., $\mathbf{L} \in \mathbb{R}^{m\times n}$) of the $\mathbf{X}(\mathbf{p})$ and the parameters $\mathbf{p}$ such that $\mathbf{X}(\mathbf{p}) = \mathbf{L} + \mathbf{E}$, where $\mathbf{E} \in \mathbb{R}^{m\times n}$ is a sparse matrix, accounting for gross errors. In particular, based on the observation that the frontal view of the face lies onto a low-rank subspace (please refer to Table 1), it can be expressed as a linear combination of a small number of precomputed orthonormal basis (i.e. $\mathbf{U} = [\mathbf{u}_1|\mathbf{u}_2|\ldots|\cdots|\mathbf{u}_k] \in \mathbb{R}^{m\cdot n\times k}$, $\mathbf{U}^T\mathbf{U} = \mathbf{I}$) that span a generic (clean) frontal view subspace, that is $\mathbf{L} = \sum_{i=1}^k \mathcal{R}(\mathbf{u}_i)c_i$. Therefore, the deformed corrupted input image can be expressed as: $\mathbf{X}(\mathbf{p}) = \mathbf{L} + \mathbf{E} = \sum_{i=1}^k \mathcal{R}(\mathbf{u}_i)c_i + \mathbf{E}$.

A natural estimator accounting for the low-rank of the frontal image and the sparsity of the error matrix is to minimize the rank of $\mathbf{L}$ and the number of non-zero entries of the $\mathbf{E}$ measured by the ℓ_0 quasi norm e.g., [8] by demanding $\mathbf{X}(\mathbf{p}) = \mathbf{L} + \mathbf{E}$. Unfortunately both rank$(\cdot)$ and ℓ_0 norm are non-convex, discrete valued functions, minimization of which is NP-hard. Furthermore, the constraint $\mathbf{X}(\mathbf{p}) = \mathbf{L} + \mathbf{E}$ is non-linear. To alleviate this problem, the nuclear- and the ℓ_1- norms are adopted as surrogates to rank and ℓ_0- norm [8]. To address the non-linearity of the above mentioned equality constraint, a first order Taylor linear approximation is applied on the vectorized form of the constraint: $\mathbf{x}(\mathbf{p} + \Delta\mathbf{p}) \approx \mathbf{x}(\mathbf{p}) + \mathbf{J}(\mathbf{p})\Delta\mathbf{p}$. where $\text{vec}(\mathbf{X}(\mathbf{p})) = \text{vec}(\mathbf{L} + \mathbf{E}) = \mathbf{U}\mathbf{c} + \mathbf{e} = \mathbf{x}(\mathbf{p})$ and $\mathbf{J}(\mathbf{p}) = \nabla\mathbf{x}(\mathbf{p})\frac{\partial W}{\partial \mathbf{p}}$ is the Jacobian matrix with the steepest descent images as its columns. Consequently, the RSF solves the following optimization problem:

$$\underset{\mathbf{L},\mathbf{e},\mathbf{c},\Delta\mathbf{p}}{\text{argmin}} \|\mathbf{L}\|_* + \lambda\|\mathbf{E}\|_1$$
$$\text{s.t.} \begin{cases} H^{(1)}(\Delta\mathbf{p}, \mathbf{c}, \mathbf{e}) = \mathbf{x}(\mathbf{p}) + \mathbf{J}(\mathbf{p})\Delta\mathbf{p} - \mathbf{U}\mathbf{c} - \mathbf{e} = 0 \\ H^{(2)}(\mathbf{L}, \mathbf{c}) = \mathbf{L} - \sum_{i=1}^k \mathcal{R}(\mathbf{u}_i)c_i = 0 \end{cases} \tag{1}$$

where λ is a positive weighting parameter that balances the norms. The set of (primal) variables is defined as $\mathcal{V} = \{\mathbf{L}, \mathbf{c}, \Delta\mathbf{p}, \mathbf{e}\}$.

2.1. Optimization

To solve (1), the *augmented* Lagrangian is introduced:

$$\mathcal{L}(\mathcal{V}, \mathcal{M}) = \|\mathbf{L}\|_* + \lambda\|\mathbf{e}\|_1 + \frac{\mu}{2}\big\|H^{(1)}(\Delta\mathbf{p}, \mathbf{c}, \mathbf{e}) + \frac{\mathbf{a}}{\mu}\big\|_2^2 + \frac{\mu}{2}\big\|H^{(2)}(\mathbf{L}, \mathbf{c}) + \frac{\mathbf{B}}{\mu}\big\|_F^2 - \frac{1}{2\mu}\left(\|\mathbf{a}\|_2^2 + \|\mathbf{B}\|_F^2\right), \tag{2}$$

where $\mathcal{M} = \{\mathbf{a} \in \mathbb{R}^{m\cdot n}, \mathbf{B} \in \mathbb{R}^{m\times n}\}$ is the set of Lagrange multipliers for the equality constraints in (1) and $\mu > 0$ is a penalty parameter. By employing the alternating directions method of multipliers (ADMM), (1) is solved by minimizing (2) with respect to each variable in an alternating fashion. Finally, the Lagrange multipliers are updated at each iteration.

Let t be the iteration index. For notation convenience (2) will be denoted as $\mathcal{L}(\mathcal{V}_{[t]}^{(i)}, \mathcal{M}_{[t]})$ when all the variables expect $\mathcal{V}^{(i)}$ are kept fixed. Accordingly, given $\{\mathcal{V}_{[t]}^{(i)}\}_{i=1}^4$, $\mathcal{M}_{[t]}$ and μ the updates of the primal variables are computed by solving the following sub-problems:

Step 1. Update L:

$$\mathcal{V}_{[t+1]}^{(1)} = \underset{\mathcal{V}^{(1)}}{\text{argmin}} \|\mathbf{L}\|_* + \frac{\mu}{2}\big\|H^{(2)}(\mathbf{L}, \mathbf{c}) + \frac{\mathbf{B}}{\mu}\big\|_F^2. \tag{3}$$

The nuclear norm regularized least squared problem (3) has the following closed-form solution:

$$\mathcal{V}_{[t+1]}^{(1)} = \mathcal{D}_{\frac{1}{\mu_{[t]}}}\left[\sum_{i=1}^k \mathcal{R}(\mathbf{u}_i)c_{i,[t]} - \frac{\mathbf{B}_{[t]}}{\mu_{[t]}}\right]. \tag{4}$$

The singular value thresholding (SVT) operator is defined for any matrix $\mathbf{Q}$ with $\mathbf{Q} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ as $\mathcal{D}_\tau[\mathbf{Q}] = \mathbf{U}S_\tau\mathbf{V}^T$ [7], with $S_\tau[\sigma]$ =sgn$(\sigma)\max(|\sigma| - \tau, 0)$ being the (element-wise) shrinkage operator [8].

Step 2. Update c:

$$\mathcal{V}_{[t+1]}^{(2)} = \underset{\mathcal{V}^{(2)}}{\text{argmin}} \frac{\mu}{2}\big(\big\|H^{(1)}(\Delta\mathbf{p}, \mathbf{c}, \mathbf{e}) + \frac{\mathbf{a}}{\mu}\big\|_2^2 + \big\|H^{(2)}(\mathbf{L}, \mathbf{c}) + \frac{\mathbf{B}}{\mu}\big\|_F^2\big). \tag{5}$$

Algorithm 1: ADMM solver

```
Data: Test image $\mathbf{X}$, initial shape parameters $\mathbf{p}_{in}$, clean
      frontal-view face subspace $\mathbf{U}$, and the parameter $\lambda$
Result: The low-rank clean image $\mathbf{L}$, the sparse error $\mathbf{e}$, the
        coefficient vector $\mathbf{c}$, and the shape parameters $\mathbf{p}$.
while not converged do
    $\mathbf{X}(\mathbf{p}) \leftarrow$ Warp and normalize the image;
    $\mathbf{J} \leftarrow$ Compute the Jacobian matrix;
    Initialize: Set $\{\mathbf{L}_{[0]}, \mathbf{e}_{[0]}, \mathbf{c}_{[0]}, \mathbf{a}_{[0]}, \mathbf{B}_{[0]}\}$ to zero
    matrices, $\mu_{[0]} = 1.25/\|\mathbf{X}(\mathbf{p})\|$ , $\rho = 1.1$;
    while not converged do
        Solve (1);
    end
    $\mathbf{p} \leftarrow \mathbf{p} + \Delta\mathbf{p}$;
end
```

(5) is a quadratic problem which for each $c_i, i \in \{1, \dots k\}$ admits a closed form solution given by:

$$c_{i,[t+1]} = \frac{\mathbf{a}_{[t]}^T \mathbf{u}_i + \mathrm{tr}(\mathbf{B}_{[t]}^T \mathcal{R}(\mathbf{u}_i))}{2\mu_{[t]}} + \frac{\hat{\mathbf{x}}^T \mathbf{u}_i + \mathbf{L}_{[t+1]}^T \mathcal{R}(\mathbf{u}_i)}{2}, \tag{6}$$

where $\hat{\mathbf{x}} = \mathbf{x}(\mathbf{p}) + \mathbf{J}(\mathbf{p})\Delta\mathbf{p}_{[t]} - \mathbf{e}_{[t]}$.

Step 3. Update $\Delta\mathbf{p}$:

$$\mathcal{V}_{[t+1]}^{(3)} = \underset{\mathcal{V}^{(3)}}{\mathrm{argmin}} \ \frac{\mu}{2}\|H^{(1)}(\Delta\mathbf{p}, \mathbf{c}, \mathbf{e}) + \frac{\mathbf{a}}{\mu}\|_2^2. \tag{7}$$

The increment of the parameters $\Delta\mathbf{p}$ is computed by solving the least square problem (7):

$$\mathcal{V}_{[t+1]}^{(3)} = -\big(\mathbf{J}(\mathbf{p})^T\mathbf{J}(\mathbf{p})\big)^{-1}\mathbf{J}(\mathbf{p})^T\big(\mathbf{x}(\mathbf{p}) - \mathbf{U}\mathbf{c}_{[t]} - \mathbf{e}_{[t]} + \frac{\mathbf{a}_{[t]}}{\mu_{[t]}}\big). \tag{8}$$

Step 4. Update e:

$$\mathcal{V}_{[t+1]}^{(4)} = \underset{\mathcal{V}^{(4)}}{\mathrm{argmin}} \quad \lambda\|\mathbf{e}\|_1 + \frac{\mu}{2}\|H^{(1)}(\Delta\mathbf{p}, \mathbf{c}, \mathbf{e}) + \frac{\mathbf{a}}{\mu}\|_2^2. \tag{9}$$

The closed-form solution of (9) is given by applying element-wise the shrinkage operator onto: $\mathbf{x}(\mathbf{p}) + \mathbf{J}(\mathbf{p})\Delta\mathbf{p} - \mathbf{U}\mathbf{c} + \mathbf{a}/\mu$, namely:

$$\mathcal{V}_{[t+1]}^{(4)} = \mathcal{S}_{\frac{\lambda}{\mu_{[t]}}}\left[\mathbf{x}(\mathbf{p}) + \mathbf{J}(\mathbf{p})\Delta\mathbf{p}_{[t+1]} - \mathbf{U}\mathbf{c}_{[t+1]} + \frac{\mathbf{a}_{[t]}}{\mu_{[t]}}\right]. \tag{10}$$

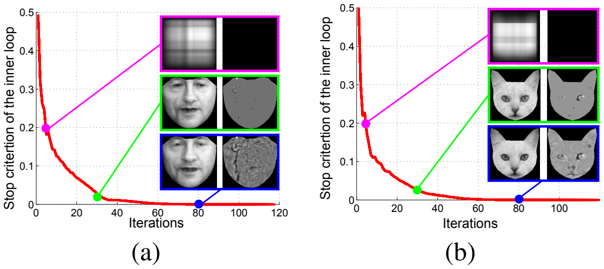
Step 5. Update Lagrange multipliers $\mathbf{a}, \mathbf{B}$ and μ : The Lagrange multipliers are updated by:

$$\begin{cases} \mathbf{a}_{[t+1]} &= \mathbf{a}_{[t]} + \mu_{[t]} \cdot H^{(1)}(\Delta\mathbf{p}_{[t+1]}, \mathbf{c}_{[t+1]}, \mathbf{e}_{[t+1]}) \\ \mathbf{B}_{[t+1]} &= \mathbf{B}_{[t]} + \mu_{[t]} \cdot H^{(2)}(\mathbf{L}_{[t+1]}, \mathbf{c}_{[t+1]}) \\ \mu_{[t+1]} &= min(\rho \cdot \mu_{[t]}, 10^{10}) \end{cases} \tag{11}$$

Convergence criteria: The inner loop of the Alg. 1 terminates when:

$$\begin{cases} \max\big(& \|\mathbf{e}_{[t+1]} - \mathbf{e}_{[t]}\|_2/\|\mathbf{x}(\mathbf{p})\|_2, \\ & \|\mathbf{L}_{[t+1]} - \mathbf{L}_{[t]}\|_F/\|\mathbf{x}(\mathbf{p})\|_2\big) \leq \epsilon_1 \\ \max\big(& \|H^{(1)}(\Delta\mathbf{p}_{[t+1]}, \mathbf{c}_{[t+1]}, \mathbf{e}_{[t+1]})\|_2/\|\mathbf{x}(\mathbf{p})\|_2, \\ & \|H^{(2)}(\mathbf{L}_{[t+1]}, \mathbf{c}_{[t+1]})\|_F/\|\mathbf{x}(\mathbf{p})\|_2\big) \leq \epsilon_2 \end{cases} \tag{12}$$

The Alg. 1 terminates when the change of the $\|\mathbf{L}\|_* + \lambda\|\mathbf{E}\|_1$ between two successive iterations is smaller than a predefined threshold ϵ_3 or the maximum number of the outers' loop iterations is reached.



(a) (b)

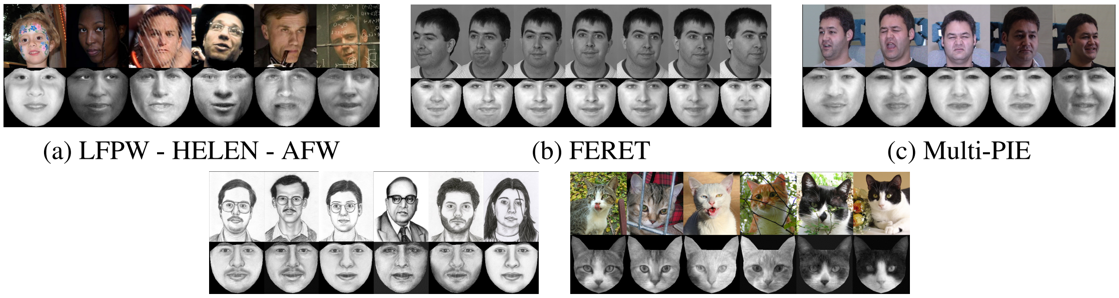
Figure 2. The convergence curve of the Algorithm's 1 inner loop in case of (a) human face and (b) cat face.

Computational Complexity: The dominant cost of each iteration of Alg. 1 is that of the Singular Value Decomposition (SVD) algorithm involved in the computation of the SVT operator in update of $\mathbf{L}$ (Step 1). Consequently, the computational complexity of Alg. 1 is $\mathcal{O}(T(min(m,n)^3 + n^2m))$, where T is the total number of iterations until convergence.

Convergence: Regarding the convergence of the Alg. 1 there is currently no theoretical proof known for the ADMM in problems with more than two blocks of variables. However ADMM has been applied successfully in non-convex optimization problems in practice [23, 25, 29]. In addition, the thorough experimental evaluation of the proposed method presented in Sec. 3, indicates that the convergence of Alg. 1 is empirically proved. In Fig. 2, the empirical convergence curves of the inner loop of Alg. 1 for the cases of human and cat faces are depicted. The low-rank and error images produced after 30, 50 and 117 iterations, respectively, are also shown.

3. Experimental Evaluation

The performance of the RSF is assessed in four different tasks: (i) *frontal view reconstruction*, (ii) *landmark localization*, (iii) *pose-invariant face recognition*, and (iv) *face verification in unconstrained (in-the-wild) conditions*, by conducting experiments on 8 databases which are presented in Table 2. For the CAT database, 350 out of 10000 cat images were re-annotated by employing a dense mark-up scheme consisting of 48 points. In case of sketches, 375



(a) LFPW - HELEN - AFW

(b) FERET

(c) Multi-PIE

(d) FS

(e) CAT

Figure 3. Reconstructed frontal views of unseen objects under controlled and in-the-wild conditions.

Database	Object	# Images	Conditions	# Points
LFPW [6]	Face	1035	In-the-Wild	68
HELEN [18]		2330		68
AFW [48]		468		68
LFW [15]		13233		-
FERET [26]		14051	Controlled	-
Multi-PIE [11]		750000		68
FS [39], [46]	Sketches	1800		35 /3
CAT [45]	Cat	10000	In-the-Wild	9

Table 2. Overview of the used databases.

face sketches (305 images taken from [39], [46] and another 53 images download from the web) were used. All images were annotated using the typical 68 points mark-up scheme employed in LFPW, HELEN, and AFW [30, 31].

3.1. Experimental setup

In all experiments, the orthonormal clean frontal subspace $\mathbf{U}^3$ was constructed by employing only frontal view face images without occlusions. The images were warped in a reference frame by using the $\mathcal{W}$. Subsequently a PCA was applied on the warped shape-free textures. Then, the first k eigen-images with the highest variance were used to form the $\mathbf{U}$. Unless otherwise stated, throughout the experiments, the parameters of the Alg. 1 were fixed as follows: $\lambda = 0.3$, $\rho = 1.1$, $\epsilon_1 = 10^{-5}$, $\epsilon_2 = 10^{-7}$, and $\epsilon_3 = 10^{-3}$.

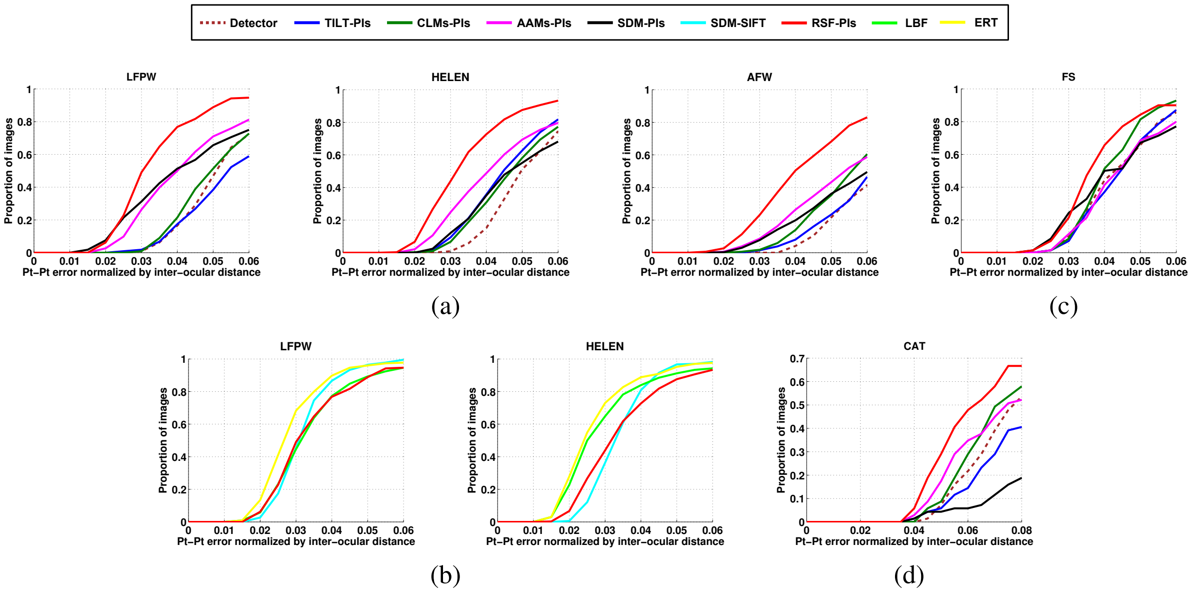
3.2. Reconstruction of frontal view

The ability of the RSF to reconstruct the frontal view from non-frontal images of unseen faces is investigated in this Section. Given the test image and initial landmarks a warped version of the image is produced by employing the $\mathcal{W}$. Next, (1) is solved iteratively. In each iteration $t+1$, a low-rank (frontalized) image ($\mathbf{L}_{[t+1]}$), an error image ($\mathbf{E}_{[t+1]}$), coefficients ($\mathbf{c}_{[t+1]}$) and increments $\Delta\mathbf{p}_{[t+1]}$ of parameters $\mathbf{p}$ are obtained. The new position of the landmarks is then computed by employing the updated parameters $\mathbf{p} \leftarrow \mathbf{p} + \Delta\mathbf{p}$. The test image is then warped using the new landmarks and (1) is solved again (INNER loop of Alg. 1). Finally, after the convergence of Alg. 1, the final frontalized test image, location of the landmarks, and error image are produced. All the frontalizations presented in this Section were created by using the $\mathbf{U}_{\mathrm{W}}$, $\mathbf{U}_{\mathrm{C}}$, and $\mathbf{U}_{\mathrm{S}}$.

In Fig. 3 (a) the frontalized views of unseen faces from the in-the-wild images are illustrated. Fig. 3 (b) and (c) depict the frontal reconstructed views from the non-frontal images of '00268' subject from FERET with 'Neutral' expression and pose $[-40° : 40°]$ and images from Multi-PIE with (i) 'Surprise' at $-30°$, (ii) 'Scream' at $-15°$, (iii) 'Squint' at $0°$, (iv) 'Neutral' at $+15°$, and (v) 'Smile' at $+30°$. The efficacy of the RSF is also assessed by creating the frontal view of face sketches and cat faces. The obtained reconstructions for these objects are depicted in Fig. 3(d) and (e). By visually inspecting the results, it is clear that the RSF is robust to many variations such as pose, expression, sparse occlusions, and lighting conditions. This attributed to the fact that the matrix ℓ_1-norm was adopted for sparse non-Gaussian noise characterization.

To quantitatively assess the quality of the frontalized images the following experiment was conducted. 'Neutral' images of 20 different subjects from Multi-PIE under poses $[-30^0$ to $30^0]$ (5 for each subject, 100 in total) were selected. The images of each subject were frontalized by employing the RSF. The Root Mean Square Error (RMSE) between each frontalized image and the real frontal image of the subject is used as the evaluation metric. The performance of the RSF with respect to RMSE is compared with that obtained by the frontalization method of the DeepFace [35]. The average RMSEs of the RSF and DeepFace were 0.0817 and 0.1025, respectively. It is worth noting that, even though DeepFace employs a 3D model to handle out-of-planar rotations, the RSF performs better without using any kind of 3D information.

[3] The employed frontal subspaces were created from training sets of the databases: $\mathbf{U}_{\mathrm{W}}$: 500 LFPW & HELEN, $\mathbf{U}_{\mathrm{L}}$: 209 LFPW, $\mathbf{U}_{\mathrm{H}}$: 284 HELEN, $\mathbf{U}_{\mathrm{S}}$: 261 FS, $\mathbf{U}_{\mathrm{C}}$: 305 CATS



(a) (c)

(b) (d)

Figure 4. Cumulative error distribution curves on LFPW, HELEN, AFW, FS, and CAT databases: (a), (c), (d): TILT-PIs, CLMS-PIs, AAM-PIs, SDM-PIs, RSF-PIs, (b) RSF-PIs, SDM-SIFT, LBF and ERT.

3.3. Landmark localization

The performance of the RSF for the generic alignment problem is assessed by conducting experiments on *(i)* in-the-wild faces, *(ii)* sketch faces and *(iii)* cat faces. To this end, the performance of the RSF is compared against to that obtained by the TILT [47], AAMs [2, 21], CLMs [32], and SDM [40]. In order to compare fairly the competing methods, the same training data, initialization, and feature representation were employed. For all experiments the simple representation of Pixel Intensities (PIs) was used. The average point-to-point Euclidean distance of N landmark points normalized by the Euclidean distance of the outer corner of eyes is used as the evaluation measure. In addition, the cumulative error distribution curve (CED) for each method is computed by using the fraction of test images for which the average error was smaller than a threshold. Finally, the implementations provided by the platform MENPO [1] were used for all compared methods.

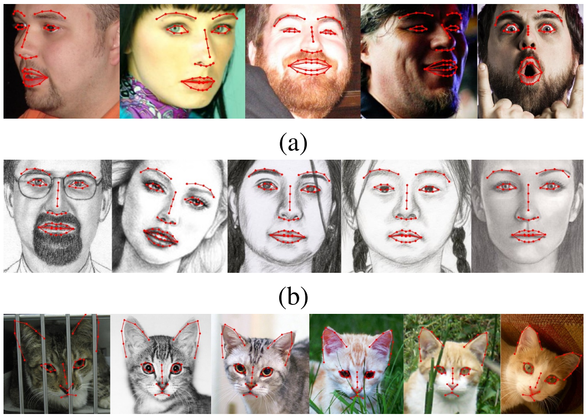
3.3.1 Aligning in-the-wild face images

Same train set and features: The in-the-wild face databases LFPW, HELEN and AFW were employed in order to assess the performance of the RSF in the problem of generic face alignment. The results produced by the detector in [42, 48] were used to initialize all the methods. The annotations provided in [30, 31] have been employed for evaluation purposes. The error for each method was computed based on $N = 49$ interior landmark points (excluding the points correspond to face boundary). Finally, the bases matrices $\mathbf{U}_{\mathrm{L}}$, $\mathbf{U}_{\mathrm{H}}$ and $\mathbf{U}_{\mathrm{W}}$ were used by the RSF.

The CEDs produced by all methods for the LFPW (test set), the HELEN (test set), and the AFW databases are depicted in Fig. 4 (a). Clearly, the RSF outperforms the TILT-PIs, the AAMs-PIs, the CLMs-PIs, and the SDM-PIs. More specifically, for normalized error of 0.05^4 the RSF yield an 20.1%, 21.5% and 24.6% improvement compared to that obtained by the AAMs-PIs across the test databases. TILT performs worst overall which can be explained to the fact that minimizes the unconstrained rank of the image ensemble. The discriminative methods SDM and CLMs yield poor performance because they were trained with only 500 frontal images. In general the discriminative methods require large amount of annotated data in order to yield powerful classifiers and functional mappings. In contrast, the AAMs, which are generative models, achieved better results than the CLMs and SDM.

State-of-the-art methods and features: In this experiment, the RSF is compared against the state-of-the-art methods SDM [40], LBF [28], and ERT [17]. The authors provided pre-trained model and code was used for the SDM, while the LBF and ERT were trained and tested by using the available implementations. In particular, the LBF and ERT were trained using the AFW and train sets of LFPW and HELEN. The parameters were set as explained in the corresponding papers. The CEDs from this experiment are shown in Fig. 4 (b). The RSF achieves comparable performance with that obtained by the competing methods, but it uses only a small set of frontal images for training.This

[4] This value was found by visually inspecting the results.



(a)

(b)

(c)

Figure 5. Sample fitting results produced by RSF-PIs for (a) in-the-wild faces, (b) sketches, and (c) cats.

is contrast to all other methods were trained on thousand images captured under several variations including different poses, illuminations and expression (i.e., train sets of the used databases). Furthermore, the SDM method takes full advantage of SIFT – a powerful hand-crafted feature – while the RSF employs only simple PIs. Fig. 5(a) illustrates fitting examples produced by RSF.

3.3.2 Aligning cat and sketch face images

While the previous experiment concerned in images of human faces, the RSF is a general method capable of aligning any object that the frontal view is that of the lowest rank. In this set of experiments, the ability of the proposed method to align cat faces and face sketches is demonstrated by using the FS and the CAT databases. The matrices $\mathbf{U}_{\mathrm{C}}$, $\mathbf{U}_{\mathrm{S}}$ were employed and the fitting error in case of CAT was calculated based on $N = 37$ interior landmark points (excluding the points of boundary). The results obtained by the compared methods are summarized in CED curves depicted in Fig. 4 (c), (d). The quality of fitting results produced by the RSF can be visually compared in Fig. 5 (b), (c). It is clear from the results that the RSF outperforms all the other methods, validating its ability to handle other symmetric objects and modalities.

Although, we tested the state-of-the-art methods LBF and ERT in the same experiment, their performances were poor. Therefore, in order to avoid the clutter of our figures we do not report their CEDs.

3.4. Pose-invariant face recognition

The performance of the RSF on pose-invariant face recognition with one gallery image per person is assessed by conducting experiments on the Multi-PIE and FERET databases. The experiment proceeds as follows. First, the frontal views of all images used in this experiment were reconstructed following the methodology described in Sec. 3.2 by employing the $\mathbf{U}_{\mathrm{W}}$. In order to remove the surrounding black pixels, the reconstructed frontal views were cropped. Subsequently, the Image Gradient Orientations (IGOs) features [37] were used for image representation. The dimensionality of IGOs was reduced by applying PCA. Finally, the classification was performed by employing the classifier in [43].

The performance of the RSF is compared to 2D based methods: LGBP [44] and PIMRF [13], 3D based methods: 3DPN [4], EGFC [20], and PAF [41], as well as the Deep learning based methods: SPAE [16] and DIPFS [49]. It should be noticed that all methods were evaluated under the fully automatic scenario, where both the bounding box of the face region and the facial landmarks were located automatically.

Results on FERET: One frontal image, denoted as 'ba', from each of the 200 subjects was used to form the gallery set, while the images captured at 6 different poses i.e., $[-40° : 40°]$ were selected as the probe images. In Table 3 the recognition rates achieved by the competing methods in the different poses are reported. Clearly, the RSF (recognition accuracy 98.58%) outperforms both the 2D and 3D state-of-the-art methods. It is worth mentioning that the PIMRF employs 200 images from the FERET database (different from the test set) in order to train the frontal synthesizer. Consequently, the different lighting conditions of the database are taken into account. This is not the case for the RSF where only frontal images from a generic in-the-wild database (i.e., the LFPW and HELEN) have been used. Even though the RSF does not use any kind of 3D information, it performs comparably to the PAF where an elaborated 3D model (trained from 4624 facial scans) has been used to find the 3D pose and extract pose adaptive features. The reported results of the EGFC [20] were not included in Table 3 as they were obtained using a semi-automatic protocol (i.e., 5 manually annotated landmark points used).

Table 3. Recognition rates achieved by the compared methods on the FERET database.

Method	$-40°$	$-25°$	$-15°$	$15°$	$25°$	$40°$	Avg
LGBP [44]	62.0%	91.0%	98.0%	96.0 %	84.0%	51.0 %	80.5 %
3DPN [4]	90.5%	98.0%	98.5%	97.5%	97.0%	91.9%	95.6%
PIMRF [13]	91.0%	97.3%	98.0%	98.5%	96.5%	91.5%	95.5%
PAF [41]	98.0%	98.5%	99.25%	99.25%	98.5%	98.0%	98.56%
RSF	96.5%	99.0%	100.0%	100.0%	100%	96%	98.58%

Results on Multi-PIE: The images of 137 subjects (Subject ID $201 : 346$) with 'Neutral' expression and poses $[-30° : +30°]$ captured under 4 different sessions were selected. The gallery was created by the frontal images of the earliest session for each subject, while the rest of images including frontal and non-frontal views were used as probes. It should be mentioned that images of first 200 subjects which include all poses (4207 in total) were not used for training purposes by the RSF. Those images were used in the 3DPN to train view-based models, in the SPAE, DIPFS to train the deep networks, and in the EGFC to train the pose estimator and matching model. The recognition accuracy achieved by the compared methods is reported in Table 4. The RSF outperforms four out of five methods that is compared to, verifying the high quality of the frontalized images. The RSF also performs comparable to the DIPFS though only using 500 frontal images outside the Multi-PIE.

Table 4. Recognition rates achieved by the compared methods on the Multi-PIE database.

Method	$-30°$	$-15°$	$0°$	$15°$	$30°$	Avg
PIMRF [13]	89.7%	91.7%	92.5%	91.0%	89.0%	90.78%
3DPN [4]	91.0%	95.7%	96.9%	95.7%	89.5%	93.76%
SPAE [16]	92.6%	96.3%	-	95.7%	94.3%	94.72%
EGFC [20]	95.0%	99.3%	-	99.0%	92.9%	96.55%
DIPFS [49]	98.5%	100%	-	99.3%	98.5%	99.07%
RSF	94.3%	98.7%	99.4%	97.3%	95.6%	97.06%

3.5. Face verification

The performance of the RSF in the face verification under in-the-wild conditions is assessed by conducting experiment in the LFW database, using the 'image-restricted, no outside data results' setting. The standard evaluation protocol, which splits the View 2 dataset into 10 folds, with each fold consisting of 300 intra-class pairs and 300 inter-class pairs, was employed.

In this experiment the basis $\mathbf{U}$ and the detector in [48] were not used since they are based on images outside the database. To create the initializations and a new $\mathbf{U}_{\text{LFW}}$, the method for automatic construction of deformable models presented in [3] was employed. The goal of this method is to build a deformable model using only a set of images with the corresponding face bounding boxes. To define the face bounding boxes without using a pre-trained detector, the deep funneled images of the LFW [14] were employed. Therefore, since these images are aligned, the exact face bounding box is known. Subsequently, a deformable model was built automatically from the training images of each fold. The created model was fitted to all images and those (from training images) with fitted shapes similar to the mean shape were selected to build the basis $\mathbf{U}_{\text{LFW}}$. In each fold the images were frontalized using the $\mathbf{U}_{\text{LFW}}$ and they were cropped subsequently. The gradient orientations ϕ_1, ϕ_2 of each image pair were extracted and the cosine of difference between them $\Delta\phi = \phi_1 - \phi_2$ was normalized to the range $[0 - 2\pi]$ and used as the feature of the pair. These features are classified by a Support Vector Machine (SVM) with an RBF kernel. The performance of the RSF is compared against that obtained by the Fisher Vector Faces [33], MRF-Fusion-CSKDA [27], and the EigenPEP [19] methods[5]. The mean classification accuracy and the corresponding standard deviation computed based on 10 folds are reported in Table 5. By inspecting Table 5 it can be seen that the RSF outperforms the MRF-MLBP and the Fisher Vector Faces and performs comparably with the recently published method EigenPEP. In [27] by using an MRF, which optimization is computationally heavy, for dense image matching and three different multi-scale features an accuracy of 0.9589 ± 0.0194 is achieved. We tested the proposed framework by using different multiple features and we observed the same accuracy improvement as in [27]. However, the scope of the conducted experiment was to validate the quality of the frontalized images and that is why we used only IGOs which is a very simple feature.

Table 5. LFW: Mean classification error and standard deviation.

Method	Mean ± Std
Fisher vector faces [33]	0.8747 ± 0.0149
EigenPEP [19]	0.8897 ± 0.0132
LFW3D-IGOs-SVM [12]	0.7928 ± 0.0175
RSF	0.8881 ± 0.0078

In order to compare the RSF with the recently proposed frontalized version of LFW named LFW3D [12], the same classification framework as before was applied. The achieved accuracy is 79.28% while the accuracy achieved by the RSF is 88.21%. This is a quite interesting result since the proposed RSF method does not use any kind of 3D information. This is due to the fact that in RSF sparse noise such as occlusions and illuminations is removed from the frontalized images.

4. Conclusions

In this paper, to the best our knowledge, we presented the first method that jointly performs landmark localization and face frontalization using only a simple statistical model of few hundred frontal images. The proposed method outperforms state-of-the-art methods for face landmark localization that were trained on thousands of images in many poses and achieves comparable results in pose-invariant face recognition and verification without using 3D elaborate models or Deep Learning-based features extraction.

Acknowledgements

This work has been funded by the EPSRC project EP/J017787/1 (4DFAB). The work of S. Zafeiriou was also partially supported by the EPSRC project EP/L026813/1 Adaptive Facial Deformable Models for Tracking (ADAManT). The work of Y. Panagakis and M. Pantic is supported by the European Community Horizon 2020 [H2020/2014-2020] under grant agreement no. 645094 (SEWA).

[5] `vis-www.cs.umass.edu/lfw`

References

[1] J. Alabort-i Medina, E. Antonakos, J. Booth, P. Snape, and S. Zafeiriou. Menpo: A comprehensive platform for parametric image alignment and visual deformable models. MM, pages 679–682. ACM, 2014.

[2] E. Antonakos, J. Alabort-i-Medina, G. Tzimiropoulos, and S. Zafeiriou. Feature-based lucas-kanade and active appearance models. 24(9):2617 – 2632, 2015.

[3] E. Antonakos and S. Zafeiriou. Automatic construction of deformable models in-the-wild. In *CVPR*, pages 1813–1820, 2014.

[4] A. Asthana, T. K. Marks, M. J. Jones, K. H. Tieu, and M. Rohith. Fully automatic pose-invariant face recognition via 3d pose normalization. In *CVPR*, pages 937–944, 2011.

[5] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic. Robust discriminative response map fitting with constrained local models. In *CVPR*, pages 3444–3451, 2013.

[6] P. N. Belhumeur, D. W. Jacobs, D. Kriegman, and N. Kumar. Localizing parts of faces using a consensus of exemplars. In *CVPR*, pages 545–552, 2011.

[7] J.-F. Cai, E. J. Candès, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010.

[8] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM*, 58(3):11, 2011.

[9] X. Cheng, C. Fookes, S. Sridharan, J. Saragih, and S. Lucey. Deformable face ensemble alignment with robust grouped-l1 anchors. In *FGR*, pages 1–7, 2013.

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

[11] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-pie. *IMAVIS*, 28(5):807–813, 2010.

[12] T. Hassner, S. Harel, E. Paz, and R. Enbar. Effective face frontalization in unconstrained images. *CVPR*, 2015.

[13] H. T. Ho and R. Chellappa. Pose-invariant face recognition using markov random fields. *TIP*, 22(4):1573–1584, 2013.

[14] G. Huang, M. Mattar, H. Lee, and E. G. Learned-Miller. Learning to align from scratch. In *NIPS*, pages 764–772, 2012.

[15] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, Univ. of Massachusetts, Amherst, October 2007.

[16] M. Kan, S. Shan, H. Chang, and X. Chen. Stacked progressive autoencoders (spae) for face recognition across poses. In *CVPR*, pages 1883–1890, 2014.

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

[18] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang. Interactive facial feature localization. In *ECCV*, pages 679–692. 2012.

[19] H. Li, G. Hua, X. Shen, Z. Lin, and J. Brandt. Eigen-pep for video face recognition. *ACCV*, 2104, 2014.

[20] S. Li, X. Liu, X. Chai, H. Zhang, S. Lao, and S. Shan. Morphable displacement field based image matching for face recognition across pose. In *ECCV*, pages 102–115. 2012.

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

[22] Y. Panagakis, C. Kotropoulos, and G. Arce. Music genre classification via joint sparse low-rank representation of audio features. *IEEE/ACM TASLP*, 22(12):1905–1917, 2014.

[23] Y. Panagakis, M. Nicolaou, S. Zafeiriou, M. Pantic, et al. Robust canonical time warping for the alignment of grossly corrupted sequences. In *CVPR*, pages 540–547, 2013.

[24] G. Papamakarios, Y. Panagakis, and S. Zafeiriou. Generalised scalable robust principal component analysis. In *BMVC*, 2014.

[25] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma. Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *TPAMI*, 34(11):2233–2246, 2012.

[26] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The feret evaluation methodology for face-recognition algorithms. *TPAMI*, 22(10):1090–1104, 2000.

[27] S. Rahimzadeh Arashloo and J. Kittler. Class-specific kernel fusion of multiple descriptors for face verification using multiscale binarised statistical image features.

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

[29] C. Sagonas, Y. Panagakis, S. Zafeiriou, and M. Pantic. RAPS: Robust and efficient automatic construction of person-specific deformable models. In *CVPR*, pages 1789–1796, 2014.

[30] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic. 300 faces in-the-wild challenge: The first facial landmark localization challenge. In *ICCV-W*, pages 397–403, 2013.

[31] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic. A semi-automatic methodology for facial landmark annotation. In *CVPR-W*, pages 896–903, 2013.

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

[33] K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman. Fisher vector faces in the wild. In *BMVC*, volume 1, page 7, 2013.

[34] Y. Sun, X. Wang, and X. Tang. Deep learning face representation from predicting 10,000 classes. In *CVPR*, pages 1891–1898, 2014.

[35] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. In *CVPR*, pages 1701–1708, 2014.

[36] G. Tzimiropoulos, J. Alabort-i Medina, S. Zafeiriou, and M. Pantic. Generic active appearance models revisited. In *ACCV*, pages 650–663. 2013.

[37] G. Tzimiropoulos, S. Zafeiriou, and M. Pantic. Subspace learning from image gradient orientations. *TPAMI*, 34:2454–2466, 2012.

[38] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma. Toward a practical face recognition system: Robust alignment and illumination by sparse representation. *TPAMI*, 34(2):372–386, 2012.

[39] X. Wang and X. Tang. Face photo-sketch synthesis and recognition. *TPAMI*, 31(11):1955–1967, 2009.

[40] X. Xiong and F. De la Torre. Supervised descent method and its applications to face alignment. In *CVPR*, pages 532–539, 2013.

[41] D. Yi, Z. Lei, and S. Z. Li. Towards pose robust face recognition. In *CVPR*, pages 3539–3545, 2013.

[42] S. Zafeiriou, C. Zhang, and Z. Zhang. A survey on face detection in the wild: past, present and future. *Computer Vision and Image Understanding*, 2015.

[43] D. Zhang, M. Yang, and X. Feng. Sparse representation or collaborative representation: Which helps face recognition? In *ICCV*, pages 471–478, 2011.

[44] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang. Local gabor binary pattern histogram sequence (lgbphs): A novel non-statistical model for face representation and recognition. In *ICCV*, volume 1, pages 786–791, 2005.

[45] W. Zhang, J. Sun, and X. Tang. Cat head detection-how to effectively exploit shape and texture features. In *ECCV*. 2008.

[46] W. Zhang, X. Wang, and X. Tang. Coupled information-theoretic encoding for face photo-sketch recognition. In *CVPR*, pages 513–520, 2011.

[47] Z. Zhang, A. Ganesh, X. Liang, and Y. Ma. TILT: transform invariant low-rank textures. *IJCV*, 99(1):1–24, 2012.

[48] X. Zhu and D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. In *CVPR*, pages 2879–2886, 2012.

[49] Z. Zhu, P. Luo, X. Wang, and X. Tang. Deep learning identity-preserving face space. In *ICCV*, pages 113–120, 2013.