# Supplementary Material for: Active Pictorial Structures

Epameinondas Antonakos Joan Alabort-i-Medina Stefanos Zafeiriou
Department of Computing, Imperial College London
180 Queens Gate, SW7 2AZ, London, U.K.
{e.antonakos, ja310, s.zafeiriou}@imperial.ac.uk

In the following sections, we provide additional material for the paper "Active Pictorial Structures". Section 1 explains in more detail the differences between the proposed Active Pictorial Structures (APS) and Pictorial Structures (PS). Section 2 presents the proofs about the structure of the precision matrices of the Gaussian Markov Random Filed (GMRF) (Eqs. 10 and 12 of the main paper). Section 3 gives an analysis about the forward Gauss-Newton optimization of APS and shows that the inverse technique with fixed Jacobian and Hessian, which is used in the main paper, is much faster. Finally, Sec. 4 shows additional experimental results and conducts new experiments on different objects (human eyes and cars). An open-source implementation of APS is available within the Menpo Project [1] in `http://www.menpo.org/`.

## 1. Differences between Active Pictorial Structures and Pictorial Structures

As explained in the main paper, the proposed model is partially motivated by PS [4, 8]. In the original formulation of PS, the cost function to be optimized has the form

$$\begin{aligned}&\arg\min_{\mathbf{s}} \sum_{i=1}^{n} m_i(\boldsymbol{\ell}_i) + \sum_{i,j:(v_i,v_j)\in E} d_{ij}(\boldsymbol{\ell}_i,\boldsymbol{\ell}_j) = \\ &= \arg\min_{\mathbf{s}} \sum_{i=1}^{n} [\mathcal{A}(\boldsymbol{\ell}_i) - \boldsymbol{\mu}_i^a]^T (\boldsymbol{\Sigma}_i^a)^{-1} [\mathcal{A}(\boldsymbol{\ell}_i) - \boldsymbol{\mu}_i^a] + \sum_{i,j:(v_i,v_j)\in E} [\boldsymbol{\ell}_i - \boldsymbol{\ell}_j - \boldsymbol{\mu}_{ij}^d]^T (\boldsymbol{\Sigma}_{ij}^d)^{-1} [\boldsymbol{\ell}_i - \boldsymbol{\ell}_j - \boldsymbol{\mu}_{ij}^d]\end{aligned} \tag{1}$$

where $\mathbf{s} = [\boldsymbol{\ell}_1^T, \dots, \boldsymbol{\ell}_n^T]^T$ is the vector of landmark coordinates ($\boldsymbol{\ell}_i = [x_i, y_i]^T$, $\forall i = 1, \dots, n$), $\mathcal{A}(\boldsymbol{\ell}_i)$ is a feature vector extracted from the image location $\boldsymbol{\ell}_i$ and we have assumed a tree $G = (V, E)$. $\{\boldsymbol{\mu}_i^a, \boldsymbol{\Sigma}_i^a\}$ and $\{\boldsymbol{\mu}_{ij}^d, \boldsymbol{\Sigma}_{ij}^d\}$ denote the mean and covariances of the appearance and deformation respectively. In Eq. 1, $m_i(\boldsymbol{\ell}_i)$ is a function measuring the degree of mismatch when part $v_i$ is placed at location $\boldsymbol{\ell}_i$ in the image. Moreover, $d_{ij}(\boldsymbol{\ell}_i, \boldsymbol{\ell}_j)$ denotes a function measuring the degree of deformation of the model when part $v_i$ is placed at location $\boldsymbol{\ell}_i$ and part $v_j$ is placed at location $\boldsymbol{\ell}_j$. The authors show an inference algorithm based on distance transform [3] that can find a global minimum of Eq. 1 without any initialization. However, this algorithm imposes two important restrictions: (1) appearance of each part is independent of the rest of them and (2) $G$ must always be acyclic (a tree). Additionally, the computation of $m_i(\boldsymbol{\ell}_i)$ for all parts ($i = 1, \dots, n$) and all possible image locations (response maps) has a high computational cost, which makes the algorithm very slow. Finally, in [8], the authors only use a diagonal covariance for the relative locations (deformation) of each edge of the graph, which restricts the flexibility of the model.

In the proposed APS, we aim to minimize the cost function (Eq. 19 of the main paper)

$$\begin{aligned}&\arg\min_{\mathbf{p}} \|\mathcal{A}(\mathcal{S}(\bar{\mathbf{s}}, \mathbf{p})) - \bar{\mathbf{a}}\|^2_{\mathbf{Q}^a} + \|\mathcal{S}(\bar{\mathbf{s}}, \mathbf{p}) - \bar{\mathbf{s}}\|^2_{\mathbf{Q}^d} = \\ &= \arg\min_{\mathbf{p}} [\mathcal{A}(\mathcal{S}(\bar{\mathbf{s}}, \mathbf{p})) - \bar{\mathbf{a}}]^T \mathbf{Q}^a [\mathcal{A}(\mathcal{S}(\bar{\mathbf{s}}, \mathbf{p})) - \bar{\mathbf{a}}] + [\mathcal{S}(\bar{\mathbf{s}}, \mathbf{p}) - \bar{\mathbf{s}}]^T \mathbf{Q}^d [\mathcal{S}(\bar{\mathbf{s}}, \mathbf{p}) - \bar{\mathbf{s}}]\end{aligned} \tag{2}$$

There are two main differences between APS and PS: (1) we employ a statistical shape model and optimize with respect to its parameters and (2) we use the efficient Gauss-Newton optimization technique. However, these differences introduce some important advantages, as also mentioned in the main paper. The proposed formulation allows to define a graph (not only tree) between the object's parts. This means that we can assume dependencies between any pair of landmarks for both the appearance and the deformation, as opposed to PS that assumes independence for the appearance and a tree structure for the deformation. As shown in the experimental results of the main paper (Sec. 3.1), this lack of restriction is very beneficial. Finally, even though the efficient Gauss-Newton APS optimization does not find a global optimum, it handles the cost function in its matricial form (not in sums as in Eq. 1) and with an inverse-compositional manner, which ends up in much faster computational time.

## 2. Precision matrix form of graphical model

Herein we provide a proof for the precision matrix formulations of Eqs. 10 and 12 of the main paper. For this purpose, let us define an undirected graph $G = (V, E)$ of $n$ vertexes, where $V = \{v_1, v_2, \dots, v_n\}$ is the set of vertexes and there is an edge $(v_i, v_j) \in E$ for each pair of connected vertexes.

### 2.1. Properties

The following properties can be easily proved.

Property 1: If $\begin{cases} f(i,j) \neq 0, \forall i,j : (v_i, v_j) \in E \\ f(i,j) = 0, \forall i,j : (v_i, v_j) \notin E \end{cases}$ then $\displaystyle\sum_{\forall i,j:(v_i,v_j)\in E} f(i,j) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} f(i,j)$.

Property 2: $\displaystyle\sum_{\forall i,j:(v_i,v_j)\in E} f(i) + f(j) = \sum_{i=1}^{n} c_i f(i)$, where $c_i = \displaystyle\sum_{\forall j:(v_i,v_j)\in E} 1 + \sum_{\forall j:(v_j,v_i)\in E} 1$ denotes the number of neighbours of vertex $v_i$.

### 2.2. Proof 1

Herein we provide a proof for the precision matrix formulation of Eq. 10. Assume that we have a set of vectors of length $k$ that correspond to each vertex, i.e. $\mathbf{x}_i = [x_1^i, x_2^i, \dots, x_k^i], \forall i : v_i \in V$. Moreover, let us assume a set of symmetrix pairwise precision matrices for each edge of the graph of size $2k \times 2k$, that have the form

$$\mathbf{Q}^{ij} = \begin{bmatrix} \mathbf{Q}_i & \mathbf{Q}_{ij} \\ \mathbf{Q}_{ij}^T & \mathbf{Q}_j \end{bmatrix}, \forall i,j : (v_i, v_j) \in E \tag{3}$$

We aim to find the structure of $\mathbf{Q}$, so that

$$\sum_{\forall i,j:(v_i,v_j)\in E} \begin{bmatrix} \mathbf{x}_i \\ \mathbf{x}_j \end{bmatrix}^T \mathbf{Q}^{ij} \begin{bmatrix} \mathbf{x}_i \\ \mathbf{x}_j \end{bmatrix} = \mathbf{x}^T \mathbf{Q} \mathbf{x} \tag{4}$$

where $\mathbf{x} = [\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_n^T]^T$.

By separating the $kn \times kn$ matrix $\mathbf{Q}$ in blocks of size $k \times k$ as

$$\mathbf{Q} = \begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} & \cdots & \mathbf{K}_{1n} \\ \mathbf{K}_{21} & \mathbf{K}_{22} & \cdots & \mathbf{K}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{K}_{n1} & \mathbf{K}_{n2} & \cdots & \mathbf{K}_{nn} \end{bmatrix} \tag{5}$$

the second part of Eq. 4 can be written as

$$\mathbf{x}^T \mathbf{Q} \mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix}^T \begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} & \cdots & \mathbf{K}_{1n} \\ \mathbf{K}_{21} & \mathbf{K}_{22} & \cdots & \mathbf{K}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{K}_{n1} & \mathbf{K}_{n2} & \cdots & \mathbf{K}_{nn} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix} = \sum_{i=1}^{n} \mathbf{x}_i^T \mathbf{K}_{ii} \mathbf{x}_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left( \mathbf{x}_i^T \mathbf{K}_{ij} \mathbf{x}_j + \mathbf{x}_j^T \mathbf{K}_{ji} \mathbf{x}_i \right) \tag{6}$$

Given the properties of Sec. 2.1, the first part of Eq. 4 can be written as

$$\sum_{\forall i,j:(v_i,v_j)\in E} \begin{bmatrix} \mathbf{x}_i \\ \mathbf{x}_j \end{bmatrix}^T \mathbf{Q}^{ij} \begin{bmatrix} \mathbf{x}_i \\ \mathbf{x}_j \end{bmatrix} = \sum_{\forall i,j:(v_i,v_j)\in E} \mathbf{x}_i^T \mathbf{Q}_i \mathbf{x}_i + \mathbf{x}_j^T \mathbf{Q}_j \mathbf{x}_j + 2\mathbf{x}_i^T \mathbf{Q}_{ij} \mathbf{x}_j = \sum_{i=1}^{n} c_i \mathbf{x}_i^T \mathbf{Q}_i \mathbf{x}_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2\mathbf{x}_i^T \mathbf{Q}_{ij} \mathbf{x}_j \tag{7}$$

By equalizing Eqs. 6 and 7 we get

$$\sum_{i=1}^{n}\mathbf{x}_i^T\mathbf{K}_{ii}\mathbf{x}_i+\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left(\mathbf{x}_i^T\mathbf{K}_{ij}\mathbf{x}_j+\mathbf{x}_j^T\mathbf{K}_{ji}\mathbf{x}_i\right)=\sum_{i=1}^{n}c_i\mathbf{x}_i^T\mathbf{Q}_i\mathbf{x}_i+\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}2\mathbf{x}_i^T\mathbf{Q}_{ij}\mathbf{x}_j\Rightarrow$$

$$\Rightarrow\begin{cases}\mathbf{x}_i^T\mathbf{K}_{ii}\mathbf{x}_i=c_i\mathbf{x}_i^T\mathbf{Q}_i\mathbf{x}_i\\ \mathbf{x}_i^T\mathbf{K}_{ij}\mathbf{x}_j+\mathbf{x}_j^T\mathbf{K}_{ji}\mathbf{x}_i=2\mathbf{x}_i^T\mathbf{Q}_{ij}\mathbf{x}_j\end{cases}\Rightarrow\begin{cases}\mathbf{x}_i^T\mathbf{K}_{ii}\mathbf{x}_i=\mathbf{x}_i^T(c_i\mathbf{Q}_i)\mathbf{x}_i\\ \mathbf{x}_i^T\mathbf{K}_{ij}\mathbf{x}_j+\left(\mathbf{x}_i^T\mathbf{K}_{ji}^T\mathbf{x}_j\right)^T=\mathbf{x}_i^T(2\mathbf{Q}_{ij})\mathbf{x}_j\end{cases}\Rightarrow\begin{cases}\mathbf{K}_{ii}=c_i\mathbf{Q}_i\\ \mathbf{K}_{ij}=\mathbf{K}_{ji}^T=\mathbf{Q}_{ij}\end{cases}\tag{8}$$

Consequently, by defining $\mathcal{G}_i=\{(i-1)k+1,(i-1)k+2,\dots,ik\}$ to be a set of sampling indices and given Eq. 3, in order for Eq. 4 to be true, the structure of $\mathbf{Q}$ is

$$\mathbf{Q}=\begin{cases}\sum\limits_{\forall j:(v_i,v_j)\in E}\mathbf{Q}_{ij}(\mathcal{G}_1,\mathcal{G}_1)+\sum\limits_{\forall j:(v_j,v_i)\in E}\mathbf{Q}_{ji}(\mathcal{G}_2,\mathcal{G}_2),\ \forall v_i\in V, & \text{at }(\mathcal{G}_i,\mathcal{G}_i)\\ \mathbf{Q}_{ij}(\mathcal{G}_1,\mathcal{G}_2),\ \forall i,j:(v_i,v_j)\in E, & \text{at }(\mathcal{G}_i,\mathcal{G}_j)\text{ and }(\mathcal{G}_j,\mathcal{G}_i)\\ 0, & \text{elsewhere}\end{cases}\tag{9}$$

### 2.3. Proof 2

Similar to the previous case, herein we provide a proof for the precision matrix formulation of Eq. 12. Again, assume that we have a set of vectors of length $k$ that correspond to each vertex, i.e. $\mathbf{x}_i=[x_1^i,x_2^i,\dots,x_k^i],\forall i:v_i\in V$. We aim to find the structure of $\mathbf{Q}$, so that

$$\sum_{\forall i,j:(v_i,v_j)\in E}\left[\mathbf{x}_i-\mathbf{x}_j\right]^T\mathbf{Q}^{ij}\left[\mathbf{x}_i-\mathbf{x}_j\right]=\mathbf{x}^T\mathbf{Q}\mathbf{x}\tag{10}$$

where $\mathbf{Q}^{ij}$ is the $k\times k$ precision matrix corresponding to $\mathbf{x}_i-\mathbf{x}_j$ and $\mathbf{x}=[\mathbf{x}_1^T,\mathbf{x}_2^T,\dots,\mathbf{x}_n^T]^T$.

By separating the $kn\times kn$ matrix $\mathbf{Q}$ in blocks of size $k\times k$ as shown in Eq. 5, the second part of Eq. 10 has the same form as shown in Eq. 6. Given the properties of Sec. 2.1, the first part of Eq. 10 can be written as

$$\begin{aligned}\sum_{\forall i,j:(v_i,v_j)\in E}\left[\mathbf{x}_i-\mathbf{x}_j\right]^T\mathbf{Q}^{ij}\left[\mathbf{x}_i-\mathbf{x}_j\right]&=\sum_{\forall i,j:(v_i,v_j)\in E}\left[\mathbf{x}_i^T\mathbf{Q}^{ij}-\mathbf{x}_j^T\mathbf{Q}^{ij}\right]\left[\mathbf{x}_i-\mathbf{x}_j\right]=\\ =\sum_{\forall i,j:(v_i,v_j)\in E}\mathbf{x}_i^T\mathbf{Q}^{ij}\mathbf{x}_i+\mathbf{x}_j^T\mathbf{Q}^{ij}\mathbf{x}_j-\mathbf{x}_i^T\mathbf{Q}^{ij}\mathbf{x}_j-(\mathbf{x}_i^T(\mathbf{Q}^{ij})^T\mathbf{x}_j)^T&=\\ =\sum_{\forall i,j:(v_i,v_j)\in E}\mathbf{x}_i^T\mathbf{Q}^{ij}\mathbf{x}_i+\mathbf{x}_j^T\mathbf{Q}^{ij}\mathbf{x}_j-2\mathbf{x}_i^T\mathbf{Q}^{ij}\mathbf{x}_j&=\sum_{i=1}^{n}c_i\mathbf{x}_i^T\mathbf{Q}^{ij}\mathbf{x}_i-\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}2\mathbf{x}_i^T\mathbf{Q}^{ij}\mathbf{x}_j\end{aligned}\tag{11}$$

By equalizing Eqs. 6 and 11 we get

$$\sum_{i=1}^{n}\mathbf{x}_i^T\mathbf{K}_{ii}\mathbf{x}_i+\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left(\mathbf{x}_i^T\mathbf{K}_{ij}\mathbf{x}_j+\mathbf{x}_j^T\mathbf{K}_{ji}\mathbf{x}_i\right)=\sum_{i=1}^{n}c_i\mathbf{x}_i^T\mathbf{Q}^{ij}\mathbf{x}_i-\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}2\mathbf{x}_i^T\mathbf{Q}^{ij}\mathbf{x}_j\Rightarrow$$

$$\Rightarrow\begin{cases}\mathbf{x}_i^T\mathbf{K}_{ii}\mathbf{x}_i=c_i\mathbf{x}_i^T\mathbf{Q}^{ij}\mathbf{x}_i\\ \mathbf{x}_i^T\mathbf{K}_{ij}\mathbf{x}_j+\mathbf{x}_j^T\mathbf{K}_{ji}\mathbf{x}_i=-2\mathbf{x}_i^T\mathbf{Q}^{ij}\mathbf{x}_j\end{cases}\Rightarrow\begin{cases}\mathbf{x}_i^T\mathbf{K}_{ii}\mathbf{x}_i=\mathbf{x}_i^T(c_i\mathbf{Q}^{ij})\mathbf{x}_i\\ \mathbf{x}_i^T\mathbf{K}_{ij}\mathbf{x}_j+\left(\mathbf{x}_i^T\mathbf{K}_{ji}^T\mathbf{x}_j\right)^T=\mathbf{x}_i^T(-2\mathbf{Q}^{ij})\mathbf{x}_j\end{cases}\Rightarrow\begin{cases}\mathbf{K}_{ii}=c_i\mathbf{Q}^{ij}\\ \mathbf{K}_{ij}=\mathbf{K}_{ji}^T=-\mathbf{Q}_{ij}\end{cases}\tag{12}$$

Consequently, by defining $\mathcal{G}_i=\{(i-1)k+1,(i-1)k+2,\dots,ik\}$ to be a set of sampling indices, in order for Eq. 10 to be true, the structure of $\mathbf{Q}$ is

$$\mathbf{Q}=\begin{cases}\sum\limits_{\forall j:(v_i,v_j)\in E}\mathbf{Q}^{ij}+\sum\limits_{\forall j:(v_j,v_i)\in E}\mathbf{Q}^{ji},\ \forall v_i\in V, & \text{at }(\mathcal{G}_i,\mathcal{G}_i)\\ -\mathbf{Q}^{ij},\ \forall i,j:(v_i,v_j)\in E, & \text{at }(\mathcal{G}_i,\mathcal{G}_j)\text{ and }(\mathcal{G}_j,\mathcal{G}_i)\\ 0, & \text{elsewhere}\end{cases}\tag{13}$$

## 3. Forward-Additive Gauss-Newton optimization

In Sec. 2.4 of the main paper, we present the inverse compositional algorithm for the optimization of the proposed model. Herein, we show the forward-additive case and prove that it is much slower than the onverse one. The general cost function to be optimized (same as Eq. 19 of the main paper) is

$$\arg\min_{\mathbf{p}} \|\mathcal{A}(\mathcal{S}(\bar{\mathbf{s}},\mathbf{p})) - \bar{\mathbf{a}}\|^2_{\mathbf{Q}^a} + \|\mathcal{S}(\bar{\mathbf{s}},\mathbf{p}) - \bar{\mathbf{s}}\|^2_{\mathbf{Q}^d} \tag{14}$$

By using an additive iterative update of the parameters as

$$\mathbf{p} \leftarrow \mathbf{p} + \Delta\mathbf{p}$$

and having an initial estimate of $\mathbf{p}$, the cost function of Eq. 14 is expressed as minimizing

$$\arg\min_{\Delta\mathbf{p}} \|\mathcal{A}(\mathcal{S}(\bar{\mathbf{s}},\mathbf{p}+\Delta\mathbf{p})) - \bar{\mathbf{a}}\|^2_{\mathbf{Q}^a} + \|\mathcal{S}(\mathbf{0},\mathbf{p}+\Delta\mathbf{p})\|^2_{\mathbf{Q}^s}$$

with respect to $\Delta\mathbf{p}$. In order to find the solution we need to linearize around $\mathbf{p}$, thus using first order Taylor series expansion at $\mathbf{p} + \Delta\mathbf{p} = \mathbf{p} \Rightarrow \Delta\mathbf{p} = \mathbf{0}$ as

$$\begin{cases} \mathcal{A}(\mathcal{S}(\bar{\mathbf{s}},\mathbf{p}+\Delta\mathbf{p})) \approx \mathcal{A}(\mathcal{S}(\bar{\mathbf{s}},\mathbf{p})) + \mathbf{J}_{\mathcal{A}}|_{\mathbf{p}=\mathbf{p}}\Delta\mathbf{p} \\ \mathcal{S}(\mathbf{0},\mathbf{p}+\Delta\mathbf{p}) \approx \mathcal{S}(\mathbf{0},\mathbf{p}) + \mathbf{J}_{\mathcal{S}}|_{\mathbf{p}=\mathbf{p}}\Delta\mathbf{p} \end{cases}$$

where $\mathbf{J}_{\mathcal{S}}|_{\mathbf{p}=\mathbf{p}} = \mathbf{J}_{\mathcal{S}}$ is the $2n \times n_{\mathcal{S}}$ shape Jacobian

$$\mathbf{J}_{\mathcal{S}} = \frac{\partial \mathcal{S}}{\partial \mathbf{p}} = \mathbf{U}$$

and $\mathbf{J}_{\mathcal{A}}|_{\mathbf{p}=\mathbf{p}} = \mathbf{J}_{\mathcal{A}}$ is the $mn \times n_{\mathcal{S}}$ appearance Jacobian

$$\mathbf{J}_{\mathcal{A}} = \nabla_{\mathcal{A}}\frac{\partial \mathcal{S}}{\partial \mathbf{p}} = \nabla_{\mathcal{A}}\mathbf{U} = \begin{bmatrix} \nabla\mathcal{F}(\mathcal{S}_1(\bar{\mathbf{s}},\mathbf{p}))\mathbf{U}_{1,2} \\ \nabla\mathcal{F}(\mathcal{S}_2(\bar{\mathbf{s}},\mathbf{p}))\mathbf{U}_{3,4} \\ \vdots \\ \nabla\mathcal{F}(\mathcal{S}_n(\bar{\mathbf{s}},\mathbf{p}))\mathbf{U}_{2i-1,2i} \end{bmatrix}$$

where $\mathbf{U}_{2i-1,2i}$ denotes the $2i-1$ and $2i$ row vectors of the basis $\mathbf{U}$. Note that we make an abuse of notation with $\nabla\mathcal{F}(\mathcal{S}_1(\bar{\mathbf{s}},\mathbf{p}))$ because $\mathcal{F}(\mathcal{S}_i(\bar{\mathbf{s}},\mathbf{p}))$ is a vector. However, it represents the gradient of a patch around landmark $i$ and it has size $m \times 2$. By substituting we get

$$\begin{aligned} &\arg\min_{\Delta\mathbf{p}} \|\mathcal{A}(\mathcal{S}(\bar{\mathbf{s}},\mathbf{p})) + \mathbf{J}_{\mathcal{A}}\Delta\mathbf{p} - \bar{\mathbf{a}}\|^2_{\mathbf{Q}^a} + \|\mathcal{S}(\mathbf{0},\mathbf{p}) + \mathbf{J}_{\mathcal{S}}\Delta\mathbf{p}\|^2_{\mathbf{Q}^s} = \\ =&\arg\min_{\Delta\mathbf{p}} \Big( [\mathcal{A}(\mathcal{S}(\bar{\mathbf{s}},\mathbf{p})) + \mathbf{J}_{\mathcal{A}}\Delta\mathbf{p} - \bar{\mathbf{a}}]^T \mathbf{Q}^a [\mathcal{A}(\mathcal{S}(\bar{\mathbf{s}},\mathbf{p})) + \mathbf{J}_{\mathcal{A}}\Delta\mathbf{p} - \bar{\mathbf{a}}] + [\mathcal{S}(\mathbf{0},\mathbf{p}) + \mathbf{J}_{\mathcal{S}}\Delta\mathbf{p}]^T \mathbf{Q}^s [\mathcal{S}(\mathbf{0},\mathbf{p}) + \mathbf{J}_{\mathcal{S}}\Delta\mathbf{p}] \Big) \end{aligned} \tag{15}$$

Taking the partial derivative with respect to $\Delta\mathbf{p}$ and solving for equality with $\mathbf{0}$ we get

$$\begin{aligned} &2{\mathbf{J}_{\mathcal{A}}}^T\mathbf{Q}^a(\mathcal{A}(\mathcal{S}(\bar{\mathbf{s}},\mathbf{p})) + \mathbf{J}_{\mathcal{A}}\Delta\mathbf{p} - \bar{\mathbf{a}}) + 2{\mathbf{J}_{\mathcal{S}}}^T\mathbf{Q}^s(\mathcal{S}(\mathbf{0},\mathbf{p}) + \mathbf{J}_{\mathcal{S}}\Delta\mathbf{p}) = \mathbf{0} \Rightarrow \\ \Rightarrow &2{\mathbf{J}_{\mathcal{A}}}^T\mathbf{Q}^a(\mathcal{A}(\mathcal{S}(\bar{\mathbf{s}},\mathbf{p})) - \bar{\mathbf{a}}) + 2{\mathbf{J}_{\mathcal{A}}}^T\mathbf{Q}^a\mathbf{J}_{\mathcal{A}}\Delta\mathbf{p} + 2{\mathbf{J}_{\mathcal{S}}}^T\mathbf{Q}^s\mathcal{S}(\mathbf{0},\mathbf{p}) + 2{\mathbf{J}_{\mathcal{S}}}^T\mathbf{Q}^s\mathbf{J}_{\mathcal{S}}\Delta\mathbf{p} = \mathbf{0} \Rightarrow \\ \Rightarrow &\Delta\mathbf{p} = -[{\mathbf{J}_{\mathcal{A}}}^T\mathbf{Q}^a\mathbf{J}_{\mathcal{A}} + {\mathbf{J}_{\mathcal{S}}}^T\mathbf{Q}^s\mathbf{J}_{\mathcal{S}}]^{-1}[{\mathbf{J}_{\mathcal{A}}}^T\mathbf{Q}^a(\mathcal{A}(\mathcal{S}(\bar{\mathbf{s}},\mathbf{p})) - \bar{\mathbf{a}}) + {\mathbf{J}_{\mathcal{S}}}^T\mathbf{Q}^s\mathcal{S}(\mathbf{0},\mathbf{p})] \end{aligned} \tag{16}$$

Thus by denoting as

$$\left.\begin{aligned} \mathbf{H}_{\mathcal{A}} &= {\mathbf{J}_{\mathcal{A}}}^T\mathbf{Q}^a\mathbf{J}_{\mathcal{A}} \\ \mathbf{H}_{\mathcal{S}} &= {\mathbf{J}_{\mathcal{S}}}^T\mathbf{Q}^s\mathbf{J}_{\mathcal{S}} = \mathbf{U}^T\mathbf{Q}^s\mathbf{U} \end{aligned}\right\} \Rightarrow \mathbf{H} = \mathbf{H}_{\mathcal{A}} + \mathbf{H}_{\mathcal{S}} \tag{17}$$

the combined $n_{\mathcal{S}} \times n_{\mathcal{S}}$ Hessian matrix and getting into account that ${\mathbf{J}_{\mathcal{S}}}^T\mathbf{Q}^s\mathcal{S}(\mathbf{0},\mathbf{p}) = \mathbf{U}^T\mathbf{Q}^s\mathbf{U}\mathbf{p} = \mathbf{H}_{\mathcal{S}}\mathbf{p}$ then the parameters increment is given by

$$\Delta\mathbf{p} = -\mathbf{H}^{-1}[{\mathbf{J}_{\mathcal{A}}}^T\mathbf{Q}^a(\mathcal{A}(\mathcal{S}(\bar{\mathbf{s}},\mathbf{p})) - \bar{\mathbf{a}}) + \mathbf{H}_{\mathcal{S}}\mathbf{p}] \tag{18}$$

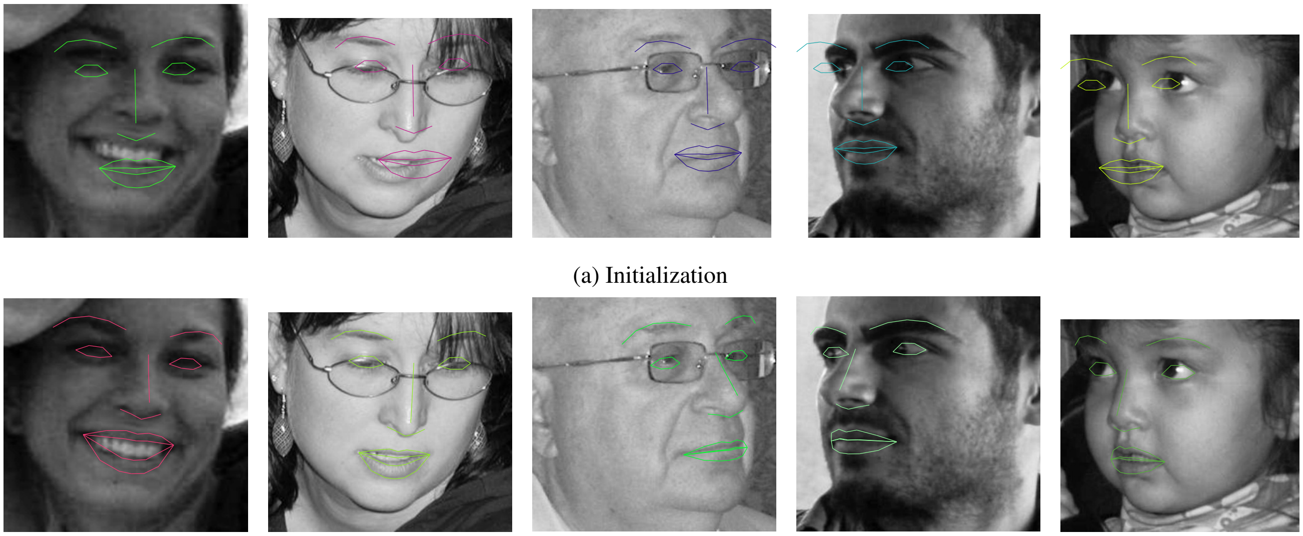
In Eq. 18, $\mathbf{H}_{\mathcal{S}}$ can be precomputed but $\mathbf{J}_{\mathcal{A}}$ and $\mathbf{H}^{-1}$ need to be computed at each iteration. Consequently, based on the costs of Tab. 1, the total computational cost is $\mathcal{O}(m^2n^2n_{\mathcal{S}} + mnn_{\mathcal{S}} + {n_{\mathcal{S}}}^3)$, which is much slower than the cost of the weighted inverse compositional algorithm with fixed Jacobian and Hessian shown in the main paper ($\mathcal{O}(mnn_{\mathcal{S}})$).

| $\mathbf{H}$ | $\mathbf{H}^{-1}$ | $\mathbf{J}_{\mathcal{A}}$ | $\mathbf{J}_{\mathcal{A}}{}^T\mathbf{\Sigma}^a$ | $\mathcal{S}(\bar{\mathbf{s}},\mathbf{p})$ | $\mathbf{H}_{\mathcal{S}}\mathbf{p}$ |
|---|---|---|---|---|---|
| $\mathcal{O}(m^2n^2n_S + mnn_S^2)$ | $\mathcal{O}(n_S{}^3)$ | $\mathcal{O}(mnn_S)$ | $\mathcal{O}(m^2n^2n_S)$ | $\mathcal{O}(2nn_S)$ | $\mathcal{O}(n_S{}^2)$ |

Table 1: The computational costs of all terms during the computation of the parameters increment. $n$ is the number of landmark points, $m$ is the length of the features' vector extracted from a patch and $n_S$ is the number of shape parameters.

## 4. Additional experiments

In this section we present additional experimental results for the proposed method. Figure 1 shows some indicative examples that correspond to the curve of Fig. 3c of the main paper. As also mentioned in the main paper, the testing database is Annotated Faces In-The-Wild (AFW) [8] and we use 49 points out of the 68-points markup of the annotations provided by the 300W competition [6, 7]. These results are indicative of the accuracy of APS.



(a) Initialization

(b) Final fitting

Figure 1: Fitting results on the AFW facial database. These are indicative results that correspond to the curve of Fig. 3c of the main paper.

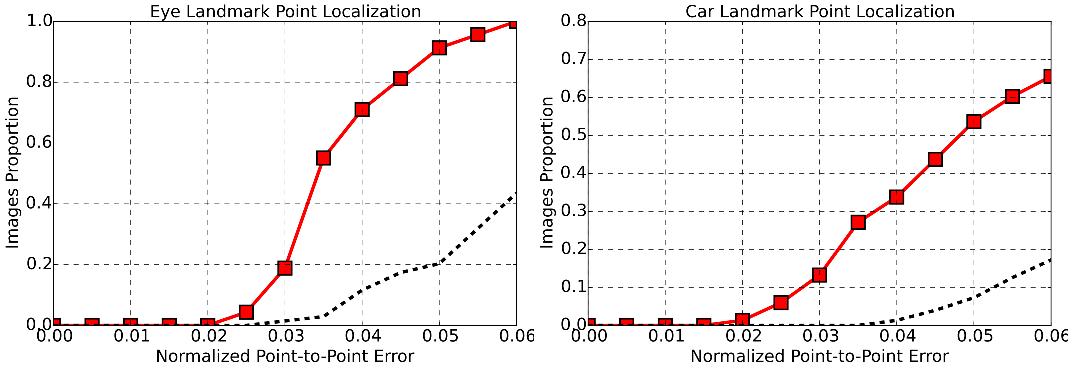
However, APS is a flexible patch-based deformable model that can also be applied to the landmark localization of other objects. Herein, we show indicative results for the case of eyes and cars. In the case of cars, we employ the sideview (view 2) images from CMU database [2, 5], which we split in 450 and 151 training and testing images repectively. For eyes, we use our internal annotated database that consists of 38 manually annotated landmarks and it has 600 and 400 training and testing images respectively. Figure 2 shows the cumulative fitting error curves for both objects. The initialization is performed as explained in Sec.3.1 of the main paper. Specifically, we add Gaussian noise to the global similarity transform retrieved from the ground truth annotations (without in-plane rotation) and applying it to the mean shape of the object. The standard deviation of the noise is set to 0.06.

Finally, Figs. 3 and 4 show some indicative fitting examples for both objects. Note that in the case of human eyes, most of the error is accumulated by the should be the upper and lower sclera, because it is a region without any distinctive features.

## References

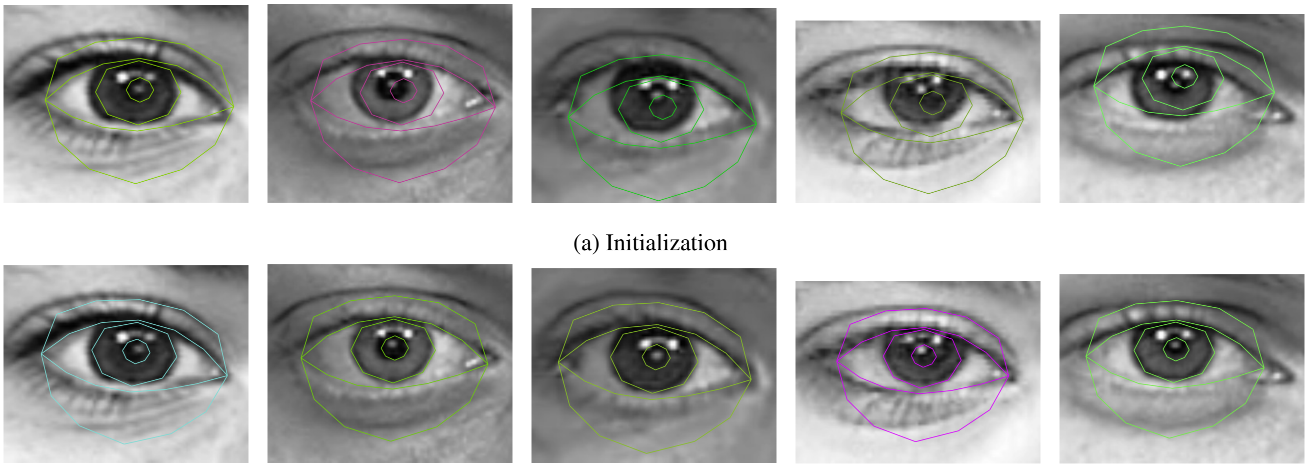
[1] J. Alabort-i Medina, E. Antonakos, J. Booth, P. Snape, and S. Zafeiriou. Menpo: A comprehensive platform for parametric image alignment and visual deformable models. In *Proceedings of the ACM International Conference on Multimedia*, MM '14, pages 679–682, New York, NY, USA, 2014. ACM.

[2] V. N. Boddeti, T. Kanade, and B. Kumar. Correlation filters for object alignment. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2291–2298, 2013.



(a) Eyes (b) Cars

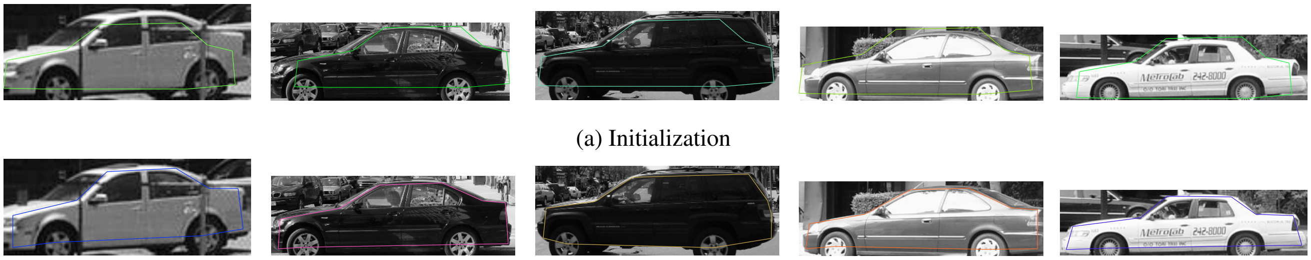
Figure 2: Fitting results for human eyes and cars.



(a) Initialization

(b) Final fitting

Figure 3: Fitting results on open eyes. These are indicative results that correspond to the curve of Fig. 2a.



(a) Initialization

(b) Final fitting

Figure 4: Fitting results on cars sideview. These are indicative results that correspond to the curve of Fig. 2b.

[3] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient matching of pictorial structures. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 66–73, 2000.

[4] P. F. Felzenszwalb and D. P. Huttenlocher. Pictorial structures for object recognition. *International Journal of Computer Vision (IJCV)*, 61(1):55–79, 2005.

[5] Y. Li, L. Gu, and T. Kanade. Robustly aligning a shape model and its application to car alignment of unknown pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 33(9):1860–1876, 2011.

[6] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic. 300 faces in-the-wild challenge: The first facial landmark localization challenge. In *Proceedings of IEEE International Conference on Computer Vision Workshopw (ICCV'W)*, pages 397–403, 2013.

[7] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic. A semi-automatic methodology for facial landmark annotation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 896–903, 2013.

[8] X. Zhu and D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2879–2886, 2012.