Imperial College London

Active Pictorial Structures

Epameinondas Antonakos, Joan Alabort-i-Medina, Stefanos Zafeiriou

IEEE 2015 Conference on Computer Vision and Pattern Recognition
CVPR 2015 BOSTON JUNE 7-12

Problem Statement

We propose **Active Pictorial Structures (APS)**, a novel generative deformable model for object alignment in-the-wild that combines the best of two worlds:

- **Pictorial Structures (PS)**

  Graph-based modeling using Gaussian Markov Random Field (GMRF).
- **Active Appearance Models (AAMs)**

  Weighted inverse-compositional Gauss-Newton with fixed Jacobian and Hessian.

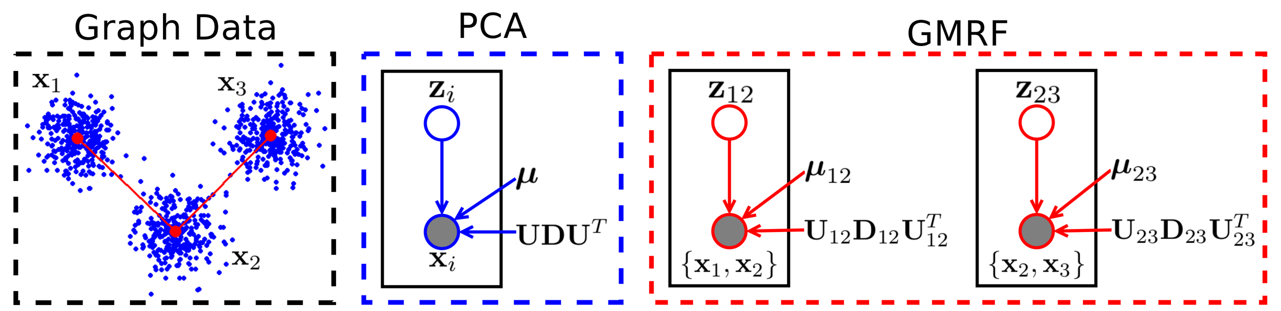
Contributions

- It is more **accurate** to model the appearance using **GMRF** rather than PCA.
- APS allow to define any graph structure; **not only tree**. We show that a tree-based shape model, as in PS, limits the model's descriptiveness and hampers the performance.
- We employ the spring-like shape model of PS as a **shape prior** in the Gauss-Newton optimization which makes the model more robust.
- We propose the best performing weighted inverse compositional Gauss-Newton algorithm with **fixed** Jacobian and Hessian. Its computational cost reduces to a single matrix multiplication per iteration and is independent of the employed graph structure.

Gaussian Markov Random Field (GMRF)

A GMRF is based on a graph $G = (V, E)$, where the vertexes stand for random variables and the edges impose statistical constraints on these random variables. It formulates the precision matrix $\mathbf{Q}$ of the data as a block-sparse matrix that has zeros at the blocks that correspond to disjoint vertexes, i.e.

$$\mathbf{Q}_{ij} = \mathbf{0}, \forall i, j : (v_i, v_j) \notin E$$



For open source code visit:
Menpo Project
http://www.menpo.org/

Active Pictorial Structures

APS consist of three GMRF-based models:

1) *Shape model:* SVD on the inverse of the precision matrix $\mathbf{Q}^s$ based on an undirected graph $G^s = (V^s, E^s)$.

   $\mathcal{S}(\bar{\mathbf{s}}, \mathbf{p}) = \bar{\mathbf{s}} + \mathbf{U}\mathbf{p}$ denotes a generated shape instance

2) *Part-based Apperance model:* Mean $\bar{\mathbf{a}}$ and precision matrix $\mathbf{Q}^a$ based on an undirected graph $G^a = (V^a, E^a)$.

   $\mathcal{A}(\mathcal{S}(\bar{\mathbf{s}}, \mathbf{p}))$: an appearance vector (concatenation of features from patches)

3) *Deformation prior:* Precision matrix $\mathbf{Q}^d$ based on a directed graph $G^d = (V^d, E^d)$.

Cost function consists of two Mahalanobis distances:

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

$$= \arg\min_{\mathbf{p}} \underbrace{[\mathcal{A}(\mathcal{S}(\bar{\mathbf{s}}, \mathbf{p})) - \bar{\mathbf{a}}]^T \mathbf{Q}^a [\mathcal{A}(\mathcal{S}(\bar{\mathbf{s}}, \mathbf{p})) - \bar{\mathbf{a}}]}_{\text{Part-based Appearance cost}} + \underbrace{[\mathcal{S}(\bar{\mathbf{s}}, \mathbf{p}) - \bar{\mathbf{s}}]^T \mathbf{Q}^d [\mathcal{S}(\bar{\mathbf{s}}, \mathbf{p}) - \bar{\mathbf{s}}]}_{\text{Deformation cost}}$$

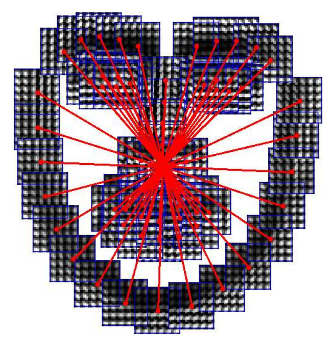
The $\mathbf{\Delta p}$ of the $\mathbf{p} \leftarrow \mathbf{p} - \mathbf{\Delta p}$ update at each iteration is a matrix multiplication (Hessian and Jacobian are fixed):

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

Pictorial Structures

PS learn a patch expert for the appearance of each part of an object and model its shape using spring-like connections between landmarks based on a tree structure.

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

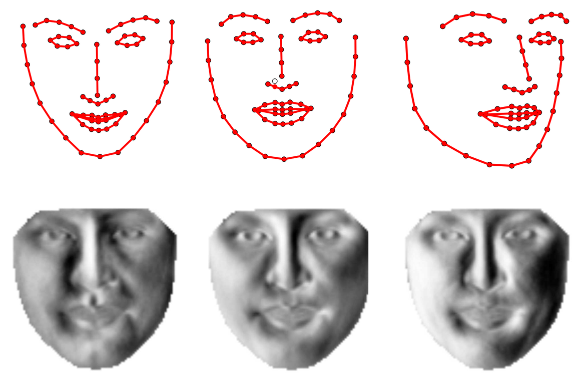


+ Globally optimal solution; no need for initialization.
− Graph can only be a tree.
− Appearance of each part is independent.
− The global optimum does not always correspond to the best shape in reality, because the tree structure restricts too much the range of possible realizable shape configurations.
− Very far from being real-time.

Active Appearance Models

AAMs are generative models that learn a parametric statistical shape and appearance model using PCA.

$$\arg\min_{\mathbf{p}, \boldsymbol{\lambda}} \|\mathcal{A}(\mathcal{W}(\bar{\mathbf{s}}, \mathbf{p})) - \bar{\mathbf{a}} - \mathbf{U}^a\boldsymbol{\lambda}\|^2$$

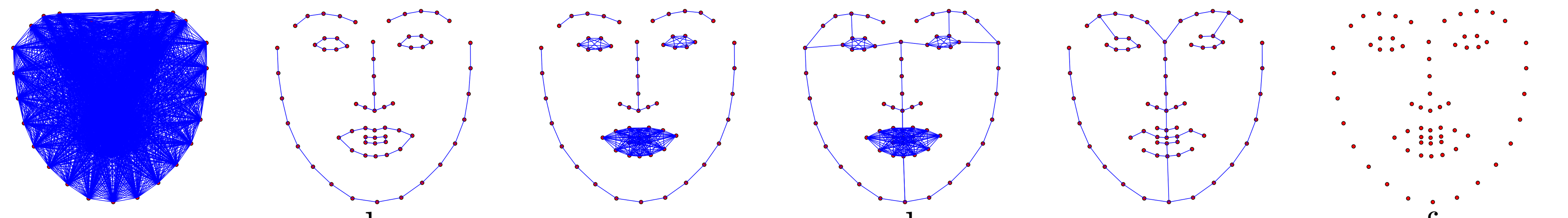


+ Accurate optimization using inverse compositional (IC) Gauss-Newton.
+ AAMs with holistic appearance and dense features are very accurate.
− Shape and appearance are modelled using PCA.
− Very sensitive to bad initializations.
− Alternating IC is very accurate but slow.
− Project-Out IC is very fast but inaccurate.

Advantages/Disadvantages of APS

+ The sums of the PS cost function become matrix multiplications at the APS cost function.
+ The deformation prior makes APS very robust to bad initializations.
+ Weighted IC with fixed Jacobian and Hessian is close to real-time.
+ The complexity is independent of the selected graph structure.
+ Many existing models can be derived from APS by changing the structure of $\mathbf{Q}^a$.
− $\mathbf{Q}^a$ is very large and requires a lot of memory.

Experiments



$\mathbf{G}^a$	*mean ± std*	*median*	≤ 0.04
(a)	0.0399 ± 0.0227	0.0324	68.3%
(b)	$\mathbf{0.0391} \pm \mathbf{0.0243}$	**0.0298**	**69.6%**
(c)	0.0506 ± 0.0371	0.0370	58.9%
(d)	0.0492 ± 0.0373	0.0354	58.9%
(e)	0.0413 ± 0.0257	0.0316	65.2%
(f)	0.0398 ± 0.0246	0.0319	66.5%
PCA	0.0716 ± 0.0454	0.0595	25.5%
Initialization	0.0800 ± 0.0280	0.0768	4.0%

$\mathbf{G}^s$	*mean ± std*	*median*	≤ 0.04
(a)	0.0495 ± 0.0273	0.0420	45.5%
(b)	0.0496 ± 0.0276	0.0438	45.5%
(c)	0.0503 ± 0.0262	0.0433	44.2%
(d)	0.0495 ± 0.0257	0.0434	44.6%
(e)	0.0519 ± 0.0306	0.0437	43.8%
(f)	0.0492 ± 0.0249	0.0437	42.9%
PCA	$\mathbf{0.0412} \pm \mathbf{0.0295}$	**0.0301**	**65.6%**
Initialization	0.0800 ± 0.0280	0.0768	4.0%

- SIFT, Training: LFPW trainset, Testing: LFPW testset + AFW, Initialization with DPM bboxes
- Comparison with other inverse compositional with fixed Jacobian and Hessian (POIC):

  DPM-AAM-POIC: G. Tzimiropoulos and M. Pantic, "Gauss-Newton deformable part models for face alignment in-the-wild", CVPR, 2014.

  BAAM-POIC: J. Alabort-i-Medina and S. Zafeiriou, "Bayesian Active Appearance Models", CVPR 2014.

  AAM-POIC: E. Antonakos, et al., "Feature-based lucas-kanade and active appearance models", IEEE TIP, 2015.
- Comparison with state-of-the-art:

  GN-DPM: G. Tzimiropoulos and M. Pantic, "Gauss-Newton deformable part models for face alignment in-the-wild", CVPR, 2014.

  DPM/PS: X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in-the-wild", CVPR 2012.

  SIFT-AAM: E. Antonakos, et. al, "Feature-based lucas-kanade and active appearance models", IEEE TIP, 2015.

  SDM: X. Xiong and F.De la Torre, "Supervised descent method and its applications to face alignment", CVPR, 2013.