Supplementary Material
Estimating Correspondences of Deformable Objects "In-the-wild"

Yuxiang Zhou* Epameinondas Antonakos* Joan Alabort-i-Medina* Anastasios Roussos*
Stefanos Zafeiriou*,†
*Department of Computing, Imperial College London, U.K.
†Center for Machine Vision and Signal Analysis, University of Oulu, Finland
{yuxiang.zhou10, e.antonakos, ja310, troussos, s.zafeiriou}@imperial.ac.uk

Introduction

In this supplementary material, we provide additional algorithmic details for our shape flow estimation, as well as additional visualisations and evaluations for the dense and patch-based AAMs that were constructed using the proposed framework.

A. Implementation of Shape Flow Estimation

As mentioned in the main paper (Section 3, Step 3), we propose to estimate the shape flow by minimising the following energy:

$$E_{sf} = \alpha \int_{\Omega} \sum_{n=1}^{N_t} \|\boldsymbol{d}(\boldsymbol{x} + \boldsymbol{u}_n(x); n) - \boldsymbol{d}(x; 0)\| \mathrm{d}\boldsymbol{x} \quad (1)$$

$$+ \beta \int_{\Omega} \sum_{n=1}^{N_t} \|\boldsymbol{u}_n(\boldsymbol{x}) - \sum_{i=1}^{R} \boldsymbol{q}_i(n) \boldsymbol{v}_i(\boldsymbol{x})\|^2 \mathrm{d}\boldsymbol{x} \quad (2)$$

$$+ \int_{\Omega} \sum_{i=1}^{R} \|\nabla \boldsymbol{v}_i(\boldsymbol{x})\| \, \mathrm{d}\boldsymbol{x} \quad (3)$$

We minimise this energy jointly with respect to $\boldsymbol{u}_n(\boldsymbol{x})$ and $\boldsymbol{v}_i(\boldsymbol{x})$, which correspond to the two sets of unknown shape flows. We implement this minimisation based on the optimisation algorithm described in [5] and the relevant publicly available code [1]. However, we modify this algorithm so that, instead of initialising the coarse-to-fine and warping iterations with a zero flow, we use Thin Plate Splines (TPS) [2] interpolation of the initial correspondence vectors described in Section 3, Step 2 of the main paper. This yields a significantly better initial location of the highly-nonconvex objective function and improves the computational efficiency, since much less coarse-to-fine pyramids are needed.

Note that in every coarse-to-fine and warping iteration, we use an initialisation that comes from the previous iteration. We approximate the data term (1) by linearising the SVS images $\boldsymbol{d}(\boldsymbol{x}; n)$ around the initialisation. After that, the energy becomes convex and we optimise it by employing alternating optimisation with respect to $\boldsymbol{v}_i(\boldsymbol{x})$ and $\boldsymbol{u}_n(\boldsymbol{x})$. The minimisation with respect to $\boldsymbol{v}_i(\boldsymbol{x})$ is decoupled for every coefficient i and corresponds to Rudin-Osher-Fatemi Total Variation denoising [7], which we solve efficiently by applying the first order primal-dual algorithm of [3]. The minimisation with respect to $\boldsymbol{u}_n(x)$ is decoupled for every pixel $\boldsymbol{x}$ and every shape index i. This minimisation is also implemented by applying the efficient primal-dual algorithm of [3].

[1] https://bitbucket.org/troussos/mfsf/downloads

B. Dense Active Appearance Models

In this section, we report additional qualitative and quantitative evaluations for the Dense Active Appearance Models (dAAMs) of faces and ears that were constructed using the proposed framework.

B.1. Principal Components and Compactness

Figure 1 visualises the first five shape and appearance principal components of ear and face dAAMs. We observe that in both ear and face cases, the variation of both shape and appearance captured by the model seem plausible.

Figure 2 plots the variance ration of the face dAAM, which provides an indication of the compactness of the model. The compactness is compared with the one of a standard sparse AAM built on the same data. Note that these are the two shape models that are compared in Figure 9-left of the main paper. We observe that our dAAM is significantly more compact than the sparse AAM, since for any given number of components, it manages to explain a larger portion of the corresponding total variance of the training set.

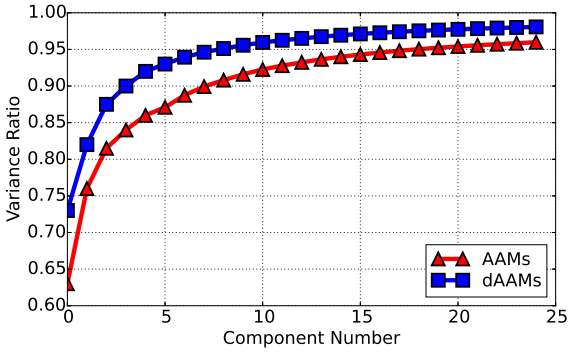

Figure 2: Compactness plots of dAAM (blue) and sparse AAM (red) models for faces. Portion of the corresponding total variance explained as a function of the number of retained principal components.

B.2. Dense Shape Reconstruction Ability

Figure 3 evaluates the dense shape reconstruction ability of the proposed dAAMs and compares it with that of standard sparse AAMs. Specifically, we use shapes with dense ground-truth annotations and reconstruct them with both AAMs and dAAMs, by projecting on the corresponding model subspace. In the case of AAMs, which only contain a sparse shape model, we densify it using a piecewise affine transformation, which is typically for texture warping of these models. We observe that dAAMs significantly outperform classic AAMs, in terms of dense shape reconstruction accuracy.

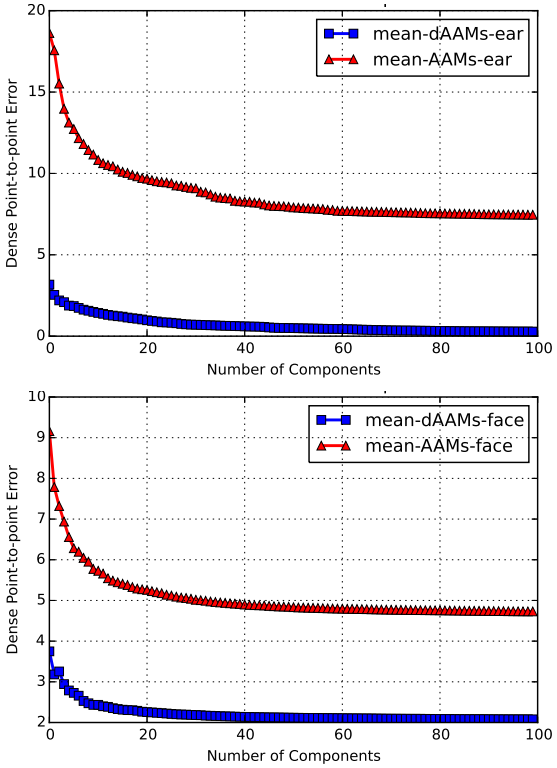

Figure 3: Dense shape reconstruction errors for ears (top) and faces (bottom), using AAMs (red) and dAAMs (blue). The average normalized dense point-to-point distance error is plotted as a function of the number of principal components of the model.

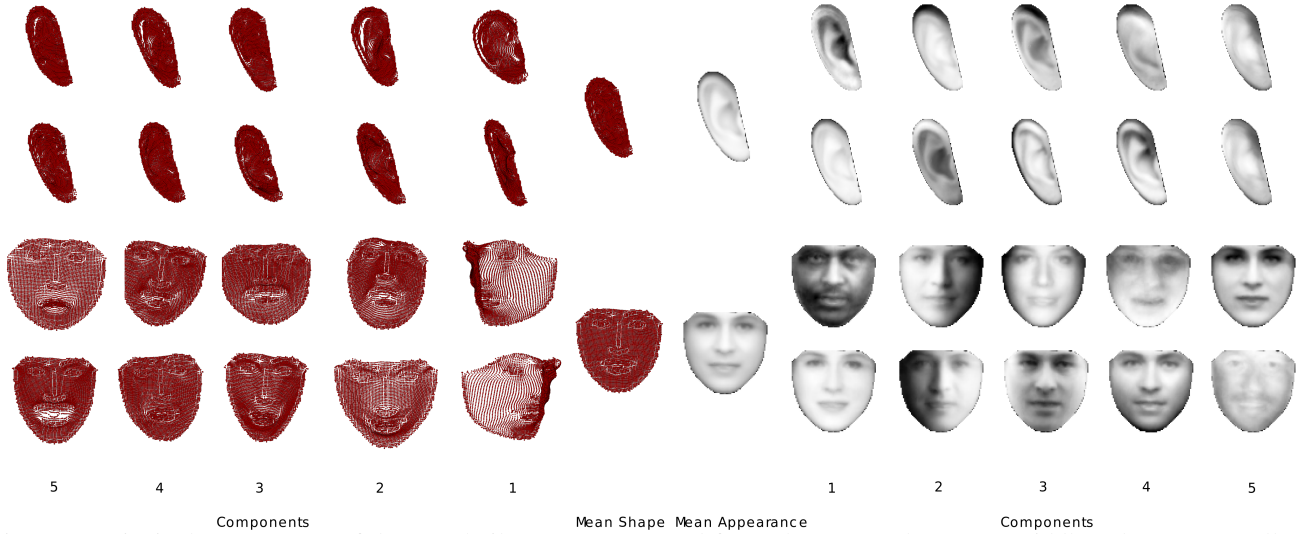

Figure 1: Principal components of dAAMs built on ears (top) and faces (bottom). The mean (middle columns) as well as the first five principal components are visualised for both shape (left) and appearance (right). ± 3 times the variance of the corresponding component is used in each case.

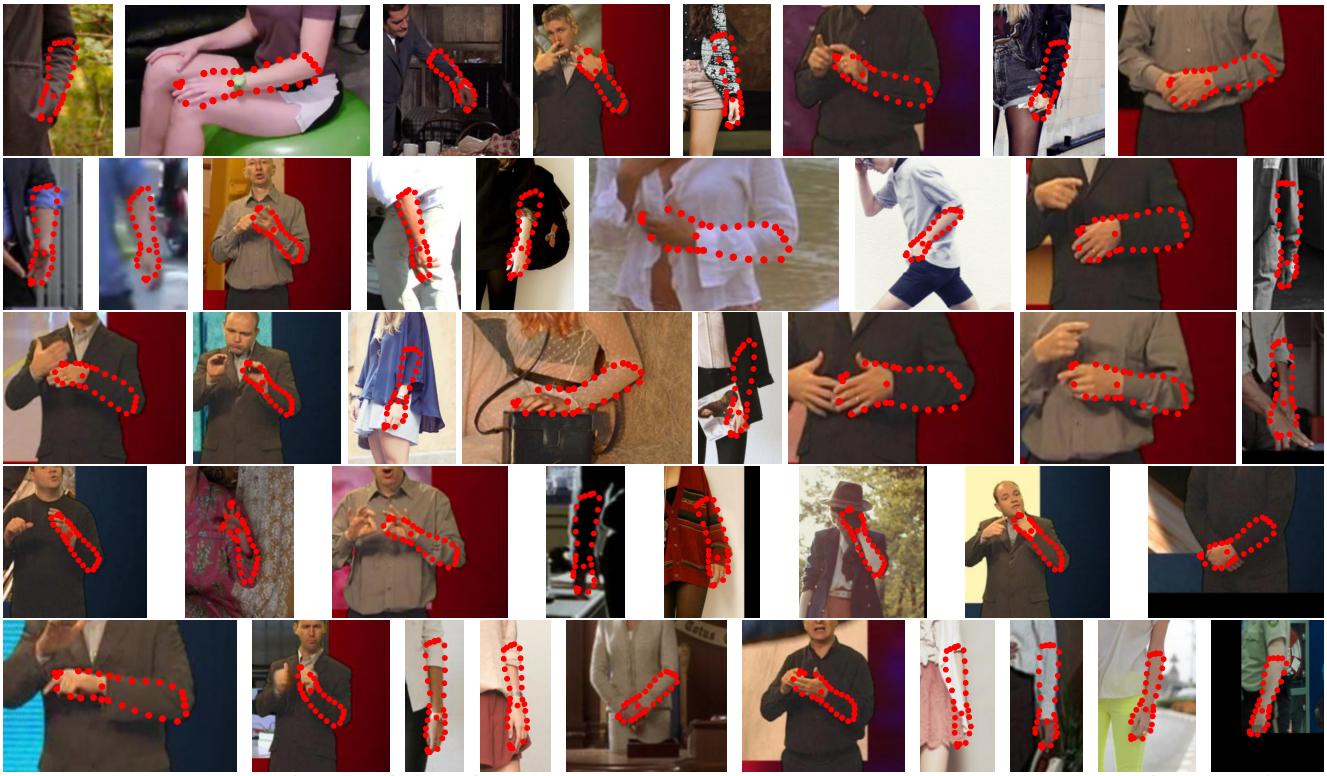

Figure 4: Demonstration of outline fitting of patch-based AAM on arms. Images are cropped to arms only for better visualization.

B.3. Dense Fitting Visualizations

In this section, we visualize some characteristic examples of fitting dAAMs that were built using the proposed framework. These results are characteristic examples that come from the quantitative evaluations reported in Section 4.1 ("Non-rigid object alignment in-the-wild") of the main paper. Figure 5 shows dAAM fitting results using a grid visualisation, for both faces and ears. We observe that the proposed method successfully captures the shape deformations of these object classes and provides a detailed shape estimation for a variety of input images.

Figure 5: Examples of fitting dAAMs that are constructed with the proposed pipeline. Results of dense fitting on images of ears (first two rows) and faces (last two rows). A grid visualisation is used.

C. Patch-based Active Appearance Models

In this section, we present additional visualisations and evaluations for the patch-based Active Appearance Model (PAAM) of arms that was constructed using the proposed framework.

C.1. Subsampling of the Outline from Dense Correspondences

As mentioned in the main paper (Section 3 - Step 4), in order to train a PAAM, we subsample the densified training shapes to only consider points on the object's outline. Some examples of this procedure are depicted in Figure 6. We manually annotate sparse outline points only on the ref-

	Wrist			Elbow		
Method	*mean*	*std*	$\leq$ 6*pt*	*mean*	*std*	$\leq$ 6*pt*
Buehler	12.08	19.94	44.5%	12.94	14.65	34.4%
Charles14	11.81	20.89	54.2%	8.30	11.00	**55.2%**
Charles13	13.78	22.39	43.3%	13.17	18.74	46.3%
Pfister14	14.69	17.89	29.7%	14.60	10.59	14.0%
Ramanan	15.59	19.04	22.6%	15.53	10.82	15.8%
Pfister15	7.62	11.04	54.1%	8.84	11.44	54.9%
Ours	**6.71**	**10.90**	**63.1%**	**8.20**	**10.54**	52.1%

Table 1: Fitting statistics on BBC Pose database for experiment 4.2 in main paper.

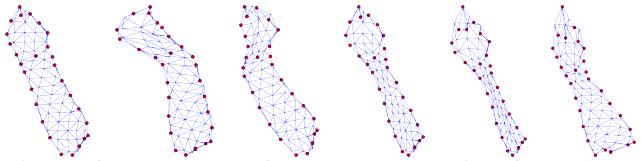

Figure 6: Examples of sparse subsampling from dense shapes of arms. The dense correspondences (visualized via a deforming grid) are established using our shape flow estimation. The sparse landmarks (red dots) on the object outline are manually annotated only on the reference shape (left most image). In all other 5 example shapes, these landmarks have been automatically "propagated" using the established dense correspondences.

erence shape. Then all other training shapes are subsampled automatically exploiting the dense correspondences that are established with our shape flow estimation. We observe that the automatic subsampling seems plausible, which is attributed to the accurate estimations of dense correspondences.

C.2. Principal Components

Figure 7 shows the mean shape and the first four principal shape components of our PAAM for arms[2]. We observe that the shape variations captured by the model are plausible and seem to produce valid shapes of human arms.

C.3. Fitting Results

Figure 4 demonstrates more fitting results produced by fitting patch-base AAM on arms using MPII [1], Fashion Pose [4], FLIC [8] and BBC Pose [6] databases. All fittings are initialised using the same method as mentioned in section 4.2 of the main paper.

In addition table 1 reports statistical measures that provide additional information to section 4.2 of the main paper. Column $\leq$ 6*pt* reports the percentage of fittings that achieved a point-to-point normalised error less than 6 pixels (same measure used in [6]). This shows that we have

[2]Note that we do not visualise the appearance variation, since this is built using SIFT features and the corresponding 36-channel feature space cannot be visualized in an intuitive way.

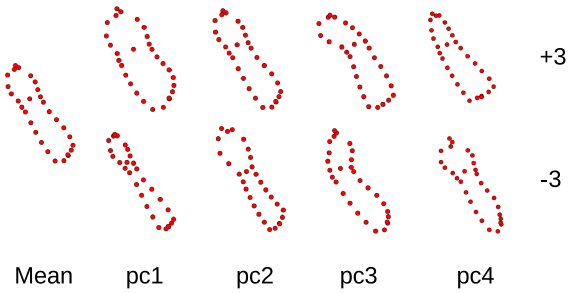

Figure 7: Principal components of our patch-based AAM for human arms. The mean (left most column) as well as the first four principal components are visualised. ± 3 times the variance of the corresponding component is used in each case.

notable improvement on estimating wrists and comparable results on estimating the elbow.

References

[1] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele. 2d human pose estimation: New benchmark and state of the art analysis. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014. 4

[2] F. J. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 11(6):567–585, 1989. 1

[3] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *JMIV*, 2011. 1

[4] M. Dantone, J. Gall, C. Leistner, and L. Van Gool. Human pose estimation using body parts dependent joint regressors. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3041–3048. IEEE, 2013. 4

[5] R. Garg, A. Roussos, and L. Agapito. A variational approach to video registration with subspace constraints. *International Journal of Computer Vision*, 104(3):286–314, 2013. 1

[6] T. Pfister, J. Charles, and A. Zisserman. Flowing convnets for human pose estimation in videos. *arXiv preprint arXiv:1506.02897*, 2015. 4

[7] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992. 1

[8] B. Sapp and B. Taskar. Modec: Multimodal decomposable models for human pose estimation. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3674–3681. IEEE, 2013. 4