Unifying Holistic and Parts-Based Deformable Model Fitting - Supplementary Material -

Joan Alabort-i-Medina Stefanos Zafeiriou
Department of Computing, Imperial College London, United Kingdom
{ja310, s.zafeiriou}@imperial.ac.uk

1. Additional Experiments

This supplementary material extends the experimental section of our main paper with two additional experiments. The first experiment reports the fitting accuracy of the 5 algorithms described in our paper (PIC and AIC for AAMs, RLMS for CLMs and PIC-RLMS and AIC-RLMS for our Unified approach) given different amounts of training data. The second experiments reports the accuracy of the same five algorithms on the challenging problem of deformable face tracking in-the-wild.

1.1. Accuracy vs amount of training data

In this experiment, we compare the accuracy of the two Unified fitting algorithms (PIC-RLMS and AIC-RLMS) against the accuracy of existing AAMs (PIC and AIC) and CLMs (RLMS) fitting strategies when all models are trained using different amounts of training data. We combine the training images of the LFPW [2] and Helen [4] datasets with half of the images from the AFW [5] and iBUG [1] datasets (66 points ground truth landmark annotations were again provided by the iBUG group[1]) to create a training dataset of approximately 3000 training images. We train all models using 64, 128, 256, 512, 1024 and all training images and report their respective fitting accuracy on a test dataset containing the test images of the LFPW and Helen datasets and the remaining halves of the AFW and iBUG datasets.

Results for this experiment are provided in Figure 1. The proposed AICRLMS algorithm is consistently the most accurate algorithm. The difference in accuracy is small (but noticeable) with respect to PIC-RLMS and AIC for small amount of training data (up to 256). For medium and large amounts of training data (from 512) AIC-RLMS substantially outperforms all other methods. Note that AIC-RLMS (and to a certain extend PIC-RLMS) is also the only method for which fitting accuracy always improves given greater amounts of training data (the accuracy of AIC picks at 1024 training images and decreases afterwards while the accuracy of RLMS remains constant after 1024 training images).

[1] http://ibug.doc.ic.ac.uk/resources/300-W/

1.2. Deformable face tracking in-the-wild

This experiments compares the accuracy of the two Unified approaches (PIC-RLMS and AIC-RLMS) against the accuracy of AAMs (PIC and AIC) and CLMs (RLMS) in the challenging problem of deformable face tracking in-the-wild. To this end, we manually annotated 5 sequences of the Youtube Celebrities dataset [3] (approximately 950 frames) with the 66 points mark-up scheme used in the previous experiments. The Youtube Celebrities database is an in-the-wild face dataset that contains videos of celebrities appearing on different TV shows. The length of most of the sequences is less than 3 seconds. For each method, we use the most accurate model in Experiment 1.1. For all algorithms the first frame of each sequence is initialized by aligning the mean shape with the ground truth shape; subsequent frames use the tracking result obtained from the previous frames as their initialization.

Quantitative results for this experiments are provided in Figure 2 (tracking statistics are only provided for those methods capable of tracking whole sequences without reinitialization.). Visual comparisons between the results obtained by each method, on five selected frames of each tracking sequence, are shown in Figures 3, 4, 5, 6 and 7. Figure 2 shows how AIC-RLMS and RMLS are the only two methods capable to successfully tracking all five sequences without reinitialization. Note, however, that AIC-RLMS is much more accurate than RMLS for all sequences. Overall, AIC-RLMS is the most accurate and robust method (notice that PIC-RLMS is slightly more accurate for Sylvester Stallone's sequence).

References

[1] http://ibug.doc.ic.ac.uk/resources/300-W/. 1

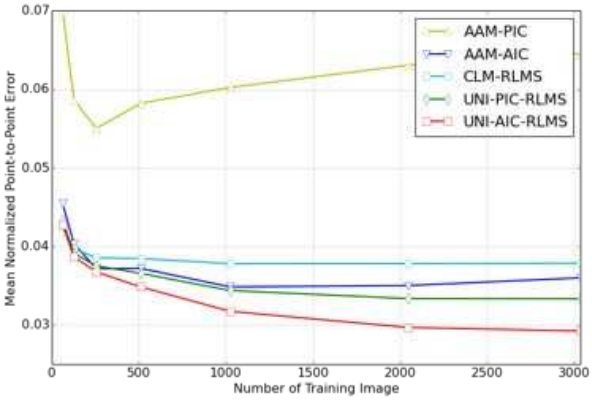
[2] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar. Localizing parts of faces using a consensus of exemplars. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011. 1

[3] M. Kim, S. Kumar, V. Pavlovic, and H. A. Rowley. Face tracking and recognition with visual constraints in real-world videos. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008. 1

[4] V. Le, B. Jonathan, Z. Lin, L. Boudev, and T. S. Huang. Interactive facial feature localization. In *European Conference on Computer Vision (ECCV)*, 2012. 1

[5] X. Zhu and D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012. 1

978-1-4673-6964-0/15/$31.00 ©2015 IEEE



(a) All methods

# Training images	Mean	Std	Median	Conv
Initialization	0.0973	0.0324	0.0932	-
64	0.0700	0.0634	0.0458	0.79
128	0.0587	0.0544	0.0386	0.85
256	**0.0550**	**0.0493**	**0.0364**	**0.88**
512	0.0682	0.0554	0.0372	0.86
1024	0.0602	0.0599	0.0370	0.84
2048	0.0631	0.0622	0.0386	0.83
3036	0.0645	0.0625	0.0374	0.81

(b) AAM-PIC

# Training images	Mean	Std	Median	Conv
Initialization	0.0973	0.0325	0.0932	-
64	0.0454	0.0351	0.0330	0.94
128	0.0403	0.0316	0.0299	0.96
256	0.0371	0.0279	0.0282	0.98
512	0.0372	0.0302	0.0276	0.98
1024	**0.0348**	**0.0276**	0.0264	**0.99**
2048	0.0350	0.0298	**0.0258**	0.97
3036	0.0360	0.0317	0.0259	0.97

(c) AAM-AIC

# Training images	Mean	Std	Median	Conv
Initialization	0.0973	0.0325	0.0932	-
64	0.0425	0.0214	0.0363	**0.99**
128	0.0397	0.0208	0.0338	**0.99**
256	0.0385	0.0202	0.0325	**0.99**
512	0.0384	**0.0206**	0.0325	**0.99**
1024	**0.0378**	**0.0206**	0.0320	**0.99**
2048	**0.0378**	0.0208	0.0318	**0.99**
3036	**0.0378**	0.0207	**0.0316**	**0.99**

(d) CLM-RLMS

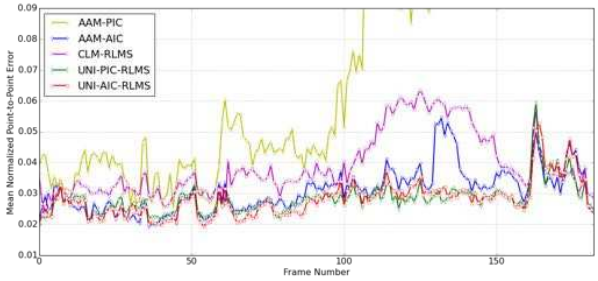
# Training images	Mean	Std	Median	Conv
Initialization	0.0973	0.0325	0.0932	-
64	0.0432	0.0237	0.0357	**0.99**
128	0.0391	0.0214	0.0330	**0.99**
256	0.0375	0.0214	0.0310	**0.99**
512	0.0365	0.0214	0.0302	**0.99**
1024	0.0344	**0.0212**	0.0381	**0.99**
2048	**0.0333**	0.0227	0.0266	**0.99**
3036	**0.0333**	0.0246	**0.0261**	**0.99**

(e) UNI-PIC-RLMS

# Training images	Mean	Std	Median	Conv
Initialization	0.0973	0.0325	0.0932	-
64	0.0427	0.0236	0.0351	**0.99**
128	0.0387	0.0214	0.0324	**0.99**
256	0.0367	0.0213	0.0301	**0.99**
512	0.0348	0.0208	0.0287	**0.99**
1024	0.0317	0.0198	0.0259	**0.99**
2048	0.0297	**0.0197**	0.0243	**0.99**
3036	**0.0292**	0.0202	**0.0237**	**0.99**

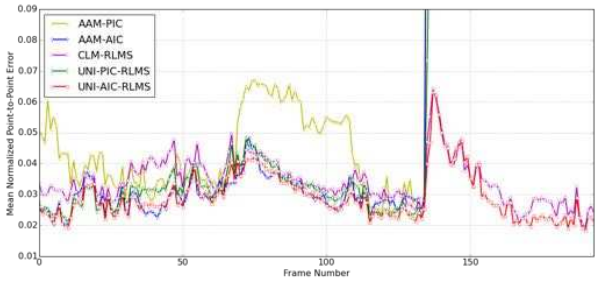
(f) UNI-AIC-RLMS

Figure 1: Mean normalized point to point error graph and fitting statistics for all methods given different amounts of training data. Note that the column "Conv" refers to the percentage of images for which the final error was smaller than the initial one.



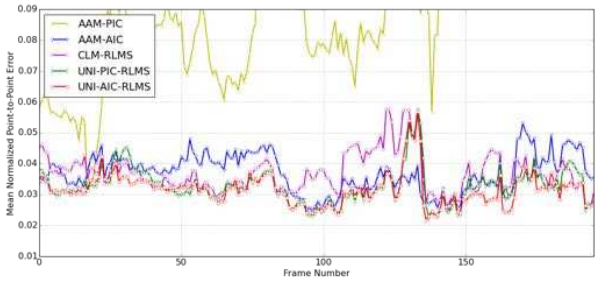
Algorithm	Mean	Std	Median	Lost
AAM-PIC	-	-	-	Yes
AAM-AIC	0.0308	0.0072	0.0300	No
CLM-RLMS	0.0390	0.0101	0.0355	No
UNI-PIC-RLMS	0.0282	**0.0050**	0.0278	No
UNI-AIC-RLMS	**0.0281**	0.0058	**0.0277**	No

(a) Angelina Jolie's first sequence.



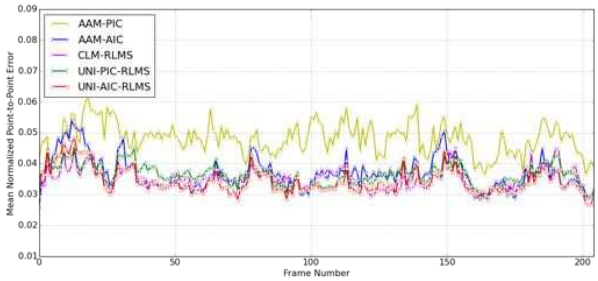
Algorithm	Mean	Std	Median	Lost
AAM-PIC	-	-	-	Yes
AAM-AIC	-	-	-	Yes
CLM-RLMS	0.0343	0.0074	0.0330	No
UNI-PIC-RLMS	-	-	-	Yes
UNI-AIC-RLMS	**0.0292**	**0.0078**	**0.0274**	No

(b) Angelina Jolie's second sequence.



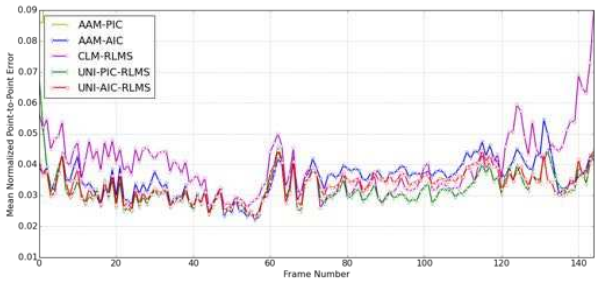
Algorithm	Mean	Std	Median	Lost
AAM-PIC	-	-	-	Yes
AAM-AIC	0.0369	0.0062	0.0370	No
CLM-RLMS	0.0370	0.0065	0.0360	No
UNI-PIC-RLMS	0.0328	0.0057	0.0320	No
UNI-AIC-RLMS	**0.0317**	**0.0052**	**0.0317**	No

(c) Adam Sandler's sequence.



Algorithm	Mean	Std	Median	Lost
AAM-PIC	0.0481	0.0051	0.0484	No
AAM-AIC	0.0367	0.0053	0.0357	No
CLM-RLMS	0.0352	0.0034	0.0350	No
UNI-PIC-RLMS	0.0364	**0.0036**	0.0363	No
UNI-AIC-RLMS	**0.0344**	0.0041	**0.0332**	No

(d) Bruce Willis' sequence.



Algorithm	Mean	Std	Median	Lost
AAM-PIC	-	-	-	Yes
AAM-AIC	0.0354	0.0058	0.0364	No
CLM-RLMS	0.0398	0.0099	0.0383	No
UNI-PIC-RLMS	**0.0322**	0.0059	**0.0312**	No
UNI-AIC-RLMS	0.0334	**0.0048**	0.0341	No

(e) Sylvester Stallone's sequence.

Figure 2: Mean normalized point to point error graphs and tracking statistics for all methods

(a) AAM-PIC

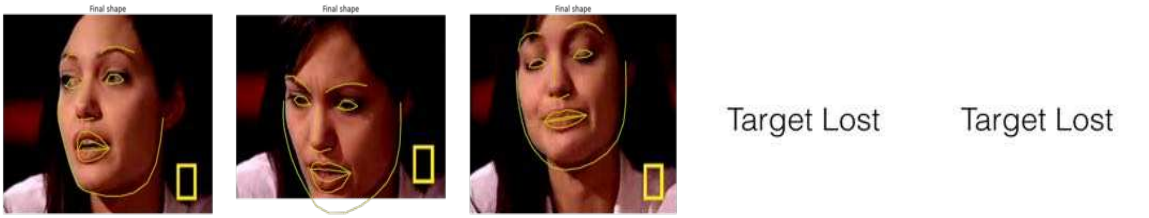
(b) AAM-AIC

(c) CLM-RLMS

(d) UNI-PIC-RLMS

(e) UNI-AIC-RLMS

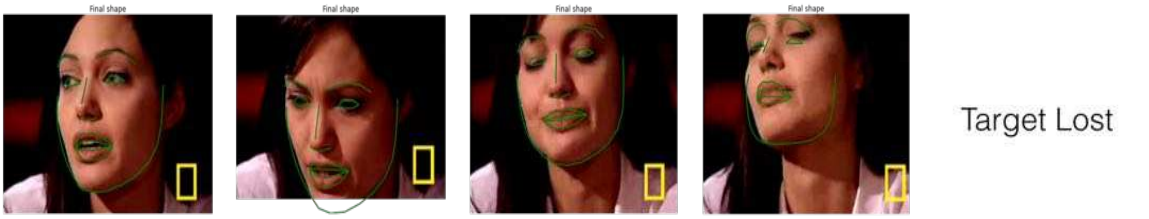
Figure 3: Selected fitted frames from Angelina Jolie's first sequence for all methods.



(a) AAM-PIC

(b) AAM-AIC

(c) CLM-RLMS



(d) UNI-PIC-RLMS

(e) UNI-AIC-RLMS

Figure 4: Selected fitted frames from Angelina Jolie's second sequence for all methods.

(a) AAM-PIC

(b) AAM-AIC

(c) CLM-RLMS

(d) UNI-PIC-RLMS

(e) UNI-AIC-RLMS

Figure 5: Selected fitted frames from Adam Sandler's sequence for all methods.

(a) AAM-PIC

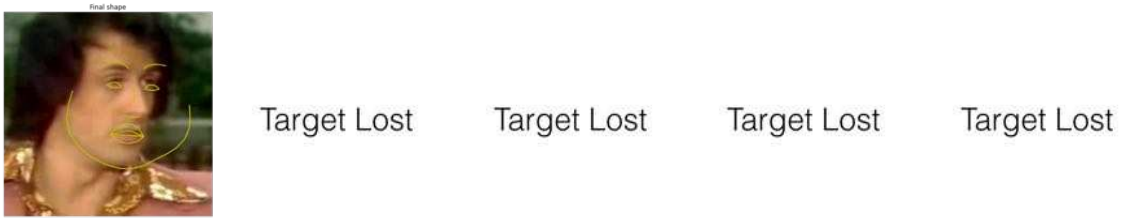
(b) AAM-AIC

(c) CLM-RLMS

(d) UNI-PIC-RLMS

(e) UNI-AIC-RLMS

Figure 6: Selected fitted frames from Bruce Willis' sequence for all methods.



(a) AAM-PIC

(b) AAM-AIC

(c) CLM-RLMS

(d) UNI-PIC-RLMS

(e) UNI-AIC-RLMS

Figure 7: Selected fitted frames from Sylvester Stallone's sequence for all methods.