22nd Iberoamerican Congress on Pattern Recognition

Benchmarking Head Pose Estimation In-The-Wild

Elvira Amador, Roberto Valle
Dep. Inteligencia Artificial
Univ. Politécnica de Madrid
Boadilla del Monte, Spain
rvalle@fi.upm.es

José Miguel Buenaposada
Dep. Ciencias Computación
Univ. Rey Juan Carlos
Móstoles, Spain
josemiguel.buenaposada@urjc.es

Luis Baumela
Dep. Inteligencia Artificial
Univ. Politécnica de Madrid
Boadilla del Monte, Spain
lbaumela@fi.upm.es

PCR http://www.dia.fi.upm.es/~pcr/research.html

Abstract

What? We review the problem of estimating head pose by regressing the yaw, pitch and roll head angles from images acquired "in-the-wild".

Why? Key preprocessing step for several tasks such as facial attributes estimation, human machine interaction, focus of attention, gaze, etc.

Contributions

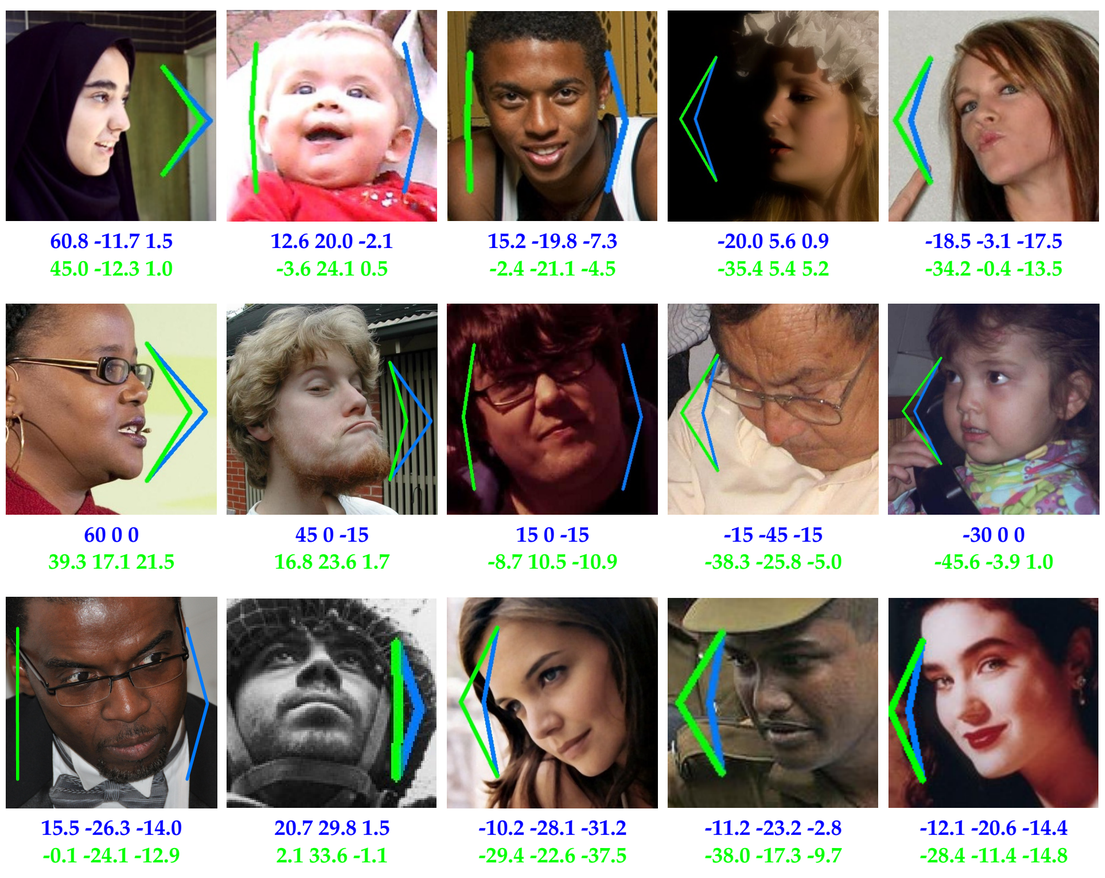
- A brief survey of the best head pose estimation algorithms.
- Definition of an evaluation methodology and publicly available benchmark to precisely compare the performance of head pose estimation algorithms.
- The establishment of the state-of-the-art on this benchmark.

Benchmarking head pose

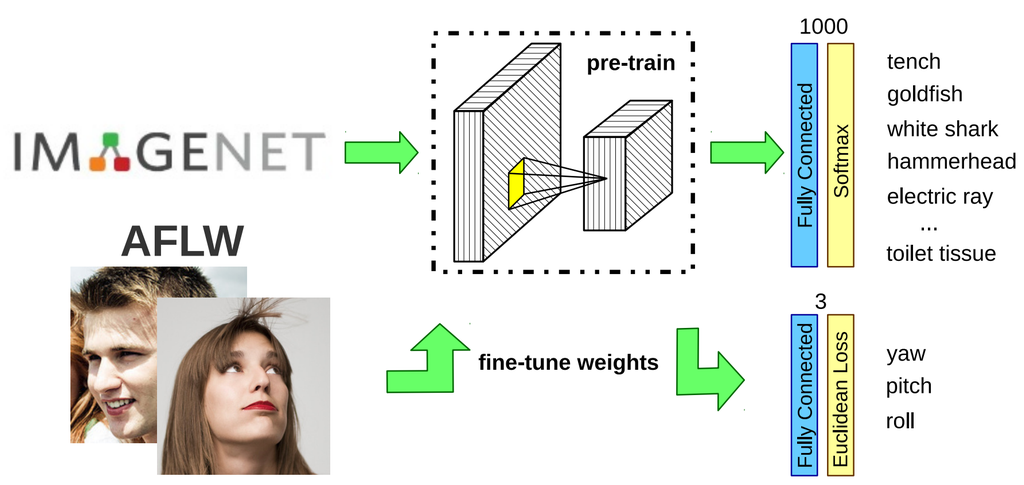
To have comparable results all algorithms should use the same train, validation and test data-sets publicly available, hence it is impossible to make a fair comparison among any of these approaches.

Method	AFLW (MAE) yaw	pitch	roll	AFW yaw	300W (MAE) yaw	pitch	roll
Peng *et al*. [5]	-	-	-	86.3%	-	-	-
Valle *et al*. [9]	12.26°	-	-	83.54%	-	-	-
Gao *et al*. [1]	6.60°	5.75°	-	-	-	-	-
Yang *et al*. [10]	-	-	-	-	4.20°	5.19°	2.42°
Ranjan *et al*. [6]	7.61°	6.13°	3.92°	97.7%	-	-	-
Kumar *et al*. [4]	6.45°	5.85°	8.75°	96.67%	-	-	-

Results



Transfer learning methodology



Experiments in-the-wild

We provide regressors based on AlexNet, VGG, GoogLeNet and ResNet.

Method	AFLW (MAE) yaw	pitch	roll	AFW yaw	300W (MAE) yaw	pitch	roll
AlexNet [3]	6.28°	5.02°	3.36°	86.32%	6.86°	6.61°	5.82°
GoogLeNet [8]	6.40°	5.31°	3.74°	**95.51%**	5.71°	7.99°	6.85°
VGG-16 [7]	6.23°	4.96°	3.35°	85.68%	6.35°	7.02°	5.98°
VGG-19 [7]	5.78°	4.79°	3.20°	94.23%	5.56°	6.35°	4.65°
ResNet-50 [2]	6.00°	4.90°	3.14°	94.44%	5.71°	5.91°	3.23°
ResNet-101 [2]	**5.59°**	**4.79°**	**2.83°**	94.44%	**5.13°**	**5.87°**	**3.03°**
ResNet-152 [2]	5.61°	4.79°	3.03°	94.01%	5.52°	6.16°	3.18°

References

[1] Gao, B.B., Xing, C., Xie, C.W., Wu, J., Geng, X.: Deep label distribution learning with label ambiguity. IEEE Trans. on Image Processing (TIP) 26(6), 2825–2838 (2016)

[2] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)

[3] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Proc. Neural Information Processing Systems (NIPS) (2012)

[4] Kumar, A., Alavi, A., Chellappa, R.: Kepler: Keypoint and pose estimation of unconstrained faces by learning efficient H-CNN regressors. In: Proc. International Conference on Automatic Face and Gesture Recognition (FG) (2017)

[5] Peng, X., Huang, J., Hu, Q., Zhang, S., Metaxas, D.N.: Three-dimensional head pose estimation in-the-wild. In: Proc. International Conference on Automatic Face and Gesture Recognition (FG) (2015)

[6] Ranjan, R., Patel, V.M., Chellappa, R.: Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. CoRR abs/1603.01249 (2016)

[7] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. CoRR abs/1409.1556 (2014)

[8] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S.E., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015)

[9] Valle, R., Buenaposada, J.M., Valdés, A., Baumela, L.: Head-pose estimation in-the-wild using a random forest. In: Proc. Articulated Motion and Deformable Objects (AMDO) (2016)

[10] Yang, H., Mou, W., Zhang, Y., Patras, I., Gunes, H., Robinson, P.: Face alignment assisted by head pose estimation. In: Proc. British Machine Vision Conference (BMVC) (2015)

Acknowledgements: The authors gratefully acknowledge computer resources provided by the Super-computing and Visualization Center of Madrid (CeSViMa) and funding from the Spanish Ministry of Economy and Competitiveness under projects TIN2013-47630-C2-2-R and TIN2016-75982-C2-2-R.