Head-pose estimation in-the-wild using a random forest

IX Conference on Articulated Motion and Deformable Objects

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

Many of the existing approaches address pose estimation in laboratory conditions. We present a real-time algorithm that estimates the head-pose from unrestricted 2D gray-scale images. We propose a classification scheme, based on a Random Forest, where patches extracted randomly from the image cast votes for the corresponding discrete head-pose.

Patch-based channel features

- Our features are the difference between the average values in two rectangles, R_1 and R_2, in a channel α

$$f(p,\theta) = \frac{1}{|R_1|}\sum_{\mathbf{q}\in R_1} \mathrm{I}^{\alpha}(\mathbf{q}) - \frac{1}{|R_2|}\sum_{\mathbf{q}\in R_2} \mathrm{I}^{\alpha}(\mathbf{q})$$

where $\mathbf{q} \in \mathbb{R}^2$ are pixel coordinates.

- Channels are gray-scale values, Sobel borders and 35 Gabor filters.

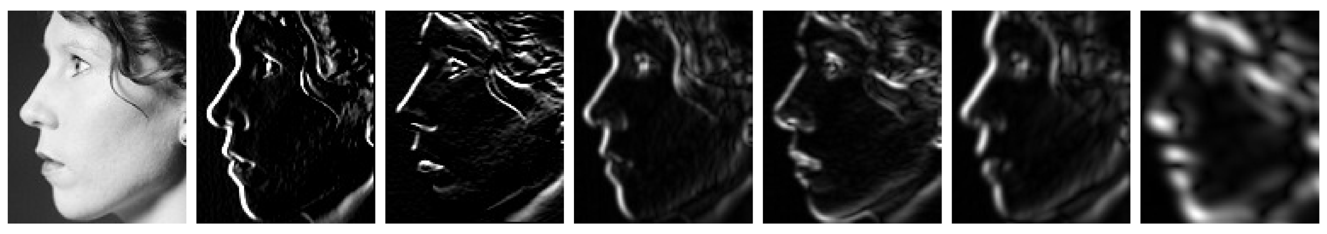

Figure 1: Sample channels used in our approach.

Training regression forest

- Train each decision tree using a randomly selected set of patches from a random subset of the training faces.
- Optimize each weak learner by selecting the $\theta = (R_1, R_2, \alpha)$, from a random pool of candidates $\phi = (\theta, \tau)$, that maximizes the information gain

$$IG(\phi) = \mathcal{H}(\mathcal{P}) - \sum_{S\in\{L,R\}} \frac{|\mathcal{P}_S(\phi)|}{|\mathcal{P}|}\mathcal{H}(\mathcal{P}_S(\phi))$$

where τ details the threshold on the feature value, $\mathcal{P}_L(\phi) = \{\mathcal{P} | f(P,\theta) < \tau\}$, $\mathcal{P}_R(\phi) = \mathcal{P} \setminus \mathcal{P}_L(\phi)$, and $\mathcal{H}(\mathcal{P}_S(\phi))$ is the class uncertainty measure.

- In our case, $\mathcal{H}(\mathcal{P}) = log(\sigma\sqrt{2\pi e})$ is the Gaussian differential entropy.

Head-pose estimation

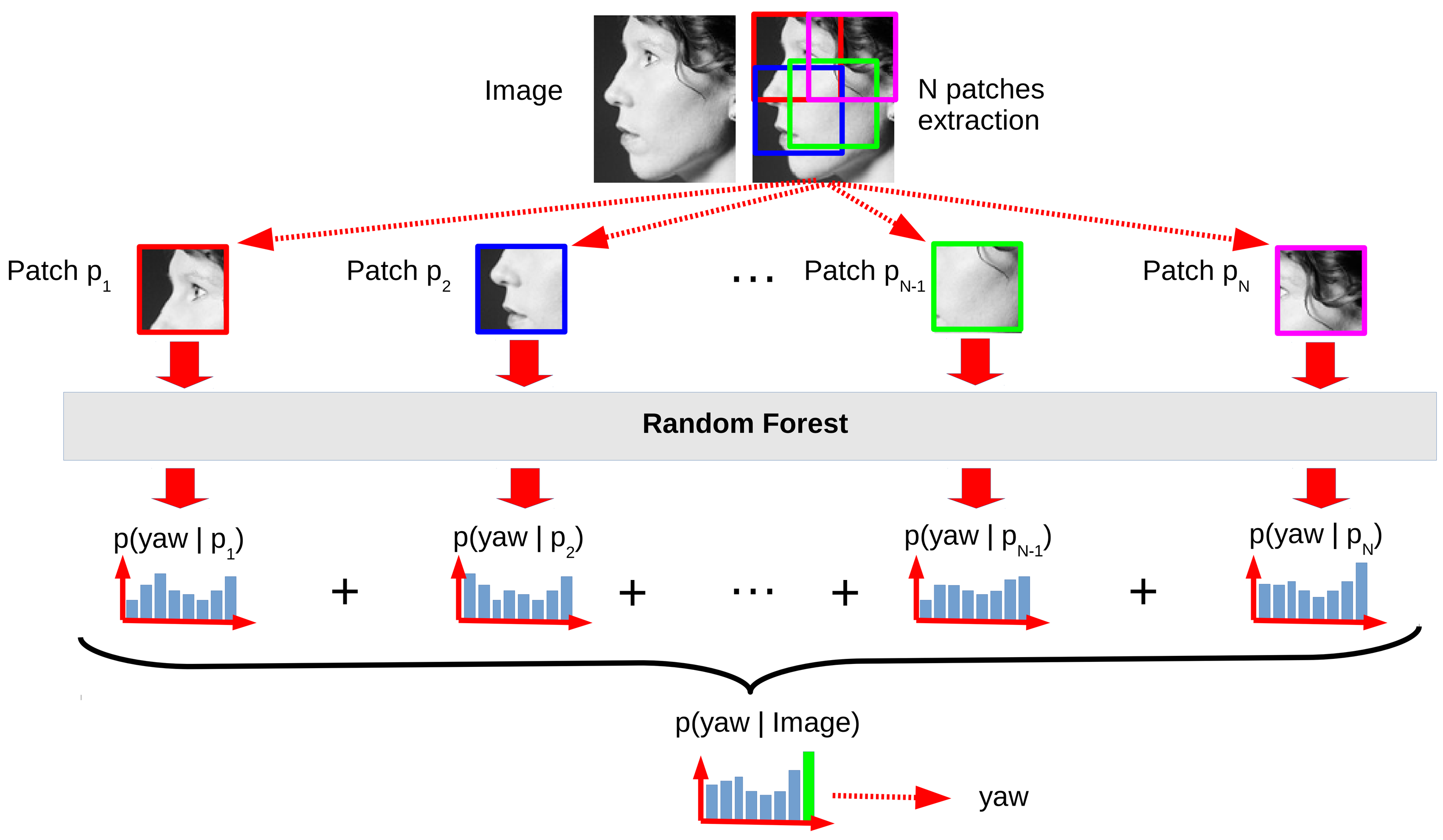

Figure 2: Head-pose prediction using different face image patches.

Experiments laboratory

Our proposal has a MAE close to the state-of-the-art in this database. All three approaches with better results use holistic HOG-based face features [2, 3, 1]. This global feature is slightly more informative, in this constrained context, for estimating face pose than the set of local patches that we use in our approach.

Method	Pointing-04 MAE	Pointing-04 Accuracy (0°)
Stiefelhagen [4]	9.5°	52.0%
Haj [2]	6.56°	67.36%
Hara [3]	5.29°	-
Geng [1]	4.24°	73.30%
Our method	**7.84°**	**55.19%**

Experiments in-the-wild

Our approach achieves the best performance, both in terms of MAE and classification accuracy. It submits a frame rate of 80 FPS on an Intel Core i7 CPU processor at 3.60GHz with 8 cores multi-threaded, 300 times faster than the second best approach, Zhu et al. [6].

Method	AFLW MAE	AFLW Accuracy (≤15°)	AFW MAE	AFW Accuracy (≤15°)
Haj [2]	-	-	-	78.7%
Zhu [6]	-	-	-	81.0%
Sundararajan [5]	17.48°	58.05%	17.20°	58.33%
Our method	**12.26°**	**72.57%**	**12.50°**	**83.54%**

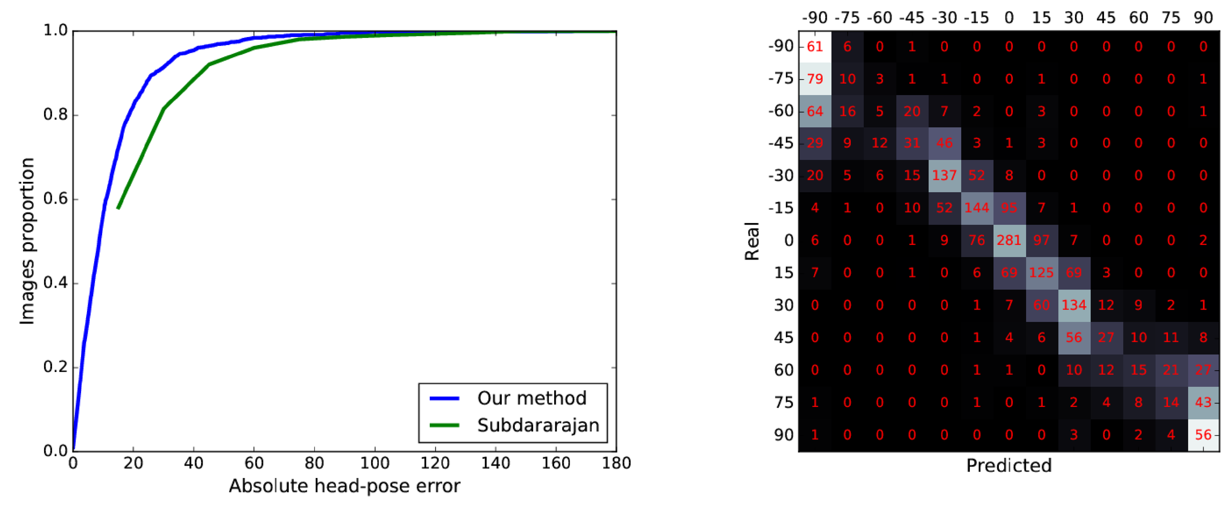

Figure 3: Cumulative head-pose error distribution and confusion matrix for AFLW.

References

[1] Geng, X., Xia, Y.: Head pose estimation based on multivariate label distribution. In: Proc. Conference on Computer Vision and Pattern Recognition (CVPR) (2014)

[2] Haj, M.A., González, J., Davis, L.S.: On partial least squares in head pose estimation: How to simultaneously deal with misalignment. In: Proc. Conference on Computer Vision and Pattern Recognition (CVPR) (2012)

[3] Hara, K., Chellappa, R.: Growing regression forests by classification: Applications to object pose estimation. In: Proc. European Conference on Computer Vision (ECCV) (2014)

[4] Stiefelhagen, R.: Estimating head pose with neural networks. In: Proc. International Conference on Pattern Recognition Workshops (ICPRW) (2004)

[5] Sundararajan, K., Woodard, D.L.: Head pose estimation in the wild using approximate view manifolds. In: Proc. Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2015)

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

Results

Figure 4: Sample results for Pointing-04, AFLW and AFW databases. Green and blue lines indicate respectively pose estimation and ground truth yaw angle.

Acknowledgements: The authors gratefully acknowledge funding from the Spanish Ministry of Economy and Competitiveness under project SPACES-UPM (TIN2013-47630).