Robust face landmark estimation under occlusion: Supplementary material

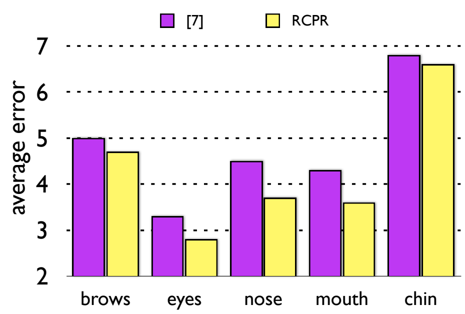
Xavier P. Burgos-Artizzu* Pietro Perona* Piotr Dollár†

*California Institute of Technology †Microsoft Research, Redmond

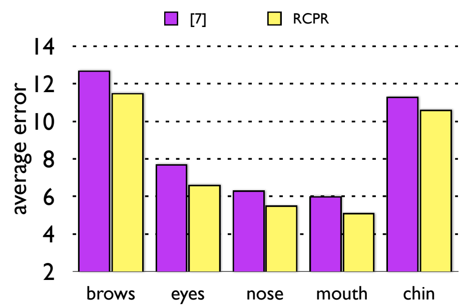
xpburgos,perona@caltech.edu pdollar@microsoft.com

1. Detailed performance comparison

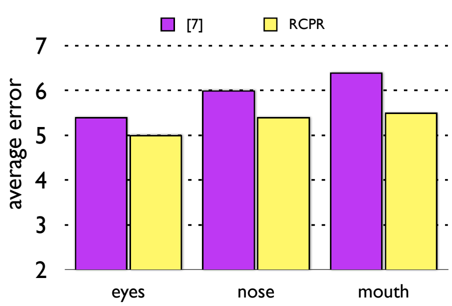
Figure 1 shows the average error of CPR [7] and RCPR on each face part for all four datasets. In the case of **COFW** we also report the error of human annotators to compare human vs. machine. RCPR consistently improves landmark estimation on all face parts on all four datasets. In **COFW**, there is still a big gap with human performance, due to the difficulty of **COFW** faces.



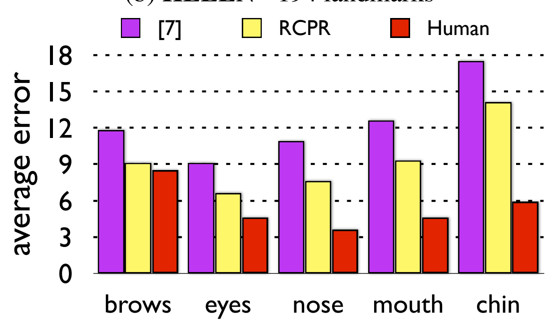
(a) **LFPW** - 29 landmarks



(b) **HELEN** - 194 landmarks



(c) **LFW** - 10 landmarks



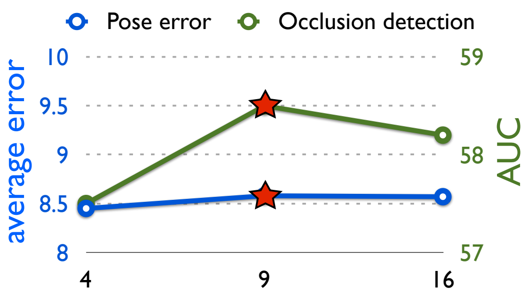
(d) **COFW** - 29 landmarks

Figure 1. CPR [7] vs RCPR average error comparison for each dataset and face part.

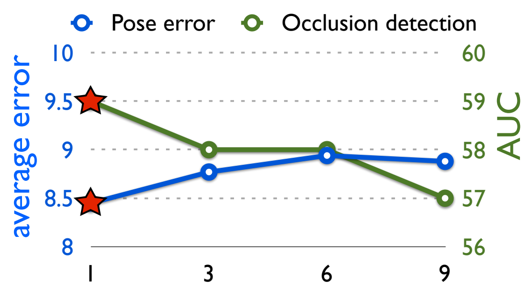
2. RCPR Parameter Sweeps

Figure 2 shows the effect of RCPR's occlusion reasoning parameters on performance on the **COFW** dataset.

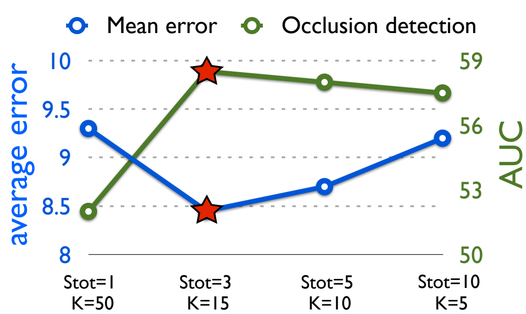
- Figure 2(a): The exact number of regions into which the face is divided has only a minor effect on the average error. However, dividing the face into 9 regions slightly improves occlusion detection.
- Figure 2(b): The number of regions from which each boosted regressor is allowed to draw features plays an important role. Allowing each regressor to draw features from the entire image (sampling regions=total regions=9) results in a degradation in performance similar to that of not using occlusion reasoning at all. On the other hand limiting each regressor to using a few or even a single region results in a higher diversity of regressors and leads to a marked improvement in both landmark estimation and occlusion detection.
- Figure 2(c): The number S_{tot} of regressors combined together through mean voting also has an important role. $S_{tot} = 3$ achieves the best performance both in terms of landmark estimation and occlusion detection (note that when using $S_{tot} > 1$ regressors, we reduce K accordingly to have approximately the same total number of regressors in the cascade).
- Figure 2(d): Analysis of RCPR (occlusion only). Reaching consensus amongst several regressors that derive visual information from different portions of the image (middle point) only improves performance slightly compared to using one regressor. However, weighting regressors according to their occlusion boosts performance, considerably reducing average estimation error.



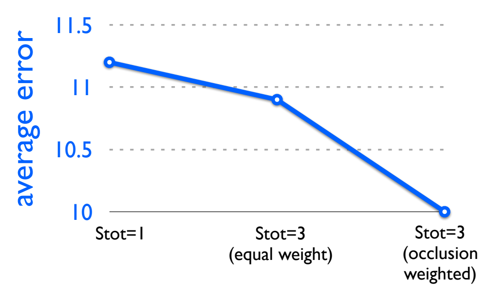
(a) number of face regions



(b) number of sampling regions



(c) S_{tot}



(d) occlusion reasoning

Figure 2. Effect of occlusion reasoning parameters on performance. **(a-c)** Left axis (blue) shows average error (lower=better). Right axis (green) shows area under the curve for occlusion detection precision/recall (higher=better). Red stars mark parameter values used for all experiments in paper. **(d)** Impact of using "visually different" regressors at each iteration and how their information is combined: occlusion-weighted mean boosts performance of RCPR (occlusion only) considerably.