Supplementary Material for the Paper: Robust Classification of Objects, Faces, and Flowers Using Natural Image Statistics

Christopher Kanan and Garrison Cottrell
Department of Computer Science and Engineering
University of California, San Diego
{ckanan,gary}@ucsd.edu

1. Demo Description

To help elucidate the ideas behind our model we provide a demonstration video, NIMBLE_Demo.mp4, with the supplementary materials. The demo is an animated version of the model discussed in the paper applied to a subset of the AR dataset [8]. It first extracts ICA features and acquires fixations/samples from 10 images, each representing a different individual. This is followed by PCA and evaluation of performance using four test images per class. The posterior probability is indicated by the bars above each class, with the correct class denoted with a green bar. The model achieves perfect accuracy using 100 fixations for both training and evaluation. A few frames from the video are shown in figure 1. The code for the demo is available at http://www.chriskanan.com/nimble

2. Software

All of the code used in the production of the results presented in this paper was written in MATLAB. The code for Efficient Fast ICA[5] can be obtained online at: `http://itakura.kes.tul.cz/zbynek/efica.htm`

3. Results on Datasets

Each dataset that we used is available online. For each dataset we perform 5-fold random cross-validation. Unless otherwise noted, per cross-validation run each class has n randomly selected training images chosen, where n is varied, and up to 30 test images randomly selected (distinct from the training images) unless fewer than 30 are available in which case all of the available images are used. After each run the mean per class accuracy (i.e., the standard Caltech-101/256 performance metric) is computed to account for a varying number of test images per class. After all runs are completed, we compute the mean accuracy and standard deviation values.

References

[1] O. Boiman, E. Shechtman, and M. Irani. In defense of Nearest-Neighbor based image classification. In *CVPR 2008*, June.

[2] P. V. Gehler and S. Nowozin. On Feature Combination for Multiclass Object Classification. In *ICCV 2009*.

[3] G. Griffin, A. Holub, and P. Perona. The Caltech-256. *Caltech Technical Report 7694*, 2007.

[4] C. Gu, J. Lim, P. ArbelŽaez, and J. Malik. Recognition using Regions. In *CVPR 2009*.

[5] Z. Koldovský, P. Tichavský, and E. Oja. Efficient Variant Of Algorithm FastICA For Independent Component Analysis Attaining The Cramér-Rao Lower Bound. *IEEE Trans. on Neural Networks*, 17:1090–1095, 2006.

[6] S. Lazebnik, C. Schmid, and J. Ponce. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. In *CVPR 2006*.

[7] Y. Liang, C. Li, W. Gong, and Y. Pan. Uncorrelated linear discriminant analysis based on weighted pairwise Fisher criterion. *Pattern Recognition*, 40:3606–3615, 2007.

[8] A. Martinez and R. Benavente. The AR Face Database. *CVC Technical Report #24*, 1998.

[9] M.-E. Nilsback and A. Zisserman. Automated flower classification over a large number of classes. In *Proc. Indian Conference on Computer Vision, Graphics and Image Processing 2008*.

[10] N. Pinto, D. Cox, and J. DiCarlo. Why is Real-World Visual Object Recognition Hard? *PLoS Computational Biology*, 4, 2008.

[11] N. Pinto, J. DiCarlo, and D. Cox. Establishing Good Benchmarks and Baselines for Face Recognition. In *ECCV 2008*.

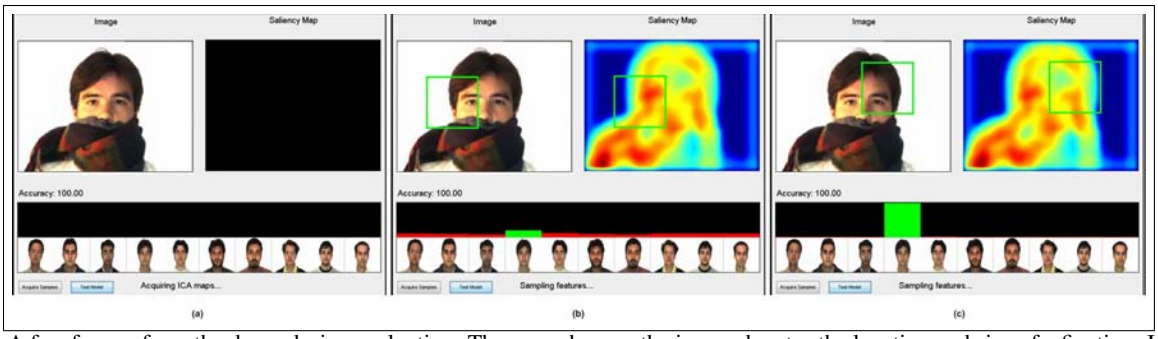

Figure 1. A few frames from the demo during evaluation. The green box on the image denotes the location and size of a fixation. In (1) the ICA maps are being computed. In (2), after a single fixation the model has a weak belief about the correct class (denoted by the green bar) and $P\left(C = 4|\mathbf{g_1}\right)$ is close to the uniform distribution, and (3) after a few more fixations the model is very confident.

Method	1	5	10	15	20	25	30
NIMBLE (This Paper) 1-Desc	33.0 ± 2.5	55.9 ± 0.7	-	70.8 ± 0.7	-	-	78.5 ± 0.4
Boiman et al. 1-Desc [1]	25.3	49.6	-	65.0 ± 1.1	-	-	72.4
Boiman et al. 5-Desc [1]	24.3	56.9	-	72.8 ± 0.4	-	-	79.1
Gehler & Nowozin 1-Desc [2]	-	46.1 ± 0.9	55.6 ± 0.5	61.0 ± 0.2	64.3 ± 0.9	66.9 ± 0.8	69.4 ± 0.4
Gehler & Nowozin 5-Desc [2]	-	54.2 ± 0.6	65.0 ± 0.9	70.4 ± 0.7	73.6 ± 0.6	75.7 ± 0.6	77.8 ± 0.4
Griffin et al. 1-Desc [3]	-	44.2^*	54.2^*	59.4^*	63.3^*	65.8^*	67.6 ± 1.4
Gu et al. 4-Desc [4]	-	45.7	-	65.0	-	-	73.1
Lazebnik et al. 1-Desc [6]	-	-	-	56.4	-	-	64.6 ± 0.8
Pinto et al. 1-Desc [10]	24	47.9	56.8	61.4	-	-	67.4
Yang et al. 1-Desc [13]	-	-	-	67.0 ± 0.5	-	-	73.2 ± 0.6

Table 1. Accuracy results on the Caltech-101 dataset corresponding to figure 5 in the paper. The number of feature types used in each approach, denoted n-Desc, is also provided. Results denoted with a star have been estimated from plots and come from the supplementary material of [2].

[12] R. Singh, M. Vatsa, and A. Noore. Face Recognition with Disguise and Single Gallery Images. *Image and Vision Computing*, 27:245–257, 2007.

[13] J. Yang, K. Yu, Y. Gong, and T. Huang. Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification. In *CVPR 2009*.

Method	1	5	10	15	20	25	30
NIMBLE (This Paper) 1-Desc	11.2 ± 1.6	22.6 ± 0.9	29.3 ± 0.4	32.9 ± 0.4	-	-	-
Boiman et al. 1-Desc [1]	8^*	19^*	27^*	-	33^*	-	37^*
Boiman et al. 5-Desc [1]	8^*	22^*	31^*	-	38^*	-	43^*
Gehler & Nowozin 1-Desc [2]	-	18.4	23.7	28.4	30.7	32.8	34.6
Gehler & Nowozin 5-Desc [2]	-	20.8	30.4	34.2	40.6	42.8	45.8
Griffin et al. 1-Desc [3]	-	18.7 ± 0.5	25.0 ± 0.5	28.4^*	31.3 ± 0.7	33.2^*	34.2 ± 0.2
Pinto et al. 1-Desc [10]	-	-	-	24	-	-	-
Yang et al. 1-Desc [13]	-	-	-	27.7 ± 0.5	-	-	34.0 ± 0.4

Table 2. Accuracy results on the Caltech-256 dataset corresponding to figure 6 in the paper. The number of feature types used in each approach, denoted n-Desc, is also provided. Results denoted with a star have been estimated from plots with most of them coming from the supplementary material of [2]. NIMBLE exceeds other approaches when using one training instance. This is partially because it samples each image multiple times.

Method	1	2	3	5	8
NIMBLE (This Paper)	92.7 ± 0.3	94.0 ± 0.2	95.1 ± 0.2	96.3 ± 0.4	98.3 ± 0.4
Liang et al. [7]	64.9	65.7	71.7	71.4	88.0
Pinto, DiCarlo, & Cox [11]	-	-	-	96^*	98^*
Singh, Vatsa, & Noore [12]	81.2	90	94.5	-	-

Table 3. Accuracy results on the AR dataset corresponding to figure 7 in the paper. All of the approaches use a single feature type. Results denoted with a star have been estimated from plots. Our performance is very close to [11], and they may have exceeded ours since we estimated their performance from plots.

Method	1	5	10	15	20	30
NIMBLE (This Paper) 1-Desc	28.0 ± 1.6	53.7 ± 0.9	62.7 ± 0.7	66.4 ± 0.4	71.4 ± 0.7	75.2 ± 0.2
Nilsback & Zisserman [9] 1-Desc	-	-	-	-	55.1	-
Nilsback & Zisserman [9] 4-Desc	-	-	-	-	72.8	-

Table 4. Results on the 102 Flowers dataset [9] corresponding to figure 5 in the paper. Nilsback and Zisserman [9] used a segmented version of their dataset, but we use the original images for our model. Nilsback and Zisserman also used multiple kernel learning with four different types of features: HSV descriptors, two types of SIFT descriptors, and a HOG descriptor whereas we use a single feature type.