

Supplementary Material

for Learning Invariance through Imitation

Graham W. Taylor, Ian Spiro, Christoph Bregler and Rob Fergus

Dept. of Computer Science, Courant Institute of Mathematical Sciences, New York University

{gwtaylor, spiro, bregler, fergus}@cs.nyu.edu

6. Additional experimental results

We include additional experimental results for the synthetic and real datasets that were not included in the main text due to space constraints.

6.1. Synthetic data

We repeated the synthetic experiments reported in §4.1 using the online, energy-based method. All methods were trained using a fixed learning rate of 0.01 and momentum of 0.9. Parameters were updated after every pair of examples. We evaluated DCG on a separate validation set of 96 sequences after every 10,000 weight updates to determine early stopping.

Again we report the mean and standard deviation over 10 repetitions with randomly initialized parameters. E-Lin and E-Conv are the linear and convnet versions of our energy-based method, respectively. The same trends observed in our experiments using the probabilistic method (Table 1) are seen in Table 2. Soft similarity is beneficial as standard DrLIM is consistently outperformed by the other methods. Additionally, the nonlinear (convnet) variant of our online method always outperforms the linear embedding.

	K=1 NN		K=5 NN		K=10 NN	
	2D	16D	2D	16D	2D	16D
DrLIM	0.28 ± .03	1.13 ± .01	0.69 ± .04	2.39 ± .01	0.91 ± .04	2.80 ± .00
E-Lin (simple)	0.29 ± .03	1.13 ± .01	0.75 ± .07	2.39 ± .01	1.00 ± .07	2.79 ± .01
E-Lin $\sigma = 0.01$	0.30 ± .03	1.14 ± .01	0.79 ± .04	2.39 ± .01	0.98 ± .07	2.79 ± .02
E-Lin $\sigma = 0.1$	0.32 ± .03	1.14 ± .01	0.80 ± .07	2.39 ± .01	1.08 ± .02	2.80 ± .01
E-Lin $\sigma = 1$	0.29 ± .03	1.13 ± .01	0.75 ± .07	2.38 ± .02	1.05 ± .06	2.78 ± .02
E-Conv (simple)	0.34 ± .09	1.12 ± .02	0.97 ± .09	2.38 ± .01	1.34 ± .24	2.81 ± .01
E-Conv $\sigma = 0.01$	0.37 ± .01	1.11 ± .01	1.14 ± .02	2.38 ± .02	1.40 ± .01	2.83 ± .01
E-Conv $\sigma = 0.1$	0.43 ± .04	1.10 ± .02	1.11 ± .10	2.39 ± .02	1.44 ± .01	2.82 ± .01
E-Conv $\sigma = 1$	0.42 ± .00	1.11 ± .00	1.09 ± .05	2.37 ± .01	1.27 ± .26	2.79 ± .04

Table 2: Image retrieval performance (measured by DCG@K) using the synthetic dataset. Online learning.

6.2. Learning an invariant mapping to capture pose

Fig. 7 shows some sample retrieval results returned by pose-sensitive embedding (simple). Figures 8-10 show additional samples from the same model. Our method can distinguish a one finger gesture (Fig. 8, row 8) from two fingers (Fig. 8, row 9) or three fingers (Fig. 8, row 10). It can cope with other people in view (Fig. 9, row 12) or people wearing masks (Fig. 9, row 2) and gloves (Fig. 10, row 1). It can match based on subtleties like the barring of teeth (Fig. 7, row 6) or palms (Fig. 7, row 1) versus the top of the hands (Fig. 7, row 2). Occasionally the model matches well to drawings or renderings that have been uploaded in lieu of imitations (Fig. 7, row 7; Fig. 8, row 15; and Fig. 9, row 14). Of course, matching is not perfect, and these figures also reveal some failures (e.g. Fig. 9, row 11 where the nearest neighbor is incorrect).

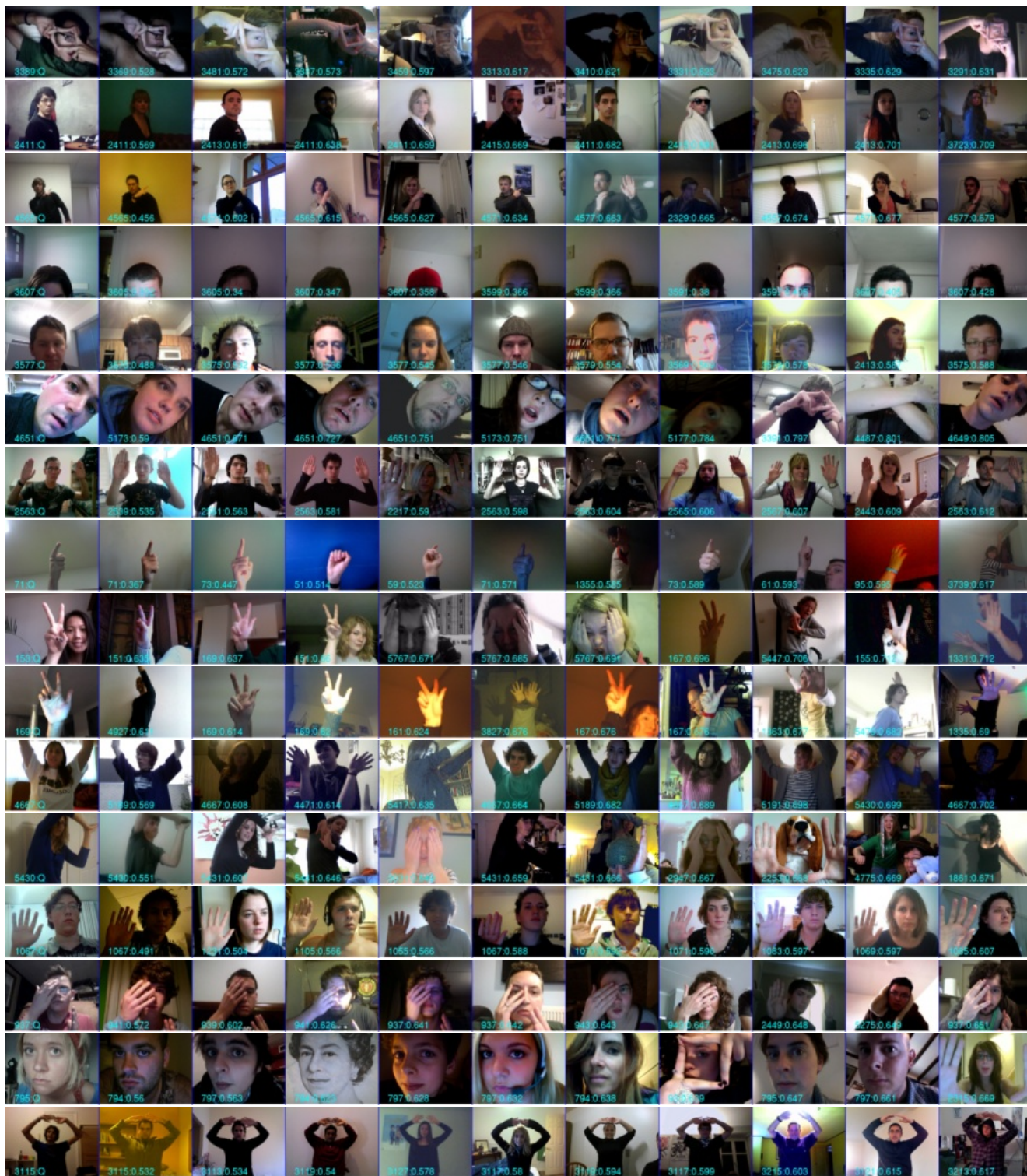


Figure 8: More sample retrieval results (1). Each row is a query. We select a test image (column 1) and find its 10 nearest neighbors using our learned embedding; PSE (simple). Text indicates seed id (left) and distance from the query (right).



Figure 9: More sample retrieval results (2). Each row is a query. We select a test image (column 1) and find its 10 nearest neighbors using our learned embedding; PSE (simple). Text indicates seed id (left) and distance from the query (right).



Images courtesy of C-Mon & Kypski. <http://www.oneframeofframe.com>

Figure 10: More sample retrieval results (3). Each row is a query. We select a test image (column 1) and find its 10 nearest neighbors using our learned embedding; PSE (simple). Text indicates seed id (left) and distance from the query (right).