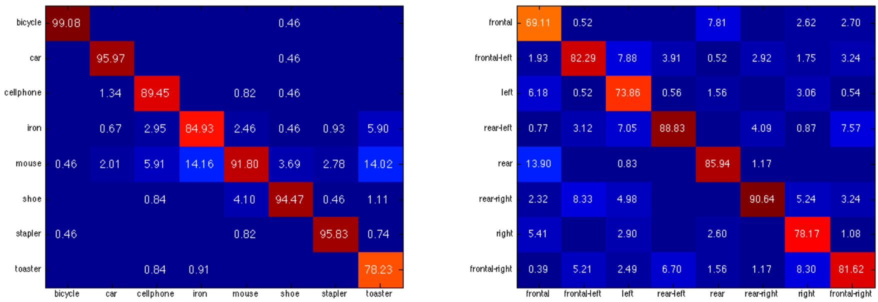
Untangling Object-View Manifold for Multiview Recognition and Pose Estimation Supplementary Materials

Amr Bakry and Ahmed Elgammal

Dept of Computer Science, Rutgers University



(**a**) (**b**)

Fig. 1. This figure shows the confusion matrix for **3DObjects** between the eight classes (**a**), and the eight views (**b**)

	View-Specific Projectors	Instance-specific Projectors	[1]	[2]
Average	90.53%	89.56%	80.07%	75.65%
Bicycle	99.54%	99.54%	99.79%	81.00%
Car	99.31%	100.00%	99.03%	69.31%
Cellphone	98.15%	96.29%	66.74%	76.00%
Iron	86.11%	90.74%	75.78%	77.00%
Mouse	52.58%	44.60%	48.60%	86.14%
Shoe	94.07%	92.59%	81.70%	62.00%
Stapler	98.10%	96.21%	82.66%	77.00%
Toaster	98.15%	99.54%	86.24%	74.26%

Table 1. Comparing our category recognition results on **3DObjects** for each category with [2] and [1]. (20 view samples used)

	View-Specific	Instance-Specific	[2]
Average	80.34%	70.08%	57.46%
Frontal	69.11%	88.35%	64.00%
Frontal-left	82.29%	78.54%	40.40%
Left	73.86%	82.13%	47.00%
Rear-Left	88.83%	80.19%	62.00%
Rear	85.94%	72.68%	53.54%
Rear-right	90.64%	75.85%	71.72%
Right	78.17%	79.23%	57.00%
Frontal-Right	81.62%	82.35%	64.00%

Table 2. Comparing our pose estimation results on **3DObjects** for each viewpoint with [2]. (20 view samples used). (Note: In [1] no such result is reported for comparison.)

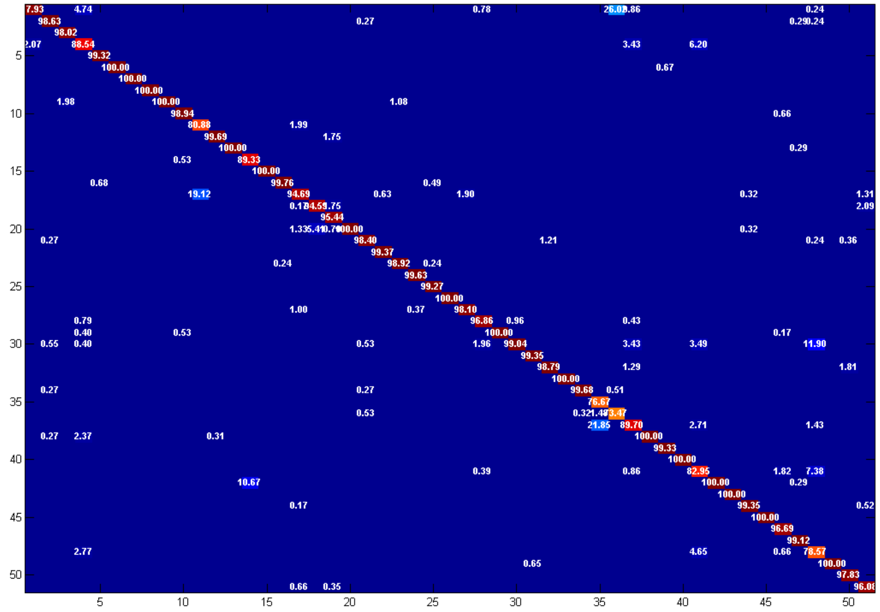


Fig. 2. This figure shows the confusion matrix for **RGBD** between the 51 classes

		Bicycle			Car		
Setting	#v	AE ≤ 22.5	AE ≤ 45	Pole-Grouped	AE ≤ 22.5	AE ≤ 45	Pole-Grouped
20x20x5x31HOG							
Non-Adaptive	4	81.59	81.59	90.55	38.27	38.27	89.38
	8	71.14	85.09	77.61	34.51	41.15	76.33
Adaptive	4	88.58	88.58	95.02	49.78	49.78	91.59
	8	82.59	89.05	88.06	45.8	49.12	85.84
15x15x3x31HOG							
Non-Adaptive	4	78.61	78.61	87.56	38.72	38.72	87.83
	8	60.2	84.08	66.67	34.29	41.59	75.22
Adaptive	4	83.58	83.58	94.03	50.66	50.66	90.27
	8	74.63	83.58	84.08	44.69	48.23	82.08

Table 3. Scalability: The table shows the ability of the model to scale and generalize to datasets with no multiview images of the same object. In this experiments, we learned the model in 3DObjects and Tested on PASCAL VOC2006 (Bicycle and Cars). The task is pose estimation. Two setting are shown: 1) Non-adaptive setting: The model is trained on 3DObjects and tested on PASCAL VOC. 2) Adaptive: The model is trained on 3DObjects, then adapted by adding the training data from PASCAL VOC, and tested on PASCAL VOC. Adapting the model involved projecting the new data (single view for each instance) using the learned view-specific projectors to obtain style vectors, then obtaining the coefficient matrix for each of these new images (manifold parameterization), then computing new tensors, and computing new view-specific projectors. Since 3DObject has 8 views and PASCAL has 4 views, we computer the pose accuracy by computing the absolute error between the estimation and ground truth and report the percentage with AE $\leq$ 22.5 and AE $\leq$ 45. We also report the pose estimation after grouping the two opposite poses (Pole grouped), which shows that most of the confusion is between opposite poses. We tested on two HOG settings. We tested with 4 and 8 view-specific projectors (#v). Without adapting the model the results are quite good. In all cases, adapting the model improves the results. The pose estimation results are worse for cars, this is mainly because the images of cars in the 3DObject dataset are different from the ones in PASCAL.

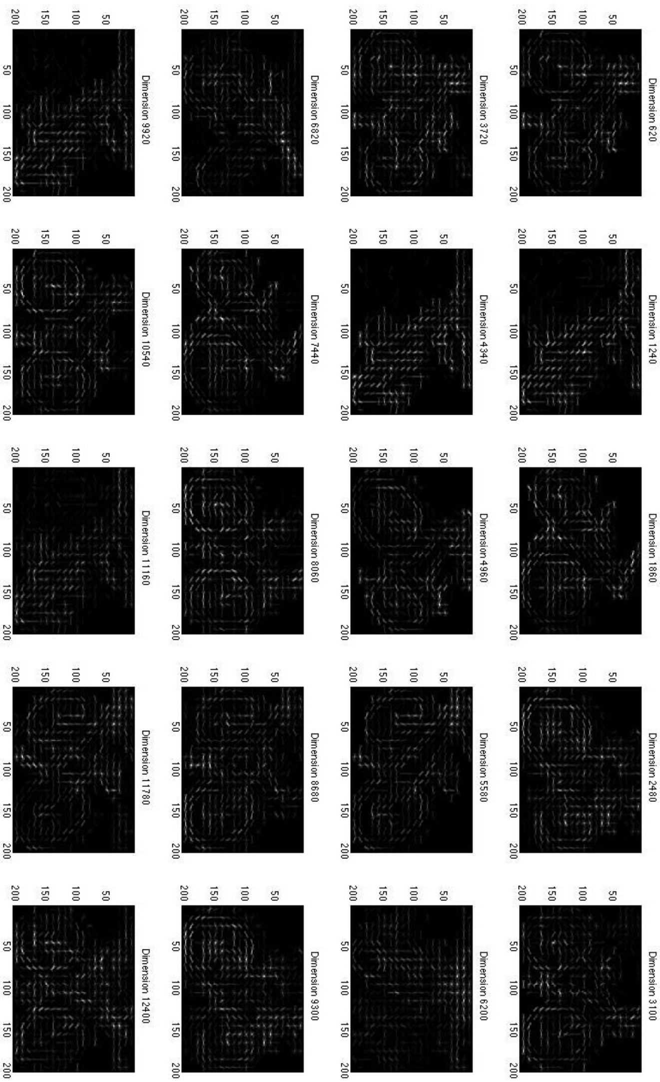


Fig. 3. Representation: Sample visualization of the columns of the matrix $B_i B_i^\dagger$ for the case of bicycles, the plots clearly show templates of different bicycles at different viewpoints.

References

1. Zhang, H., El-Gaaly, T., Elgammal, A., Jiang, Z.: Joint object and pose recognition using homeomorphic manifold analysis. In: AAAI. (2013)
2. Savarese, S., Fei-Fei, L.: 3d generic object categorization, localization and pose estimation. ICCV (2007)