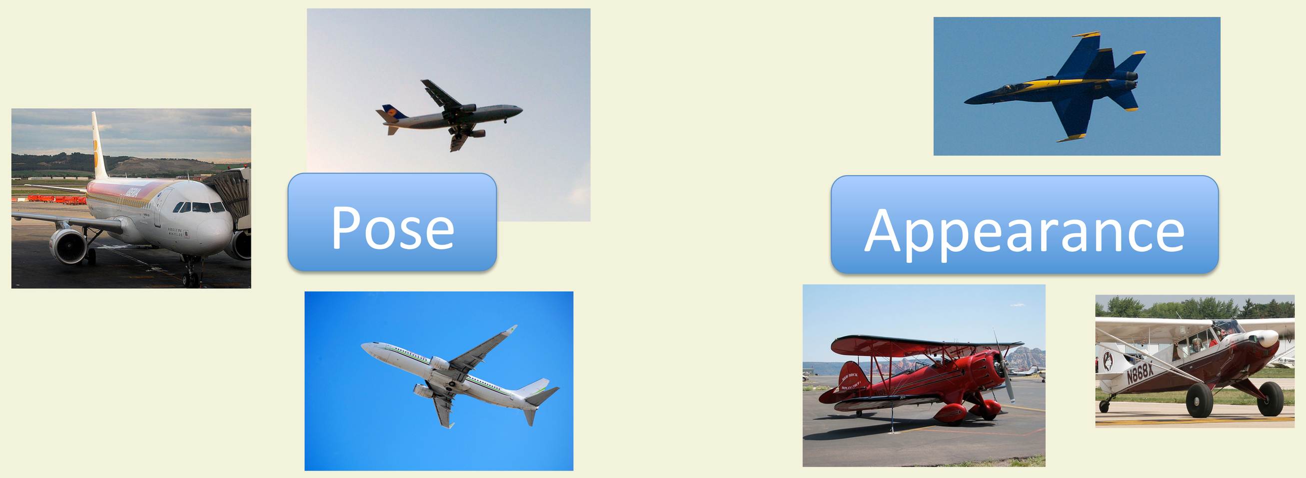
Detecting Objects using Deformation Dictionaries

Bharath Hariharan[1], C. Lawrence Zitnick[2], Piotr Dollár[2]
[1]Berkeley, [2]Microsoft Research

Overview

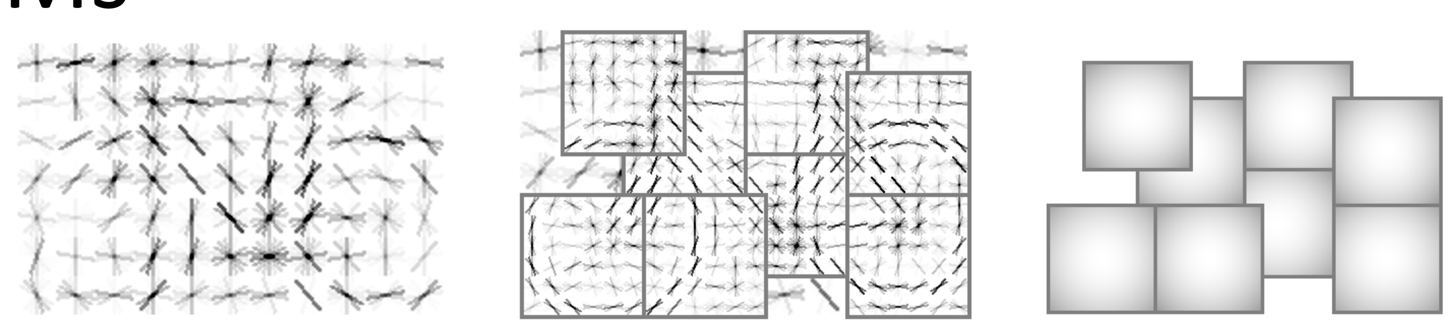
Objects within general categories vary significantly from one instance to the next. We can group the numerous factors contributing to these changes into two broad categories:



Mixture models treat each factor independently which requires lots of data to train, and may lead to overfitting.

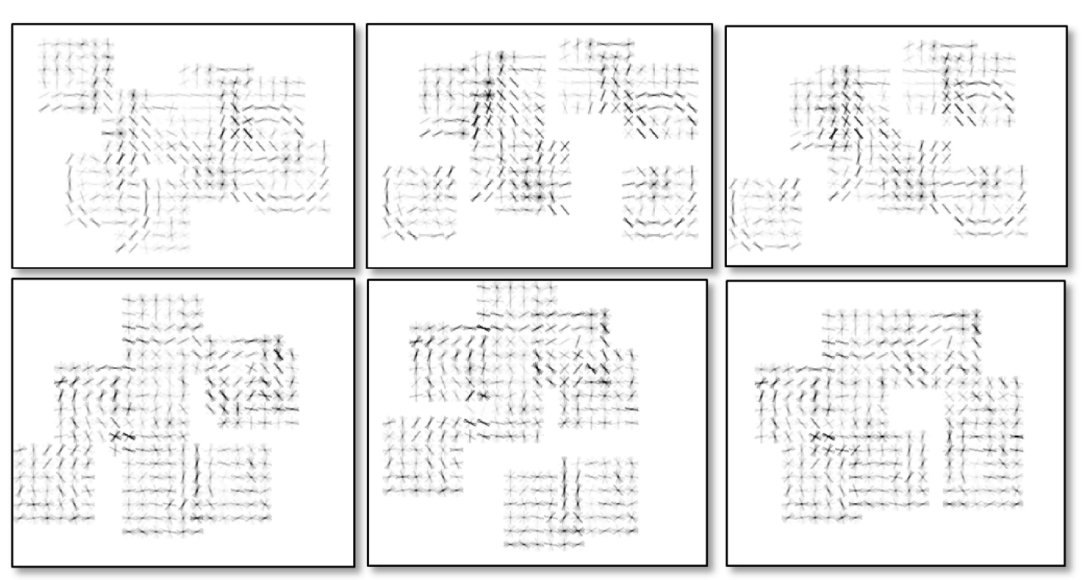
Contribution : We propose using one appearance model that deforms according to pose

DPMs



- Each part placement gives a new template
- *Exponentially many* templates that *share parameters*

✓ Can generalize to unseen poses
✗ May not capture all relevant deformations
✗ Are mostly implausible:



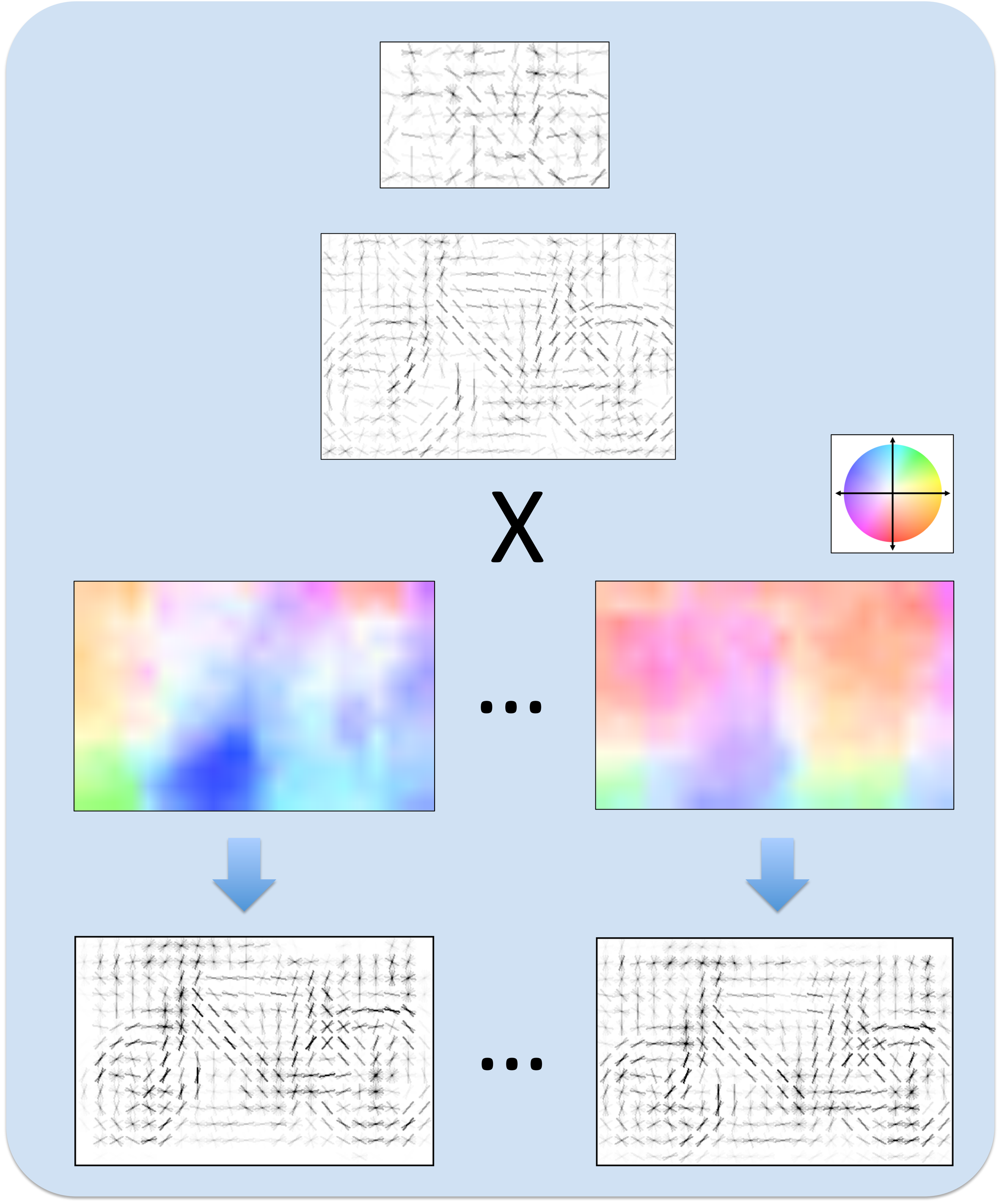
Can a small data-driven deformation dictionary capture most plausible deformations?

Our model

Each component has:

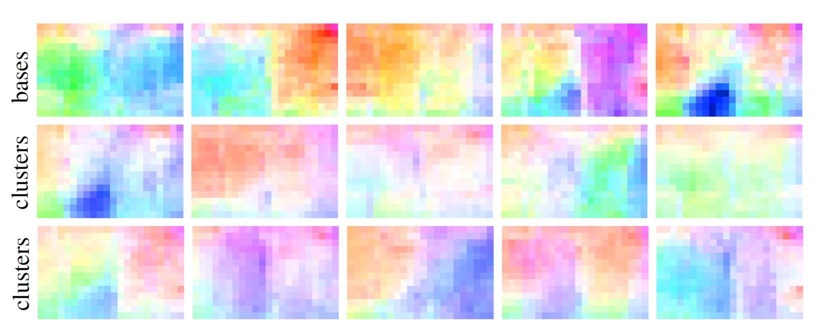
1. A coarse template w_r
2. A single fine template w_f
3. A dictionary of deformations $\mathcal{D}$

Deformed templates $D^T w_f$

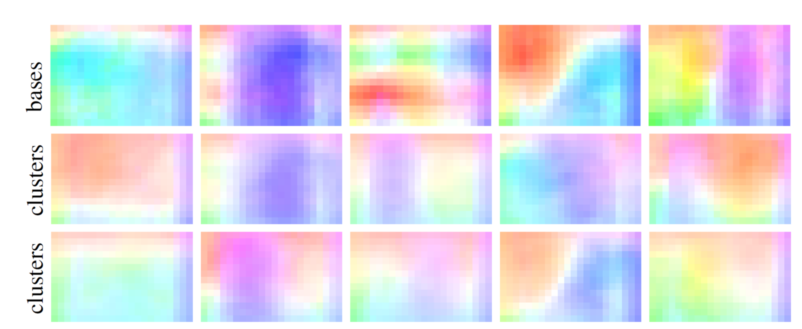


Generating the dictionary

- Whiten all exemplar HOGs
- Estimate deformation from mean to exemplars
  Use block matching with small HOG patches
- Do PCA and project to top principal components
- Cluster using k-means

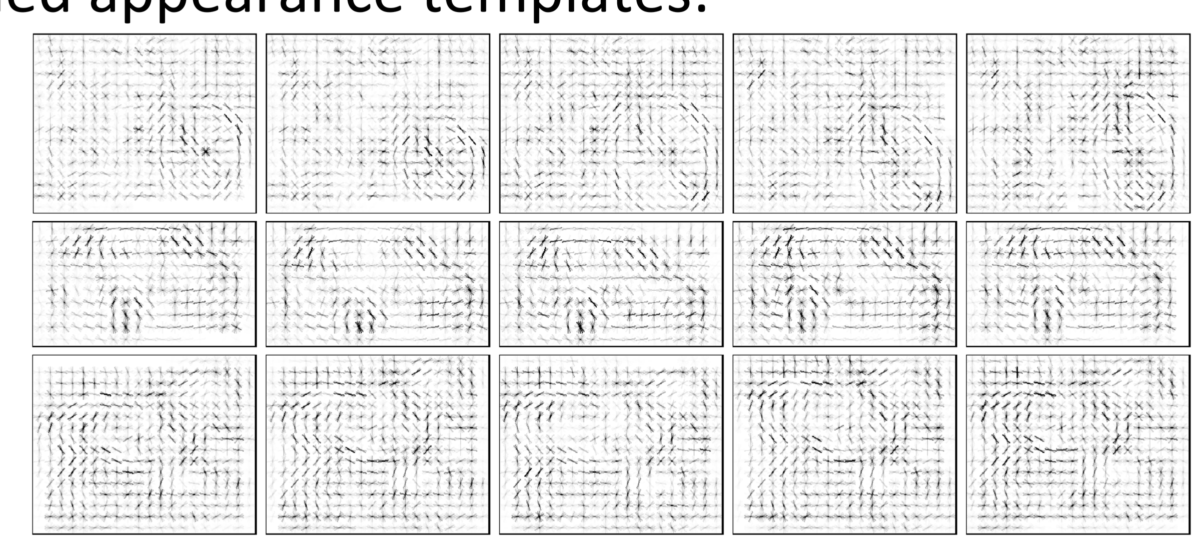


The k = 5 estimated flow bases for bicycles. Middle/Bottom: the m = 10 centroids obtained by clustering bicycle deformations.



The k = 5 estimated flow bases for all categories. Middle/Bottom: the m = 10 centroids obtained by clustering bicycle deformations.

Deformed appearance templates:



Results

Example detections:

The two detections in each pair share the template but have different deformations.

Parameter sharing:

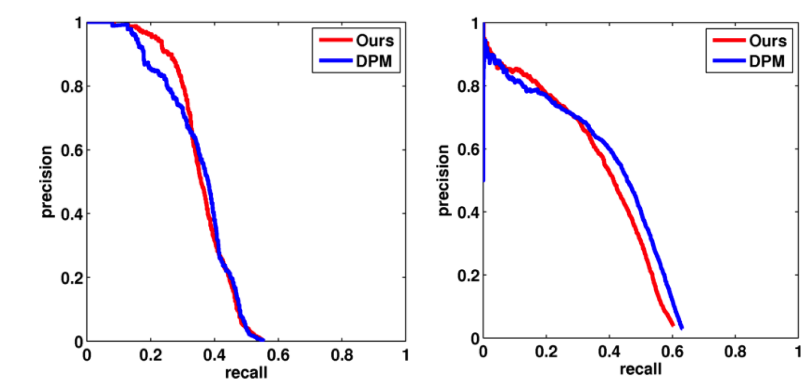
n-comp: Each component has just a low-resolution root filter.

nm-comp: A standard mixture model with as many root filters as we have high-resolution templates.

n-comp x m-fine: Each component has a root filter and m high-res templates trained without parameter sharing.

method	bike (100 samples)	bike (full data)
n-component	30.4	40
nm-component	26	–
n-comp$\times m$ fine	36.2	44.8
Ours	**38.9**	**46.5**

Full experiments on PASCAL VOC 2012:



Precision recall curves for car (left) and person (right). We do well in the high precision regime.

	n-comp	n-comp ×m-fine	DPM	ours	ours -common
plane	27.3	32.0	**40.5**	36.6	35.8
bike	40.0	44.8	45.2	46.5	**46.9**
bird	1.3	3.0	1.9	4.2	**4.4**
boat	4.5	4.3	4.0	**5.4**	4.9
bottle	16.4	17.6	19.2	20.1	**20.7**
bus	47.3	55.1	53.0	54.9	**55.3**
car	27.3	**36.4**	35.3	36.3	35.5
cat	9.1	17.8	18.6	**19.8**	19.7
chair	7.1	9.8	**14.6**	11.7	11.5
cow	5.9	11.3	10.3	**13.4**	12.3
table	2.9	6.0	3.9	**8.9**	8.0
dog	5.8	7.1	**10.7**	7.3	7.7
horse	23.5	33.5	32.9	**35.6**	34.7
moto	27.1	26.4	**31.4**	31.1	30.8
person	28.2	36.7	**38.5**	36.4	36.8
plant	1.7	4.3	4.0	**6.6**	5.1
sheep	23.5	25.6	25.5	**27.0**	27.0
sofa	5.4	7.8	11.6	11.0	**13.6**
train	22.0	33.3	32.7	**35.0**	32.9
tv	21.3	33.4	**34.2**	32.7	31.9
mean	17.4	22.3	23.4	**24.0**	23.8

Sharing deformation dictionaries:

category	category-specific	common	super-category	DPM
Bicycle	46.5	46.9	**47**	45.2
Motorbike	31.1	30.8	**31.6**	31.4
Cow	13.4	12.3	**13.7**	10.3
Sheep	**27**	**27**	26.9	25.5

Conclusion

We have proposed using a discrete set of deformations. However, we can also search for the optimal deformation within the space defined by our set of 5 PCA bases. Using a greedy search technique, we were able to obtain similar results to that of our discrete model. While the discrete approach is more computationally efficient, it may prove beneficial to search in a continuous space of deformations for some object categories.