SPATIAL INVARIANCE FOR SCENE CLASSIFICATION

Dunlop H.
hdunlop@digitalsmiths.com

Introduction

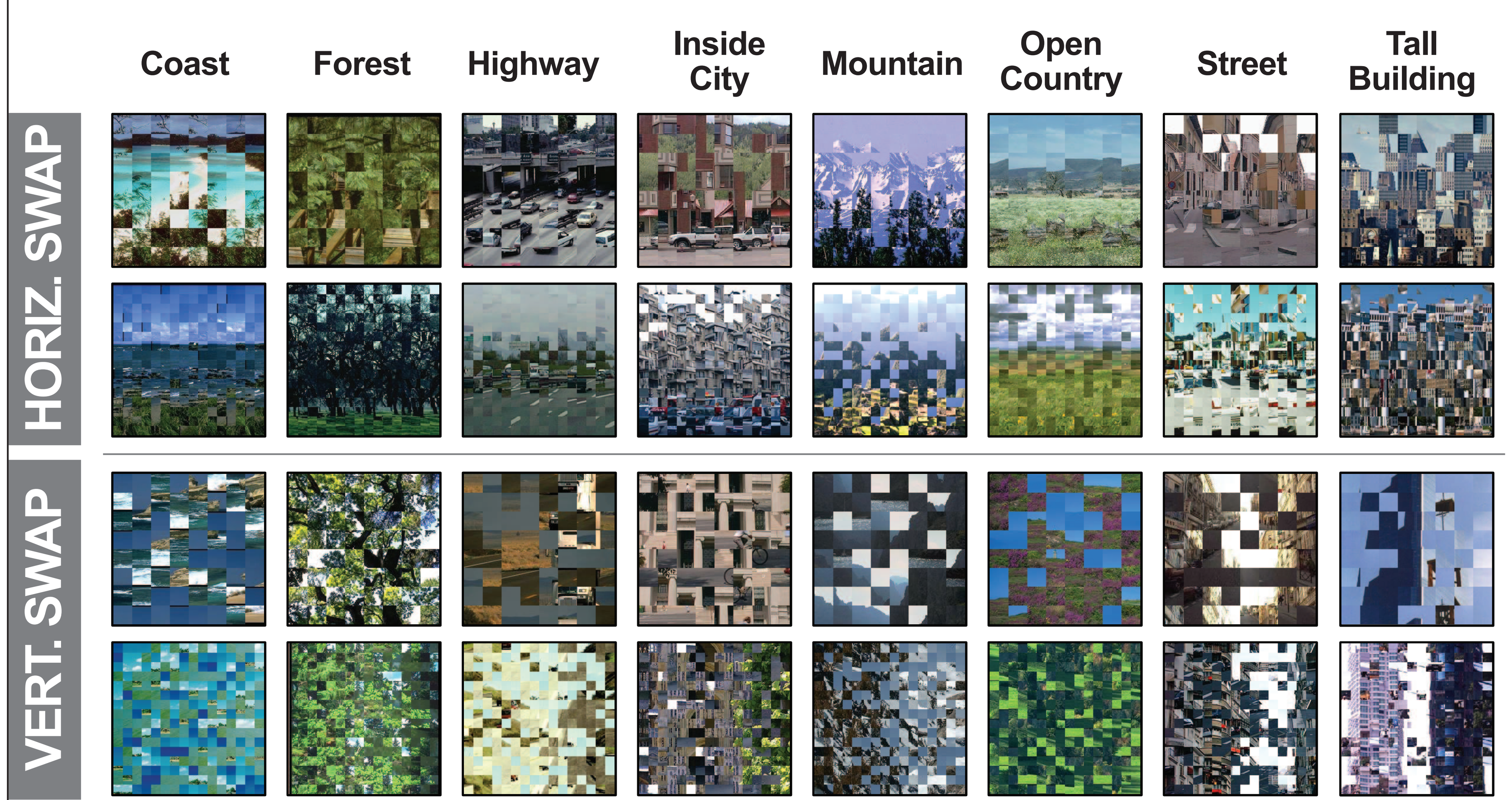
We present a method for categorizing scenes using properties from local, intermediate, and global scales with a learned spatially invariant model. By applying a spatial pyramid to regions of varying sizes and in different spatial locations, we are able to better characterize the appearance of images. Adding spatial invariance that is modeled for each scene category, these mid-sized region descriptors complement well the local and global methods already in common use for scene classification.

Problem:

- Bag of words model is too loose.
- Spatial pyramid [1] is too rigid.

Approach:

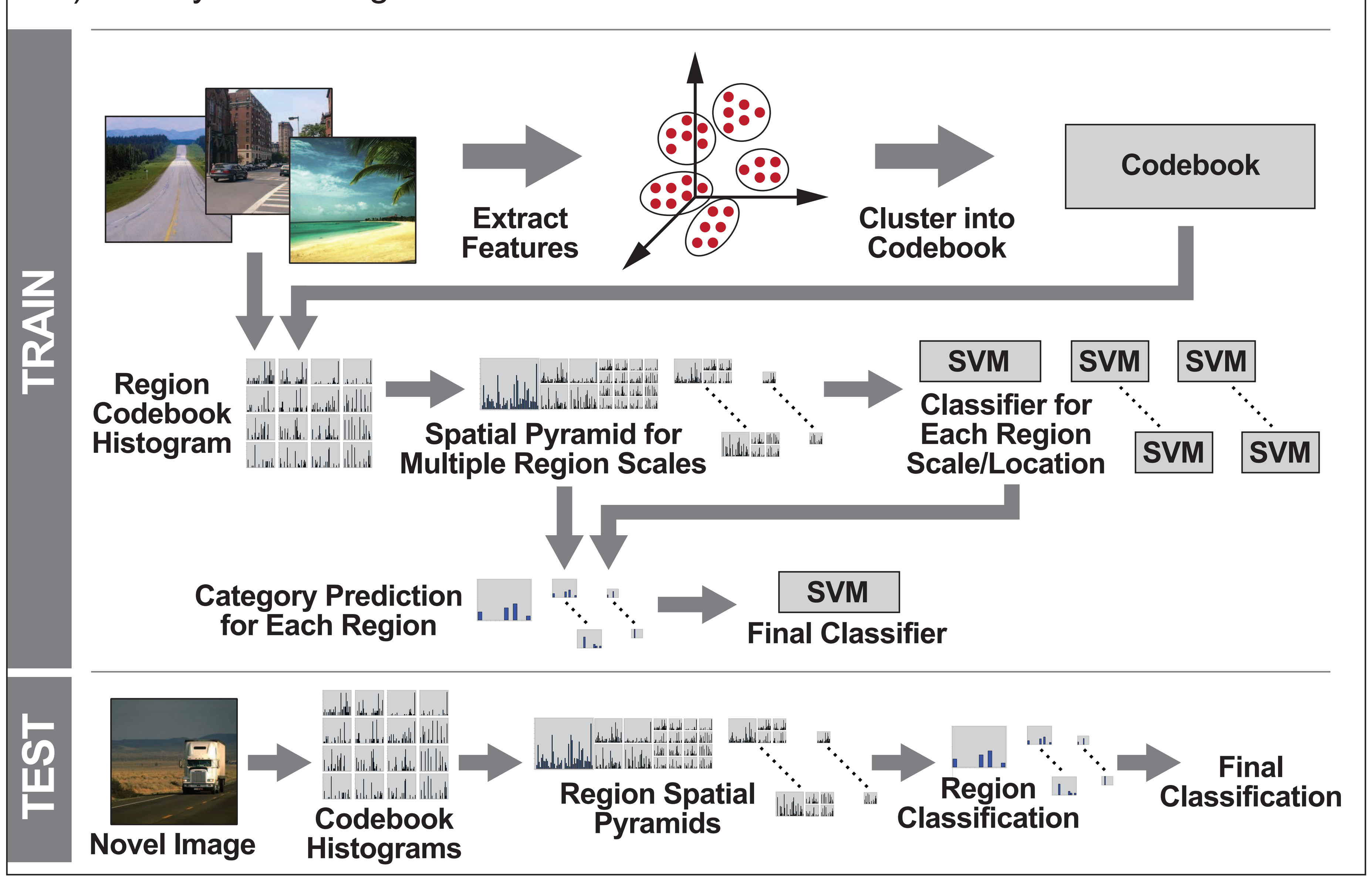
- Classify intermediate-sized regions in addition to the whole image.
- Allow horizontal invariance.
- Learn how much vertical invariance to apply to each category.



First two rows: after randomly swapping regions from the same row, image categories are still identifiable. Last two rows: randomly swapping regions from the same column produces images that are less coherent for most categories. Conclusion: horizontal invariance may be useful, vertical is less so.

Spatial Pyramid Region Descriptor for Classification

1) Extract dense SIFT features and cluster into a bag of words.
2) Represent multiple scales of regions by a spatial pyramid [1].
3) Classify regions into coast, forest, highway, etc.
4) Classify whole image.



Spatial Invariance

The amount of spatial invariance is dependent on the scene category. Generally, invariance is much stronger in the horizontal direction than vertical.

Four models:

- No invariance:
  No spatial invariance is used.
- Full x invariance:
  Horizontal invariance is assumed and the classifiers are trained accordingly.
- Full x, partial y invariance:
  Horizontal invariance is assumed and vertical invariance is learned.
- Partial x and y invariance:
  Horizontal and vertical invariance are both learned.

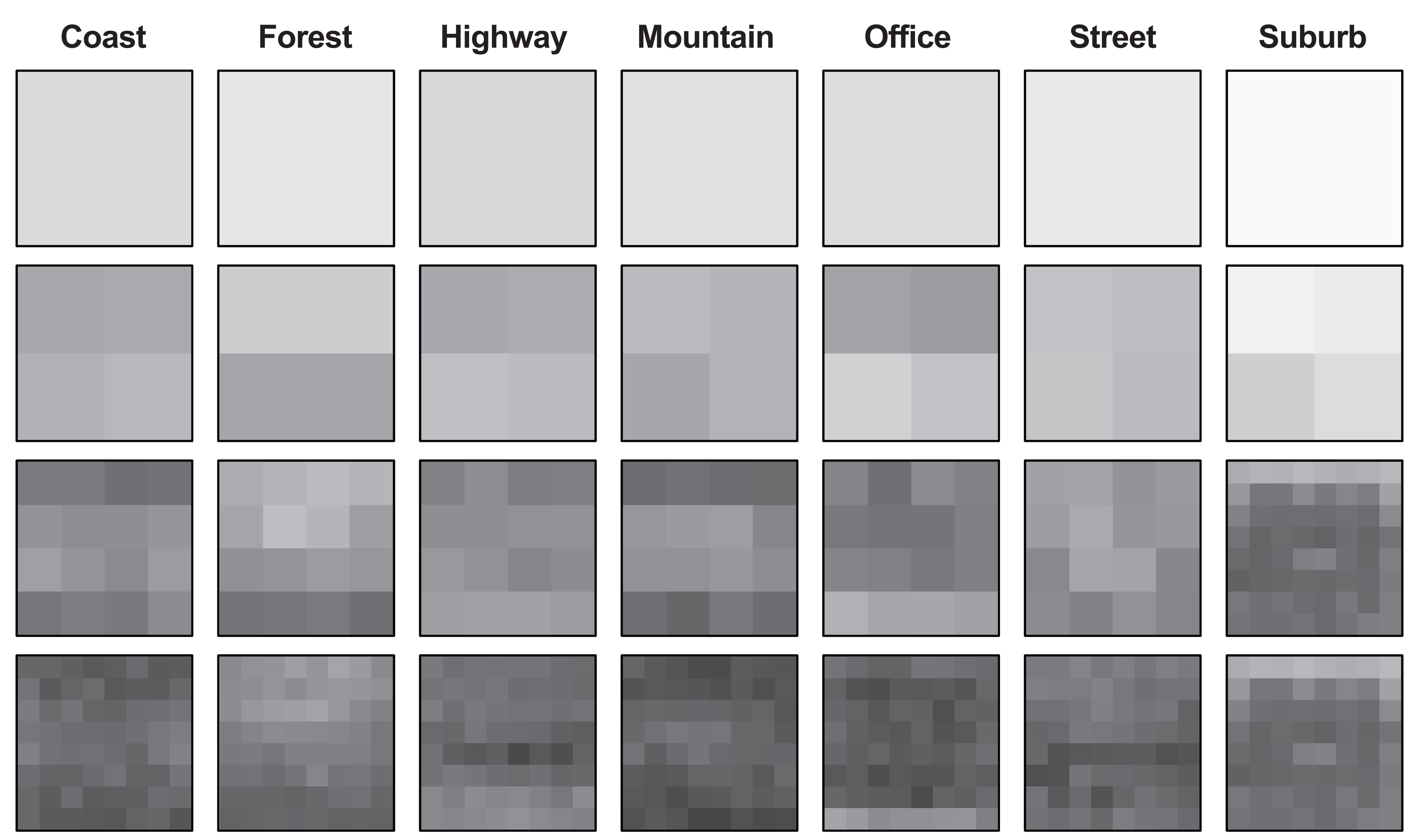
Experiments

Three data sets of increasing complexity:

- Scene-8: outdoor categories [2]
- Scene-15: 7 additional indoor/outdoor categories [1]
- Indoor-67: indoor categories [3]

Each spatial invariance model is tested with 1 to 4 region scales. The results demonstrate that spatial invariance improves performance on standard data sets and out performs other methods on these same data sets.

Data Set	Scales	No Invariance	Full x	Full x, Partial y	Partial x and y	Other Authors
Scene-8 [2]	1	87.5%	-	-	-	87.8% [4] pLSA+SPM+color
	2	87.0%	86.2%	88.6%	87.2%	
	3	87.6%	87.2%	88.2%	88.2%	
	4	88.5%	88.4%	**89.2%**	88.3%	
Scene-15 [1]	1	81.7%	-	-	-	88.1% [5] SPM, many feature types
	2	80.7%	81.2%	81.5%	82.0%	
	3	82.3%	82.6%	83.6%	83.4%	83.7% [4] PLSA+SPM, SIFT only
	4	82.8%	82.8%	**84.2%**	83.0%	
Indoor-67 [3]	1	38.9%	-	-	-	25.0% [3] ROI+GIST segmentation
	2	37.6%	39.1%	40.7%	38.9%	
	3	38.9%	41.8%	**42.7%**	40.3%	



A comparison of the performance of the region classifiers at different scales (with no horizontal or vertical invariance). The intensity of each region is proportional to the accuracy that the classifier achieved. Each category has a unique distribution of regions that are most informative.

References

[1] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In Proc. CVPR, 2006.
[2] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. IJCV, 42(3):145–175, 2001.
[3] A. Quattoni and A. Torralba. Recognizing indoor scenes. In Proc. CVPR, 2009.
[4] A. Bosch, A. Zisserman, and X. Munoz. Scene classification via pLSA. In Proc. ECCV. 2006.
[5] J. Xiao, J. Hays, K. Ehinger, A. Oliva, and A. Torralba. SUN database: Large-scale scene recognition from abbey to zoo. In Proc. CVPR, 2010.

Digitalsmiths®
Fanatical Video Discovery.™