BİL-722 Advanced Topics in Computer Vision

Presentation: Çağdaş Baş

Paper: Randomized Spatial Partition for Scene Recognition

Authors: Yuning Jiang, Junsong Yuan, and Gang Yu

RANDOMIZED SPATIAL PARTITION FOR SCENE RECOGNITION



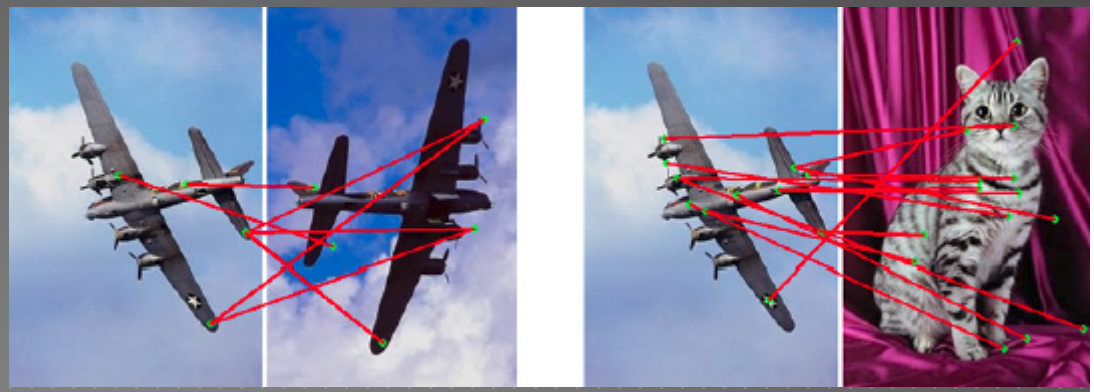



Images from: SUN Dataset

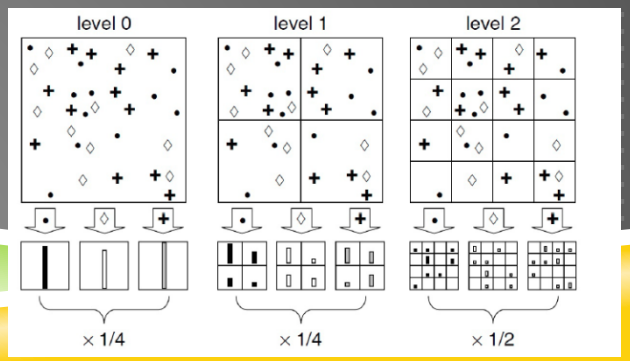
Related Work

- Behmo et. al. (ECCV2010)
  - Feature points are independent

- Zhang et. al. (ACM2009)
  - Only relative positions are considered



- Lazebnik et. al. (CVPR2006)
  - Fixed partitioning



Spatial Pyramid Matching

Lazebnik et. Al. (CVPR2006)

- Build pyramid in image space, quantize feature space

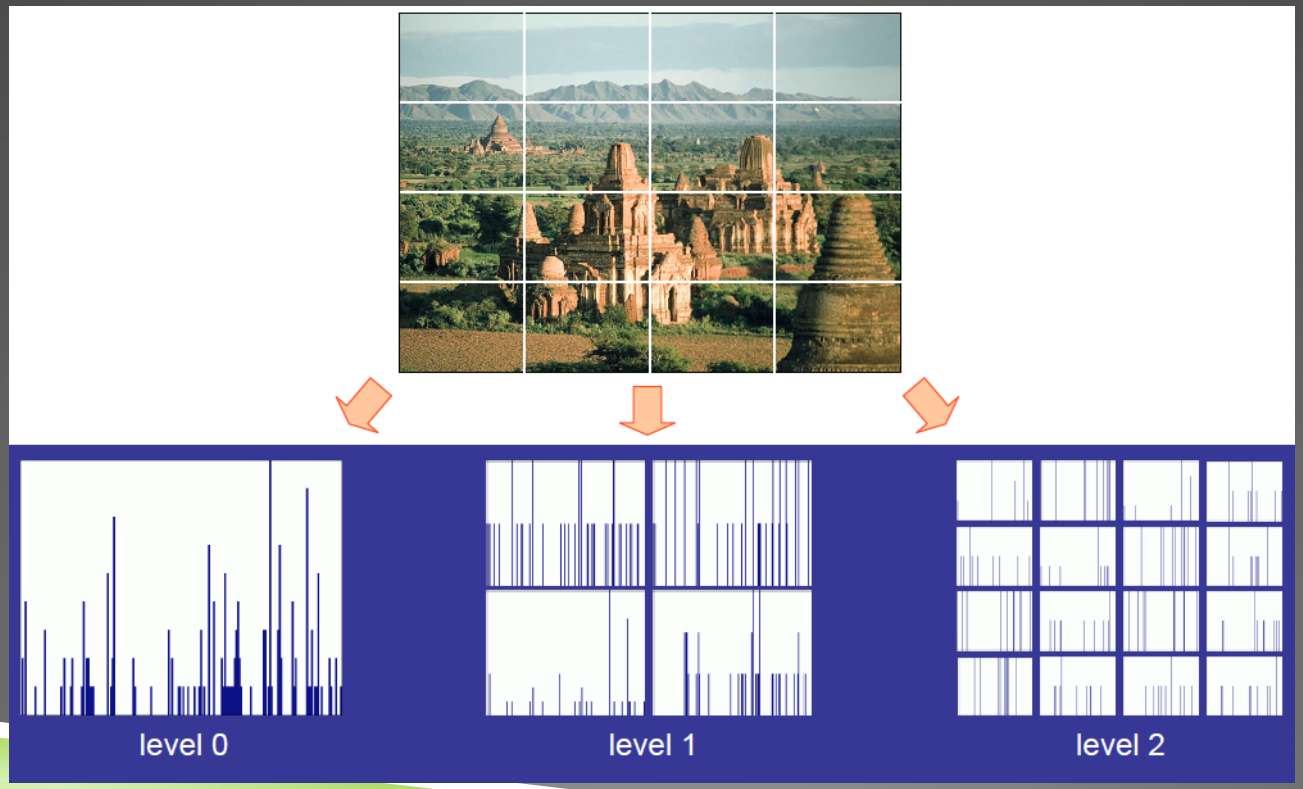


Figure Credit: Svetlana Lazebnik, Cordelia Schmid, Jean Ponce

Pyramid matching

Indyk & Thaper (2003), Grauman & Darrell (2005)

Find maximum-weight matching (weight is inversely proportional to distance)

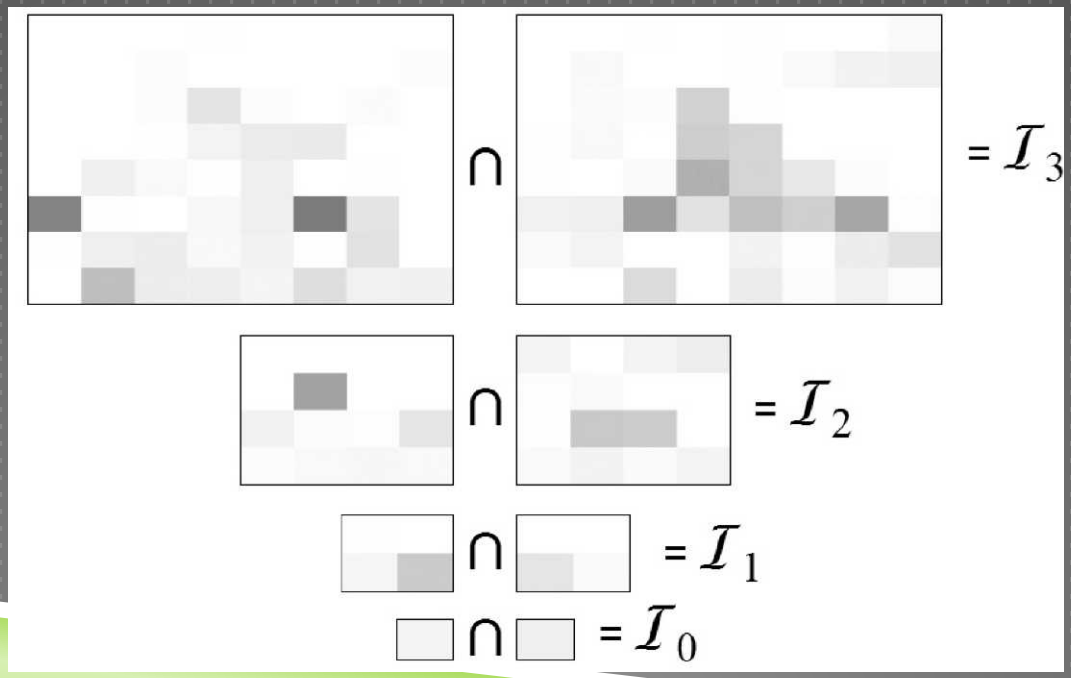
Original images:

Feature histograms:
Level 3

Level 2

Level 1

Level 0



Slide Credit: Svetlana Lazebnik, Cordelia Schmid, Jean Ponce

Limitations of SPM

- Different spatial layouts for similar scenes (a)
- Intra-class variations due to scale, viewpoint and rotation (b)

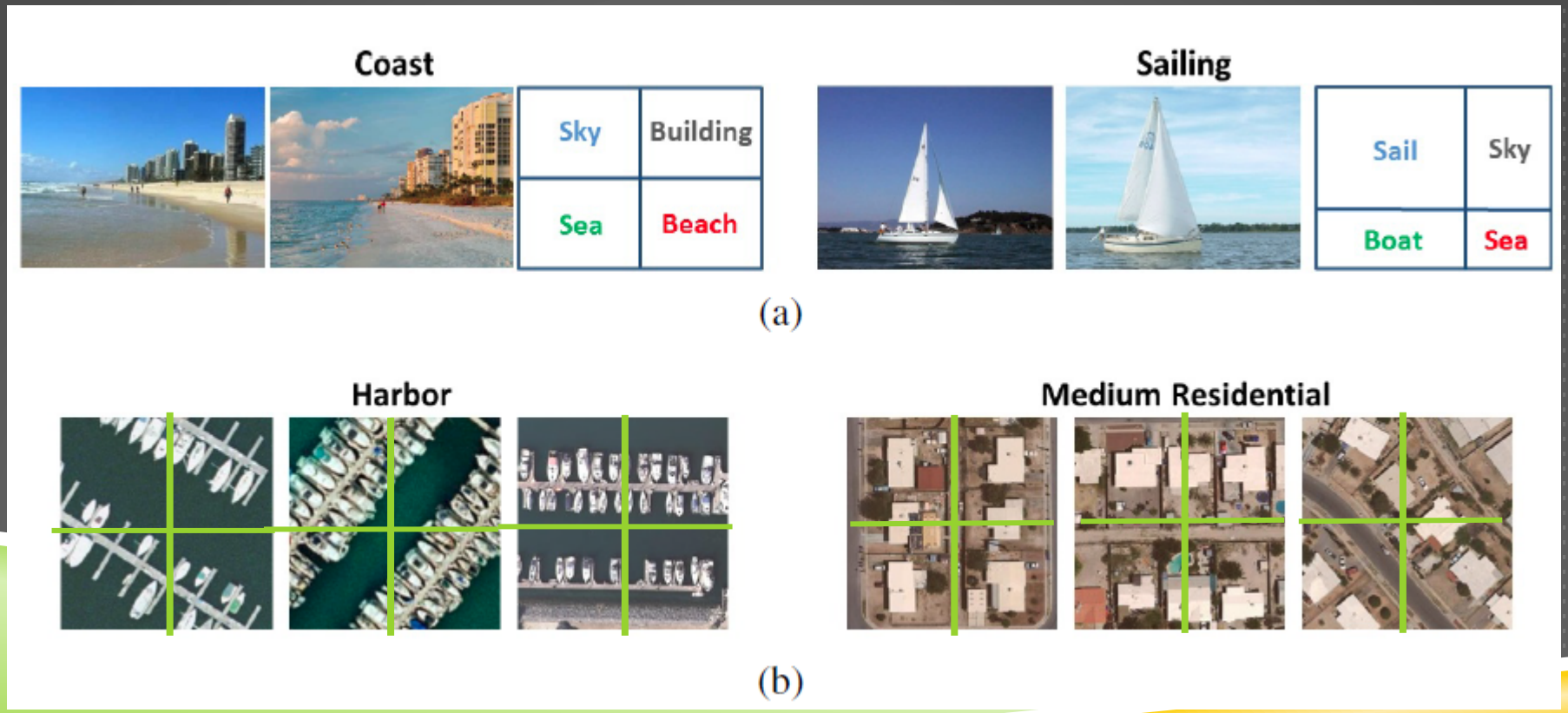
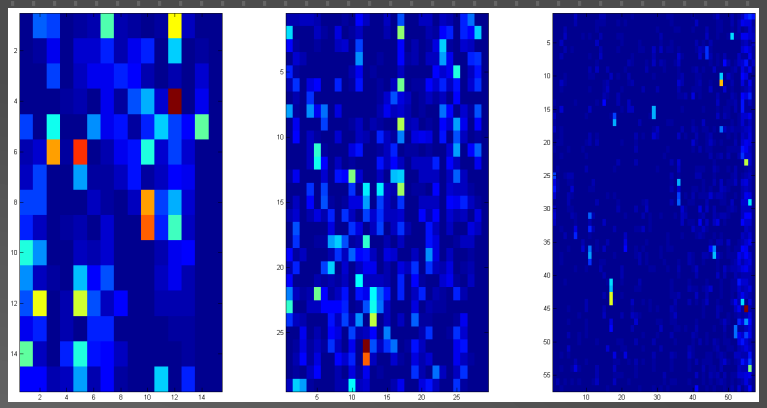
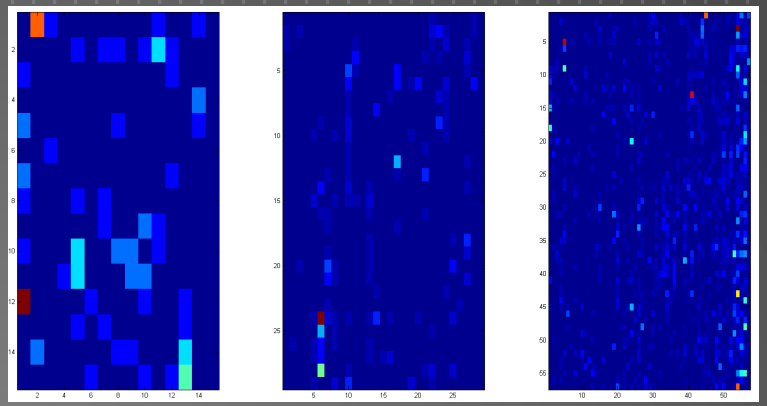
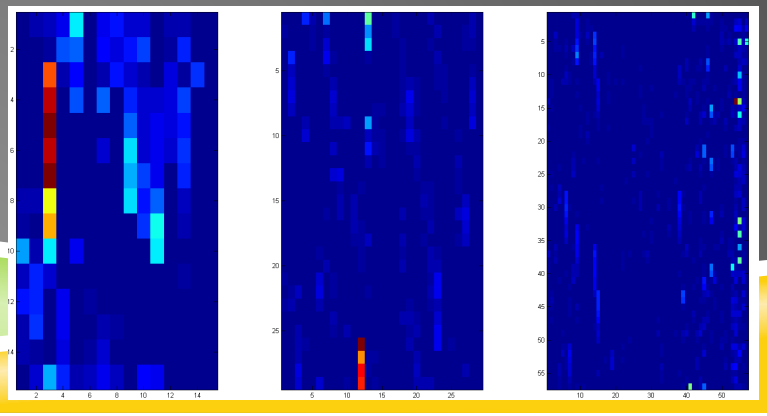


Figure Credit: Yuning Jiang, Junsong Yuan, Gang Yu

Limitations of Spm







Images from: SUN dataset

Randomized Patterns

- Instead of making fixed partitions, randomize partition for each class

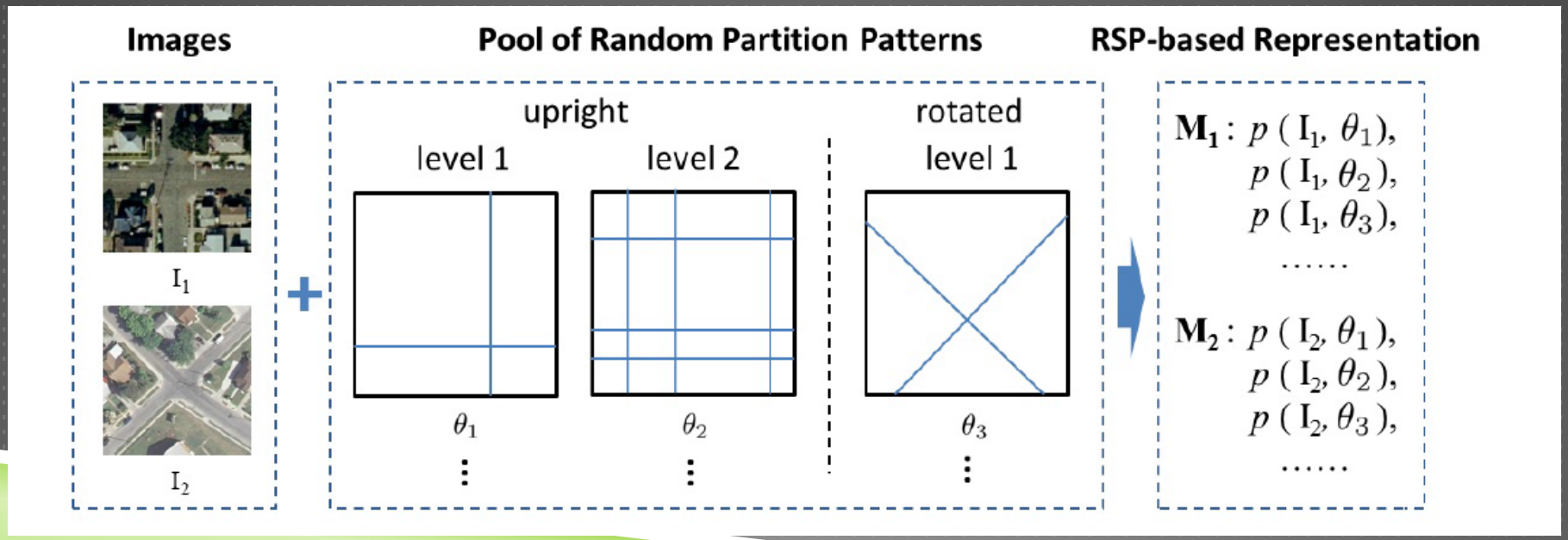
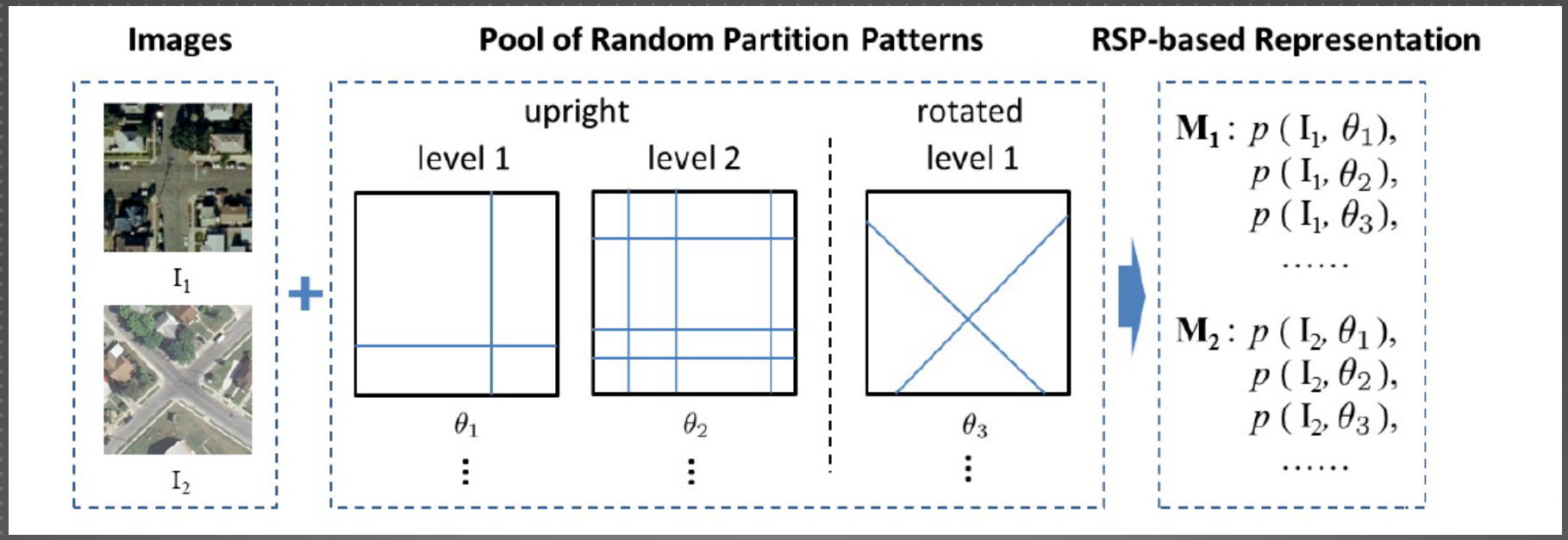


Figure Credit: Yuning Jiang, Junsong Yuan, Gang Yu

Choosing Right Pattern: Optimal Pattern Selection



- Find best pattern for each level with lowest error with SVM classification:

$$err_\theta^c = \sum_{c_v = c} \mathbf{I}(f_\theta^c(I_v) \neq c) + \sum_{c_v \neq c} \mathbf{I}(f_\theta^c(I_v) = c),$$

Figure Credit: Yuning Jiang, Junsong Yuan, Gang Yu

Choosing Right Pattern: Boosting

1. **For all** $\theta \in \Theta$:
   - Randomly sample a subset $\Phi_\theta \subset \Phi$, and represent the images in Φ_θ in pattern θ.
   - Train a multi-class classifier $f_\theta(\cdot)$ on the random subset Φ_θ using SVM.
2. Initialize the weight $w_i = \frac{1}{CN_{c_i}}$ for each images I_i, where N_{c_i} is the number of the images with label c_i; the current iteration number $j = 0$; the current accuracy $\sigma^{(0)} = 0$.
3. **While** $(\sigma^{(j)} < \sigma_{target})$
   - $\forall i = 1, \ldots, N$, $w_i \leftarrow \frac{w_i}{\sum_{i=1}^{N} w_i}$; $j \leftarrow j + 1$.
   - $\forall \theta \in \Theta$, calculate its classification error on Φ: $err_\theta = \sum_{I_i \in \Phi} w_i \cdot \mathbf{I}(f_\theta(I_i) \neq c_i)$.
   - Select the pattern $\theta^{(j)}$ with minimum error $err^{(j)}$, and then calculate the weight for $\theta^{(j)}$ as:
   $$\alpha^{(j)} = \log \frac{1 - err^{(j)}}{err^{(j)}} + \log(C - 1).$$
   - $\forall i = 1, \ldots, N$, $w_i \leftarrow w_i \cdot \exp(\alpha^{(j)} \cdot \mathbf{I}(f_{\theta^{(j)}}(I_i) \neq c_i))$.
   - Generate the strong classifier as:
   $$F(I) = \arg\max_c \sum_{m=1}^{j} \alpha^{(m)} \cdot \mathbf{I}(f_{\theta^{(m)}}(I) = c),$$
   and calculate its classification accuracy on Φ: $\sigma^{(j)} = \sum_{I_i \in \Phi} \mathbf{I}(F(I_i) = c_i)/N$.

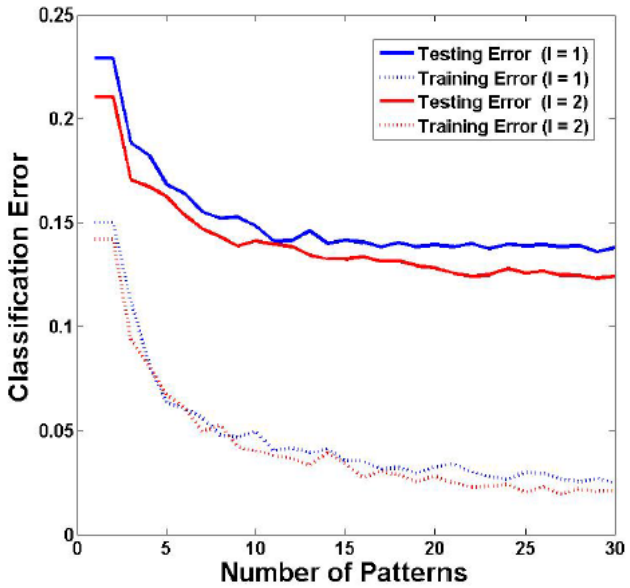
Algorithm Credit: Yuning Jiang, Junsong Yuan, Gang Yu

Experimental Results

- Error rates for 15-Scene dataset

level	Original SPM [11]	ORSP Classifier	BRSP Classifier
$l = 0$	74.8%	-	-
$l = 1$	78.8%	80.7%	86.4%
$l = 2$	79.7%	82.6%	87.1%
$l = 0, 1, 2$	81.4%	83.9%	87.2%

Table 1. Comparison with SPM at different levels on the 15-scene dataset.



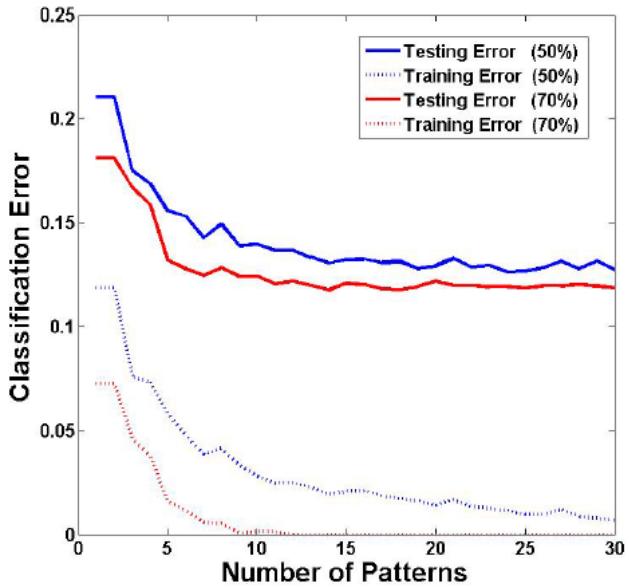


Figure Credit: Yuning Jiang, Junsong Yuan, Gang Yu

Experimental Results
15-scene Dataset

- 15-scene dataset

Method	Avg. Accuracy
SPM + SIFT with 400 clusters [11]	81.4%
SPM + SIFT with 400 concepts [28]	83.3%
DSP + SIFT with 1000 clusters [23]	80.7%
SP-pLSA + SIFT with 1200 topics [29]	83.7%
CENTRIST + RBF-SVM [12]	83.9%
CENTRIST + LCC + Boosting [27]	87.8%
RSP + Optimal Selection	83.9%
RSP + Boosting ($\lvert\Phi_\theta\rvert/\lvert\Phi\rvert = 50\%$)	87.2%
RSP + Boosting ($\lvert\Phi_\theta\rvert/\lvert\Phi\rvert = 70\%$)	**88.1**%

Figure Credit: Yuning Jiang, Junsong Yuan, Gang Yu

Experimental Results
15-scene Dataset – Cont.

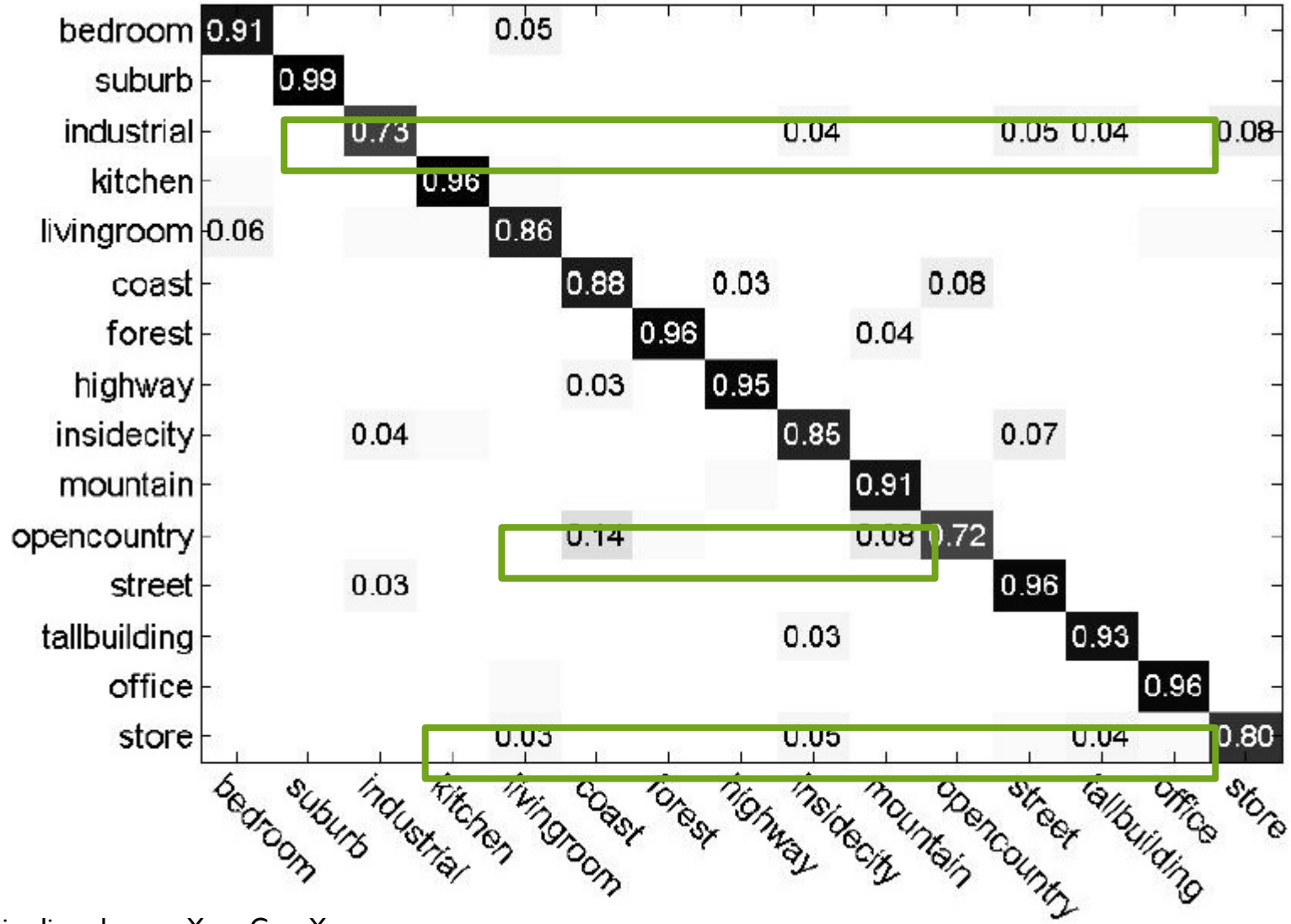


Figure Credit: Yuning Jiang, Junsong Yuan, Gang Yu

Mis-Classifications For 15-scene Dataset

- Images for coast-opencountry from 15-scene dataset
- Images for livingrom-bedroom from 15-scene dataset
- Images for insidecity-industrial from 15-scene dataset

Experimental Results
8-Scene Dataset

Method	Avg. Accuracy
Scene Model + SIFT [14]	$\approx 60\%$
Scene Model + Object Model + SIFT [14]	73.4%
PACT + RBF-SVM [12]	78.2%
SPM + RBF-SVM	74.0%
RSP + Optimal Selection	77.9%
RSP + Boosting	**79.6%**

Figure Credit: Yuning Jiang, Junsong Yuan, Gang Yu

Experimental Results
21-land-use Dataset

- Aerial images of scenes

	BoVW	SPMK [11]	SPCK [1]	SPCK+ [1]	SPCK++ [1]	ORSP	BRSP
Acc.	71.9	74.0	73.1	76.1	77.3	75.5	**77.8**

Figure Credit: Yuning Jiang, Junsong Yuan, Gang Yu