Fu-xiang Lu, Jun Huang, 2015. Beyond bag of latent topics: spatial pyramid matching for scene category recognition. *Frontiers of Information Technology & Electronic Engineering*, **16**(10):817-828. [doi:10.1631/FITEE.1500070]

Beyond bag of latent topics: spatial pyramid matching for scene category recognition

Key words: Scene category recognition, Probabilistic latent semantic analysis, Bag-of-words, Adaptive boosting

Contact: Fu-xiang Lu
E-mail: lufux@lzu.edu.cn
ORCID: http://orcid.org/0000-0002-5810-7631

Motivation/Main ideas

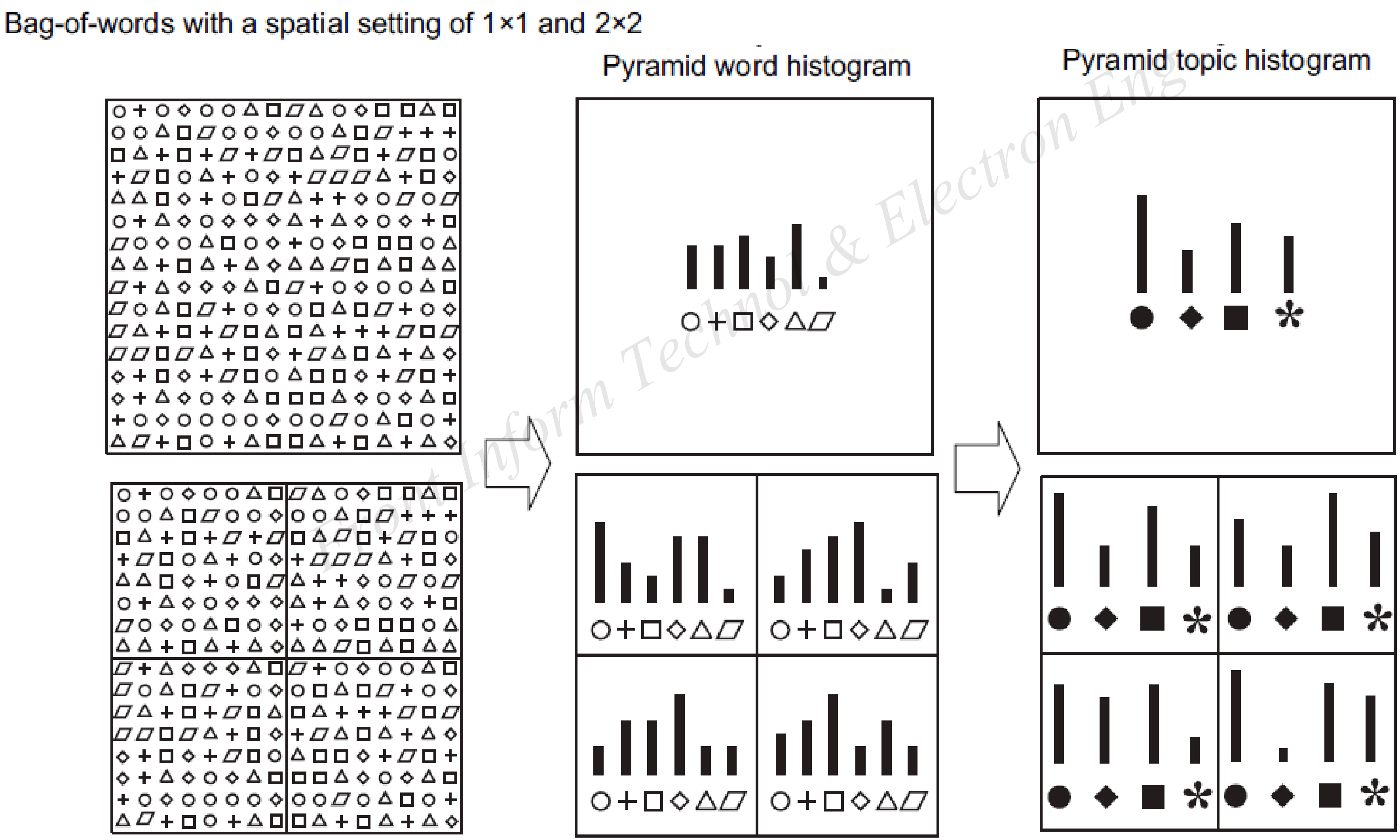
- **Motivation**

  Establish a robust method for the representation and subsequent recognition of scene categories, so as to achieve higher recognition accuracies in the case of a large number of scene categories.

- **Main ideas**
  - Introduce spatial position relationships among latent topics to probabilistic latent semantic analysis (pLSA).
  - Make performance improvement for real-world applications by combining various interest point detectors and region descriptors.
  - Use a two-stage framework to perform multi-class classification.

Method (I)

1. Main steps of computing the pyramid topic histogram for an image



Method (II)

2. Two-stage multi-class classification

Stage 1: For each of possible detector/descriptor pairs, use adaptive boosting classifiers to select the most discriminative topics and further compute posterior probabilities of a novel image from those selected topics.

Stage 2: use the prod-max rule to combine information coming from multiple sources and assigns the novel image to the scene category with the highest 'final' posterior probability. For M pyramid topic histograms and K classes, let $H_{mk}(\boldsymbol{x}^m)$ be AdaBoost classifiers learned from the training set by using Algorithm 3. Then, the prod-max rule is defined as follows:

$$y = \arg\max_{k \in \{1,2,\cdots,K\}} \prod_{m=1}^{M} H_{mk}\left(\boldsymbol{x}^m\right)$$

Major results (I)

Table 2 The average of per-category recognition rates over OT, LP, and LSP obtained using the pyramid topic histogram based on the grid detector and SIFT descriptor

D	Recognition rate (%)		
	OT	LP	LSP
1	81.4 ± 0.2	71.8 ± 0.8	65.8 ± 0.8
2	82.3 ± 0.7	75.0 ± 1.1	70.2 ± 0.1
3	84.6 ± 0.1	80.1 ± 0.9	75.4 ± 0.1

Major results (II)

Table 3 The average of per-category recognition accuracies obtained using each of the individual pyramid topic histograms and their combination for OT, LP, and LSP

Channal number	Recognition accuracy (%)		
	OT	LP	LSP
1	77.2 ± 0.6	69.1 ± 0.5	61.2 ± 1.5
2	83.4 ± 0.8	74.3 ± 0.6	69.2 ± 0.6
3	84.6 ± 0.1	80.1 ± 0.9	75.4 ± 0.6
4	81.0 ± 0.3	77.7 ± 1.0	75.5 ± 0.7
5	80.8 ± 0.6	74.9 ± 0.3	68.1 ± 1.3
6	80.5 ± 0.5	77.7 ± 0.9	75.5 ± 0.4
prod-max	$\mathbf{88.8} \pm 0.2$	$\mathbf{86.7} \pm 0.2$	$\mathbf{83.7} \pm 0.5$

Conclusions

- Pyramid topic histograms were proposed to represent an image.
  - Given an interest point detector and a local region descriptor, the pyramid histogram consistently outperforms standard pLSA.
  - Significant improvement can be made by combining various interest point detectors and local region descriptors involved in the bag-of-words representation.
- A two-stage framework was proposed to perform multi-class classification. The prod-max fusion rule employed in the framework works well in the case of a large number of scene categories.