Orientational Pyramid Matching for Recognizing Indoor Scenes

Lingxi Xie[1], Jingdong Wang[2], Baining Guo[2], Bo Zhang[1] and Qi Tian[3]
[1]Department of Computer Science and Technology, Tsinghua University, Beijing, China
[2]Microsoft Research, Beijing, China
[3]Department of Computer Science, University of Texas at San Antonio, Texas, USA

ABSTRACT

Scene recognition is a basic task towards image understanding. Spatial Pyramid Matching (SPM) has been shown to be an efficient solution for spatial context modeling. Although SPM is very efficient in scene recognition tasks, it still fails to discriminate categories with similar layouts.

In this paper, we introduce an alternative approach, Orientational Pyramid Matching (OPM), for orientational context modeling. Our approach is motivated by the observation that the 3D orientations of objects are a crucial factor to discriminate indoor scenes. The major novelty lies in that OPM uses the estimated 3D orientations to form the pyramid and produce the pooling regions, which is unlike SPM that uses the spatial positions to form the pyramid. Experimental results on challenging scene classification tasks show that OPM achieves the performance comparable with SPM and that OPM and SPM make complementary contributions so that their combination gives the state-of-the-art performance.

NOVELTY

In this paper, we propose a novel framework named Orientational Pyramid Matching (OPM) to incorporate the 3D orientation information into scene classification models.

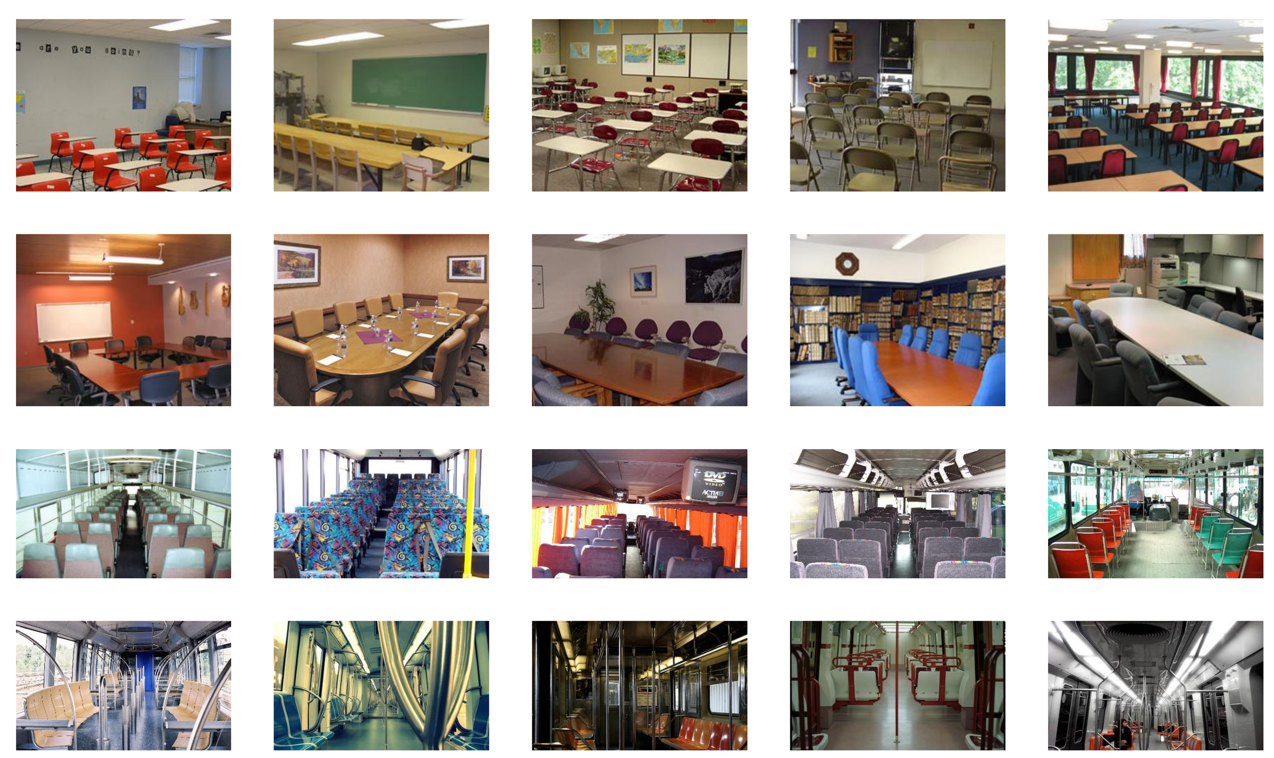
1. Our motivation comes from the observation that orientation of objects, e.g., *chairs*, is the main difference between some very similar scene categories, e.g., *classroom* vs. *meeting room*.
2. We propose to extract 3D orientational features using a similar manner of that in the SPM model. Each local descriptor is assigned with a 3D orientation and the image is partitioned into several regions based on the 3D orientation. Individual pooling process is performed on each region.
3. To extract 3D orientational information, we use a naïve yet efficient method [15]. It is a data-driven approach that assigns orientation to each local patch based on the nearest-neighbor regressor. The orientation in training images is labeled manually.

Experimental results on two challenging scene recognition datasets, i.e., Indoor-67 and SUN-397, verify that our algorithm achieves the state-of-the-art performance.

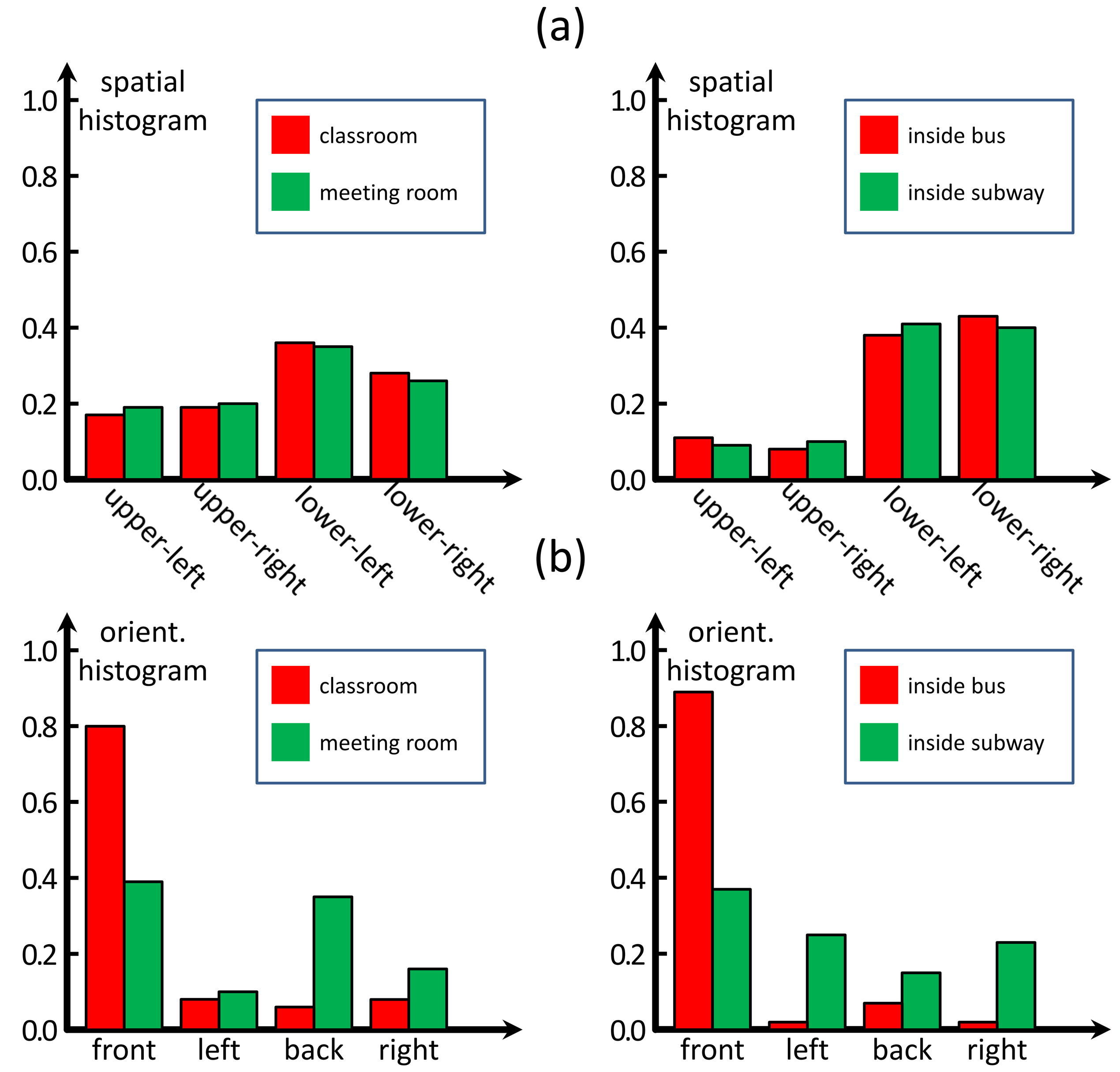
The success of our algorithm suggests that orientational features are indeed useful for scene recognition tasks, since it provides complementary information to spatial features. It also encourages the computer vision society to develop more accurate and efficient orientational assignment algorithms.

THE PROPOSED FRAMEWORK

Motivation



(a)

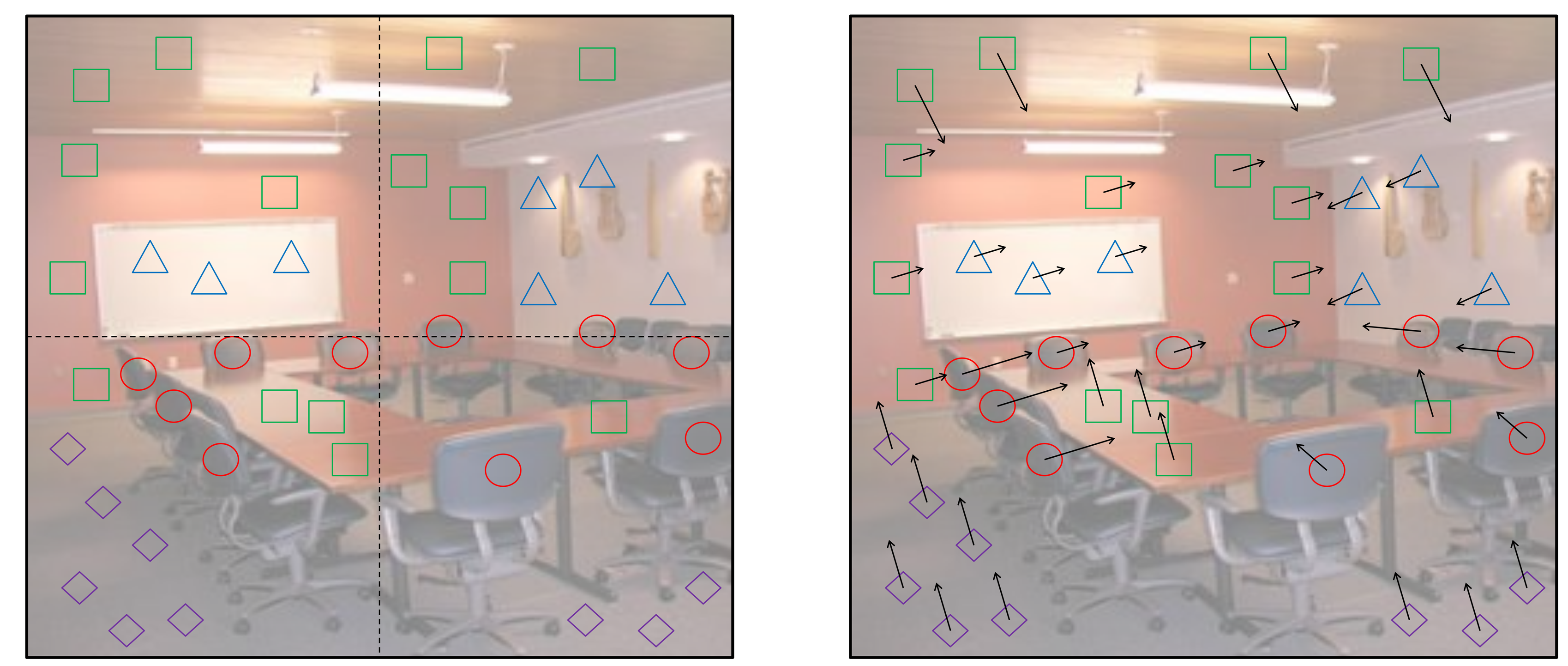


(b)

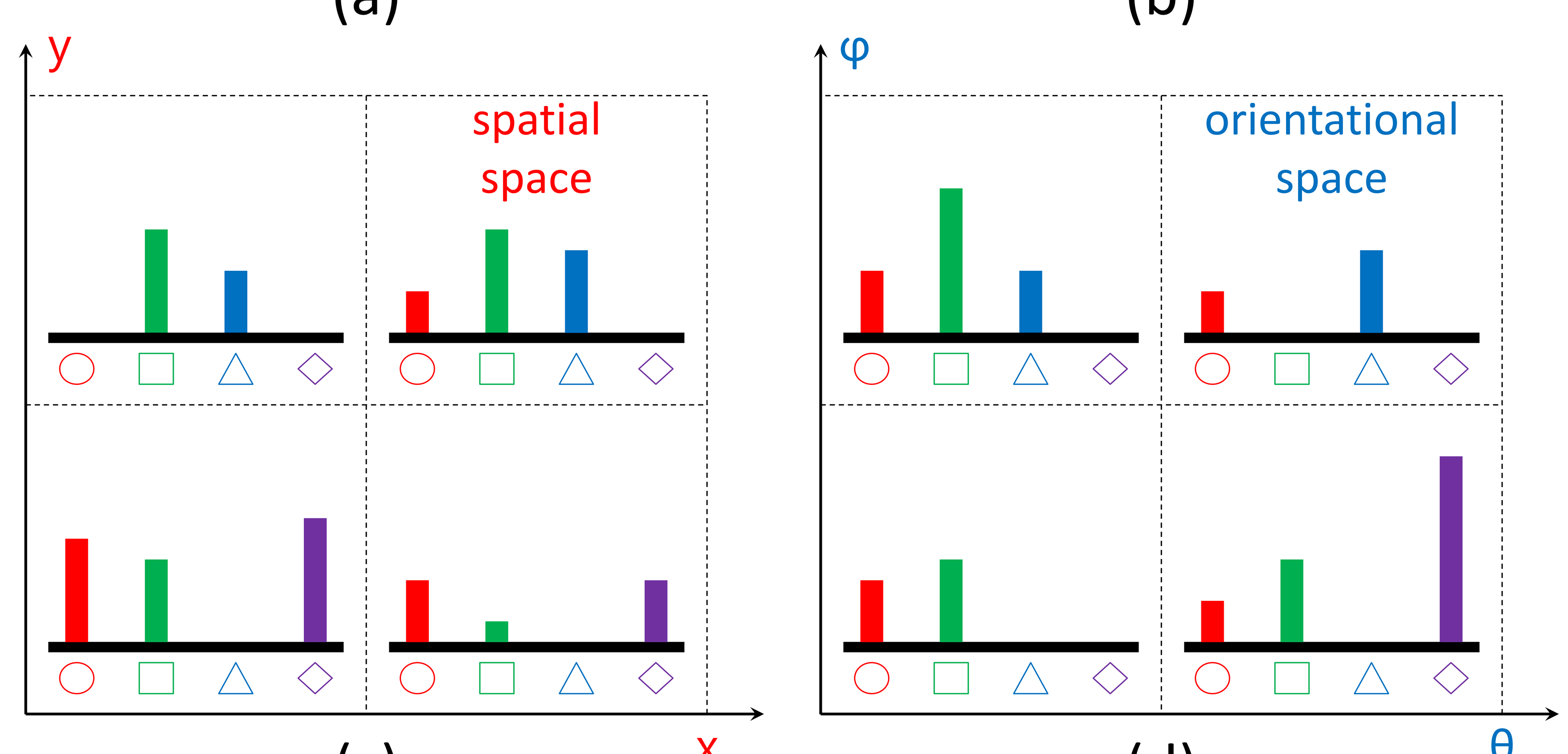
(c)

The orientation of ***chairs*** is the major difference between:
classroom vs. *meeting room*
bus carriage vs. *subway carriage*

Orientational Pyramid Matching



(a) (b)



(c) (d)

	bedroom	classroom	compt. room	hosptl. room	living room	meetn. room	office	waitn. room
bedroom	44.6	1.1	2.7	14.6	23.4	2.8	4.8	6.0
classroom	0.9	71.8	8.5	2.1	1.8	7.9	3.6	3.3
compt. room	2.6	15.6	44.1	9.4	5.9	6.8	8.2	7.4
hosptl. room	11.0	1.4	5.2	61.4	3.3	3.3	8.1	6.2
living room	18.2	4.6	3.4	5.0	51.0	3.6	8.8	5.5
meetn. room	3.6	12.2	9.1	11.0	5.5	45.5	3.9	9.3
office	5.9	9.0	19.3	8.6	16.2	8.3	27.2	5.5
waitn. room	8.6	5.6	10.0	13.1	10.3	9.4	8.9	34.6

	bedroom	classroom	compt. room	hosptl. room	living room	meetn. room	office	waitn. room
bedroom	52.7	0.6	1.2	11.7	22.6	1.8	4.1	5.3
classroom	0.9	80.9	5.8	0.6	3.0	4.8	0.6	3.3
compt. room	5.1	19.1	44.1	4.1	3.5	3.2	10.3	4.1
hosptl. room	6.2	1.4	2.4	71.4	0.5	7.1	5.2	5.7
living room	17.1	2.5	2.6	4.7	58.2	2.6	6.8	5.3
meetn. room	1.4	10.8	4.6	6.7	6.9	59.5	3.3	6.9
office	5.5	6.9	14.8	6.9	15.2	6.2	36.2	8.3
waitn. room	5.4	3.9	5.9	10.6	8.9	9.3	8.7	47.3

RESULTS

Indoor-67 Accuracy

	Quat. [32]	June. [23]	SPM(Sc)	OPM(Sc)	COMB(Sc)
	26.0	56.66	57.83	48.83	59.57

	Koba. [24]	June. [23]	SPM(FV)	OPM(FV)	COMB(FV)
	58.91	63.10	61.22	51.45	63.48

SUN-397 Accuracy

	Xiao [42]	Sanc. [34]	SPM(FV)	OPM(FV)	COMB(FV)
	38.0	43.2	43.58	34.61	45.91

Conclusions

We propose a novel Orientational Pyramid Matching (OPM) algorithm to capture the orientational contexts in the images, and combine the OPM features with SPM features to capture the complementary information for scene recognition. State-of-the-art classification performance is achieved on both MIT Indoor-67 and SUN-397 datasets. In the future, we will investigate the combination of OPM with many other approaches, and look forward to some more accurate orientation assignment algorithms to improve the OPM performance.

REFERENCES

Key references are numbered as they appear in the paper.
[15] O. Haines and A. Calway. Detecting Planes and Estimating their Orientation from a Single Image. BMVC, 2012.
[25] S. Lazebnik, C. Schmid, and J. Ponce. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. CVPR, 2006.
[27] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. IJCV, 2004.
[31] F. Perronnin, J. Sanchez, and T. Mensink, Improving the Fisher Kernel for Large-scale Image Classification. ECCV, 2010.
[40] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-Constrained Linear Coding for Image Classification. CVPR, 2010.
[44] L. Xie, Q. Tian, M. Wang, and B. Zhang. Spatial Pooling of Heterogeneous Features for Image Classification. TIP, 2014.

ACKNOWLEDGE.

This work was supported by the National Basic Research Program (973 Program) of China (Grant Nos. 2013CB329403, 2012CB316301 and 2014CB347600), the National Natural Science Foundation of China (Grant Nos. 61128007, 61332007, 61273023 and 91120011), the Beijing Natural Science Foundation (Grant No. 4132046), and the Tsinghua University Initiative Scientific Research Program (Grant No. 20121088071). This work was also supported in part to Dr. Qi Tian by ARO grant W911NF-12-1-0057, Faculty Research Awards by NEC Laboratories of America, and 2012 UTSA START-R Research Award, respectively.