Integrated Indoor-outdoor Scene Classification based on Local and Global Descriptors

Junsic Youn, and Sanghoon Sull
School of Electrical Engineering, Korea University
jsyoun@mpeg.korea.ac.kr and sull@mpeg.korea.ac.kr

Abstract

In this paper, we propose an indoor-outdoor scene classification method based on a systematic integration of a local shape vocabulary, a local color vocabulary and global color feature vectors. Existing methods based on the Bag-of-Words (BOW) use only local descriptors whereas other scene classification approaches use only global descriptors. However, we utilize both of local and global descriptors using decision-level fusion method. Experimental results on a variety of datasets are presented to demonstrate the improved performance of the proposed approach over the existing methods.

Keywords: scene classification, decision-level fusion.

1. Introduction

The BOW approach generally consists of three steps: local-level descriptor extraction, vocabulary construction, and an image representation using a vocabulary histogram. Since an image is described by a vocabulary histogram, only local-level descriptors are used in the BOW approach. Most scene classification methods are based on the SIFT descriptor [1] for the shape feature. Other type utilizes global-level descriptors (e.g. edge histogram) [2] to represent an image.

In this paper, we propose an indoor-outdoor scene classification method based on the systematic integration of a local shape feature vector, a local color feature vector, and a global color feature vector. The three feature vectors are fed separately into three classifiers providing three posterior probabilities. Finally, based on the decision- level fusion method, the color feature vector with the higher posterior probability between the local and global color feature vectors is probabilistically integrated with the local shape feature vector. It is demonstrated that the proposed method yields better performance than previous work.

2. BOW for Local-Level Descriptors and Clustering for Global-Level Descriptors

Fig. 1-(a) illustrates an overall procedure for learning. For indoor-outdoor scene classification, we use two different low-level features of shape and color. As the shape feature, a SIFT descriptor [1] is used because of its invariance to image scale and rotation. The top row of the Fig. 1-(a) illustrates the steps for a shape feature vector x_S. We extract the SIFT descriptor on both of keypoints and grid-sampled points. Then, a shape descriptor vocabulary is constructed through k-means clustering algorithm. The middle row of the Fig. 1-(a) shows the steps for a local color feature vector x_{LC}. We use the hue histogram for the color feature. The steps for the local color feature are similar to those of the shape feature. We also utilize a spatial layout to generate the local color descriptor vocabulary histogram x_{LC}. A global color feature vector x_{GC} is also generated by concatenating the hue histograms from the same layout as the local feature vector as shown in the third row in Fig. 1-(a). After x_S, x_{LC}, and x_{GC} are obtained, they are fed separately into three classifiers. We adopt the Support Vector Machine (SVM) classifier since it could provide an estimate of a posterior probability [3].

3. Decision-Level Fusion Strategy

We classify a scene by inferring the world state w that maximizes a posterior probability $P(w|x)$ given a feature vector x computed from an input image [4]. In this paper, we utilize three classifiers each of which provides a value of posterior probability. In order to properly combine the outputs of three classifiers, we use the decision-level fusion method [5]. Fig. 3-(b) shows the decision-level fusion strategy in our work. The max rule and the weight criterion are used to select

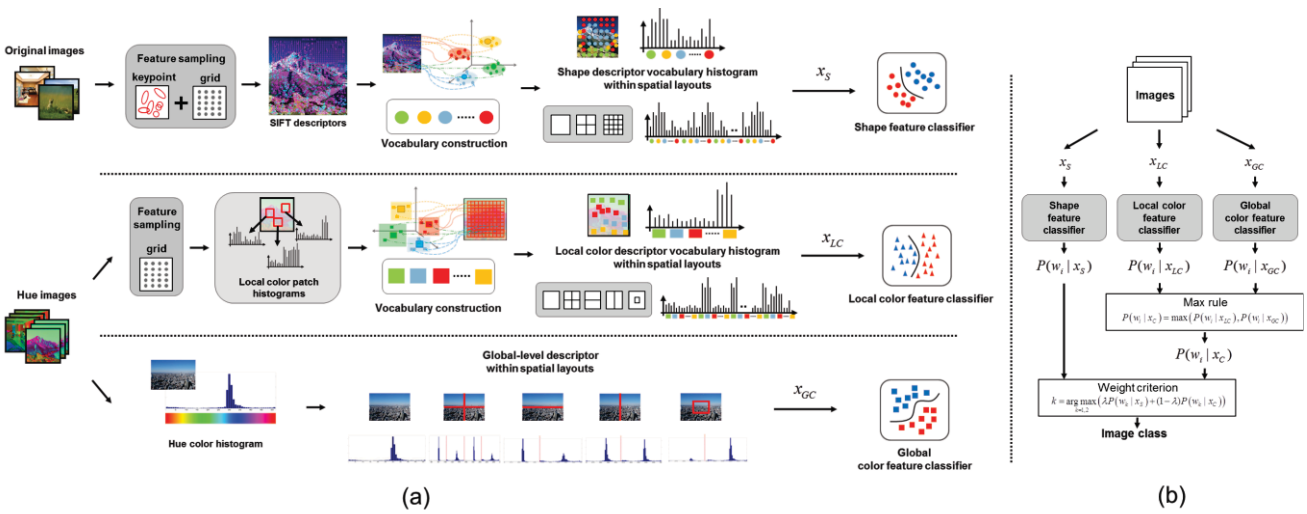

Figure 3: Overview of learning stage for indoor-outdoor scene classification. Top row: local shape feature. Middle row: local color feature. Bottom row: global color feature.

one color feature and classify an image, respectively. An image, $X=\{x_S, x_C\}$ is represented by the shape feature vector x_S and the color feature vector x_C. The more discriminating color feature vector for an image is selected using the max-rule between the local color feature vector x_{LC} and the global color feature vector x_{GC}. Finally, an input image X is classified as either indoor ($k=0$) or outdoor ($k=1$), $\{w_k\}_{k=1}^{2}$, by the weight criterion where the two probabilities from two classifiers are combined with the weight, λ.

4. Experimental Results

Our proposed method was tested on two different datasets: PASCAL VOC Challenge 2010 [7], SUN [8] which consist of various classes for scene classification. For the indoor-outdoor classification, we randomly sampled training and test images as an indoor and an outdoor class from the each datasets.

As in Table 1, the latest BOW approach [6] shows better performance compared with the other method [2]. However, the proposed method provides the improved overall performance over the two methods.

Table 1: Performance Comparison

Datasets	Class	Proposed	[6]	[2]
PASCAL VOC	In	**91.67**	85	80
	Out	92.5	**94.17**	80
	total	**92.08**	89.58	80
SUN	In	**94**	86	79
	Out	89	**93.5**	77
	total	**91.5**	89.75	78

5. Conclusion

In this paper we proposed an integrated indoor-outdoor scene classification method. Our method is based on a probabilistic integration of a local shape descriptor vocabulary histogram, a local color descriptor vocabulary histogram, and a global color histogram.

Acknowledgements

This research was funded and supported by Samsung Electronics Co., Ltd.

References

[1] D. G.Lowe. "Distinctive image features from scale-invariant keypoints." *IJCV*, 60(2):91-110, 2003.
[2] W. Kim, J. Park, and C. Kim, "A Novel Method for Efficient Indoor-Outdoor Image Classification." *Journal of Signal Processing Systems* 61(3): 251-258, 2010.
[3] Lin, H.T., Lin, C.J., Weng, R.C. "A note on platt's probabilistic outputs for support vector machines." *Machine Learning.* 68, 267–276, 2007.
[4] S. J. D. Prince. "Computer vision: models, learning, and inference." *Vol. 2. Cambridge University Press*, 2012.
[5] P. K. Atrey, M. A. Hossain, A. E. Saddik, and M. S. Kankanhalli. "Multimodal fusion for multimedia analysis: a survey." *Multimedia Systems,* 16(1):345-379, 2010.
[6] A. Bolobvinou, I. Pratikakis, and S. Perantonis. "Bag of spatio-visual words for context inference in scene classification." *Pattern Recognition*, 2012.
[7] M. Everingham, L. Van Gool, C. K. I.Williams, J.Winn, and A. Zisserman. "The PASCAL Visual Object Classes (VOC) challenge." *IJCV,* 88(2):303-338, Jun 2010
[8] T. Xiao, J. Hays, K. Ehinger, A. Oliva, and A. Torralba. "Sun database: Large-scale scene recognition from abbey to zoo." *In CVPR*, 2010