A New Bag of Words LBP (BoWL) Descriptor for Scene Image Classification

Sugata Banerji, Atreyee Sinha and Chengjun Liu
Department of Computer Science, New Jersey Institute of Technology, Newark, NJ.

College of Computing Sciences

Abstract

This paper explores a new Local Binary Patterns (LBP) based image descriptor that makes use of the bag-of-words model to improve classification performance for scene images. Experiments with three challenging image datasets show that the proposed BoWL descriptor yields significantly higher classification performance than LBP, and also results better than or at par with some other popular image descriptors.

Introduction

The LBP descriptor captures the variation in intensity between neighboring pixels [1], [2]. Lately, part-based methods have been very popular among researchers. Here the image is considered a collection of parts. After feature extraction, similar parts are clustered to form a visual vocabulary and a histogram of the parts is used to represent the image. This is known as a ”bag-of-words model” [3].

An Innovative Bag of Words LBP (BoWL) Descriptor for Scene Image Classification

Dense Sampling: The image is divided into a large number of equal sized overlapping blocks using a uniform grid and each block is used for feature extraction. This process is explained in Figure 1.

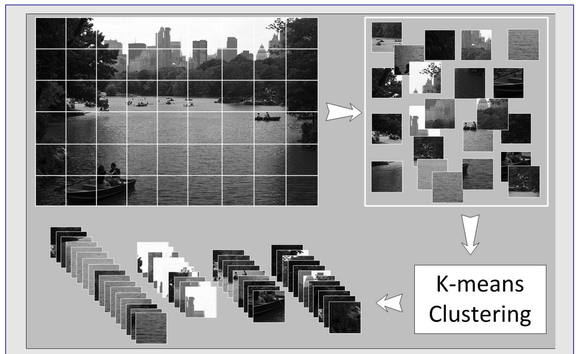

Fig. 1. A grayscale image is broken down into small image patches which are then quantized into a number of visual words and the image is represented as a histogram of words.

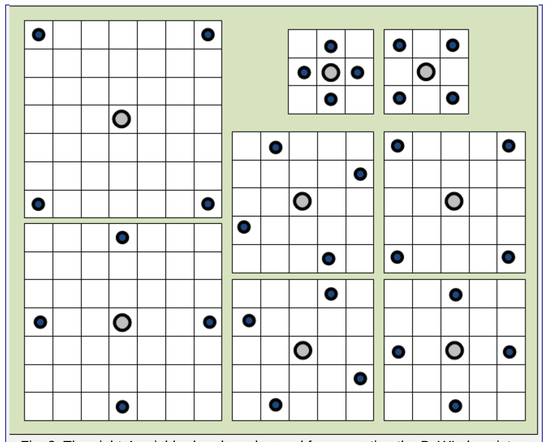

Fig. 2. The eight 4-neighborhood masks used for computing the BoWL descriptor.

A Modified LBP: Figure 2 shows the eight 4-pixel neighborhoods used for generating the multi-neighborhood LBP descriptor used here. Each of these neighborhoods produces a 16-bin histogram, and eight such histograms from different neighborhoods are concatenated to generate the 128-dimensional feature vector describing each image patch.

DCT Smoothing: The Discrete Cosine Transform (DCT) can be used to transform an image from the spatial domain to the frequency domain. In the proposed method, the original image is transformed to the frequency domain and the lowest 6.25%, 25% and 56.25% frequencies are used, respectively, to regenerate the image. The original image and the three images thus formed undergo the same process of dense sampling and eight-mask LBP feature extraction.

Quantization: The features are quantized into a visual vocabulary using K-means clustering.

We perform classification using a Support Vector Machine (SVM) classifier with a Hellinger kernel [4].

Experimental Results

The UIUC Sports Event dataset [5] contains 1,574 images from eight sports event categories. The MIT Scene dataset (also known as OT Scenes) [6] has 2,688 images classified as eight categories. The Fifteen Scene Categories dataset [7] is composed of 15 scene categories with 200 to 400 images in each. Figure 3 shows sample images from each dataset. Figures 4 and 5 show the improvement achieved by BoWL over LBP. Table I shows a comparison of our results with those obtained by other researchers.

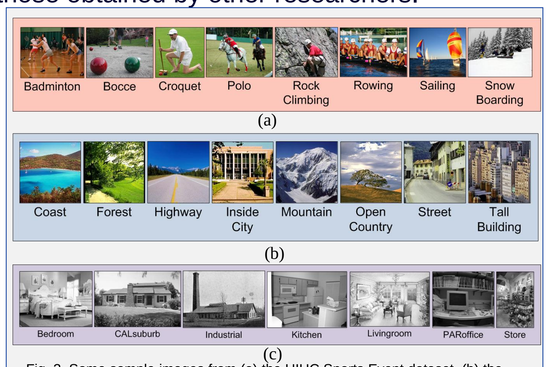

Fig. 3. Some sample images from (a) the UIUC Sports Event dataset, (b) the MIT Scene dataset, and (c) the Fifteen Scene Categories dataset

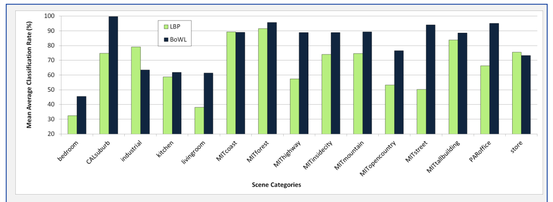

Fig. 4. The comparative classification performance of the LBP and the BoWL descriptors on the 15 categories of the Fifteen Scene Categories dataset.

TABLE I: Comparison with Other Methods on the Three Datasets (%)

Method	UIUC Sports Event	MIT Scene	Fifteen Scene
SIFT+GGM [5]	73.4	-	-
OB [8]	76.3	-	-
KSPM [9]	-	-	76.7
KC [10]	-	-	76.7
CA-TM [11]	78.0	-	-
ScSPM [9]	-	-	80.3
SIFT+SC [12]	82.7	-	-
SE [6]	-	83.7	-
HMP [12]	85.7	-	-
C4CC [13]	-	86.7	-
BoWL+SVM (Proposed)	**87.7**	**91.6**	**80.7**

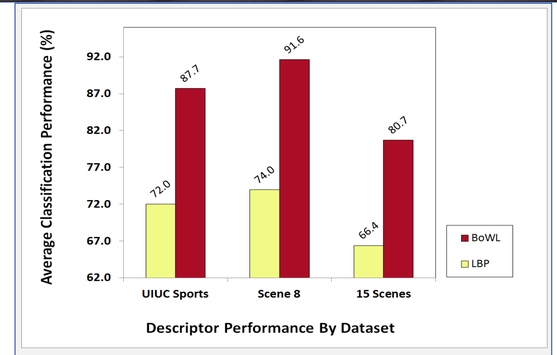

Fig. 4. Comparison of the mean average classification performance of the conventional LBP descriptor and the proposed BoWL descriptor using the SVM classifier on the three datasets

Conclusion

We proposed a new BoWL feature vector that enhances the popular LBP descriptor using a DCT and bag-of-words based representation. Experimental results on three large representative datasets show the supremacy of the proposed method over the traditional LBP method, and also some other popular methods for image category classification.

References

[1] Ojala, T., Pietikainen, M., Harwood, D.: A comparative study of texture measures with classification based on feature distributions. Pattern Recognition 29(1) (1996) 51–59

[2] Banerji, S., Verma, A., Liu, C.: Novel color LBP descriptors for scene and image texture classification. In: 15th International Conference on Image Processing, Computer Vision, and Pattern Recognition, Las Vegas, Nevada (July 18-21 2011) 537–543

[3] Yang, J., Jiang, Y., Hauptmann, A., Ngo, C.: Evaluating bag-of-visual-words representations in scene classification. In: Multimedia Information Retrieval. (2007) 197–206

[4] Vedaldi, A., Fulkerson, B.: Vlfeat – an open and portable library of computer vision algorithms. In: The 18th Annual ACM International Conference on Multimedia. (2010)

[5] Li, L.J., Fei-Fei, L.: What, where and who? classifying event by scene and object recognition. In: IEEE International Conference in Computer Vision. (2007)

[6] Oliva, A., Torralba, A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. International Journal of Computer Vision 42(3) (2001) 145–175

[7] Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: IEEE Conference on Computer Vision and Pattern Recognition, New York, NY, USA (2006)

[8] Li, L.J., Su, H., Xing, E.P., Fei-Fei, L.: Object bank: A high-level image representation for scene classification & semantic feature sparsification. In: Neural Information Processing Systems, Vancouver, Canada (December 2010)

[9] Yang, J., Yu, K., Gong, Y., Huang, T.: Linear spatial pyramid matching using sparse coding for image classification. In: IEEE Conference on Computer Vision and Pattern Recognition, Singapore (December 4-6 2009) 1794–1801

[10] Van Gemert, J., Veenman, C., Smeulders, A., Geusebroek, J.M.: Visual word ambiguity. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(7) (2010) 1271–1283

[11] Niu, Z., Hua, G., Gao, X., Tian, Q.: Context aware topic model for scene recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA (June 16-21 2012) 2743–2750

[12] Bo, L., Ren, X., Fox, D.: Hierarchical matching pursuit for image classification: Architecture and fast algorithms. In: Advances in Neural Information Processing Systems. (December 2011)

[13] Bosch, A., Zisserman, A., Munoz, X.: Scene classification via pLSA. In: The European Conference on Computer Vision. (2006)