

# Combining Local Binary Patterns for Scene Recognition

Minguan Song, Ping Guo\*

Image Processing and Pattern Recognition Laboratory,  
Beijing Normal University, Beijing 100875, China  
Email: mgsongbnu@gmail.com; pguo@ieee.org

**Abstract**—Recently, spatial principal component analysis of census transform histograms (PACT) was proposed to recognize instance and categories of places or scenes in an image. An improved representation called Local Difference Binary Pattern (LDBP) also was proposed and performed better than that of PACT. LDBP is based on the comparisons between center pixel and its neighboring pixels, but the relationship among neighbor pixels is not considered. In this paper, we propose to combine Local Neighbor Binary Pattern (LNBP) with LDBP to construct a spatial representation for scene recognition, because that LNBP can provide complementary information regarding neighboring pixels for LDBP. Experiments on widely used datasets demonstrate that the performance of image recognition is further improved with proposed method.

**Index Terms**—scene recognition, spatial pyramid matching, local binary pattern

## I. INTRODUCTION

Scene recognition is an important task in computer vision and has attracted considerable attention in recent years, it refers to the problem of recognizing the semantic category (e.g. bedroom, mountain, or coast) of a single image [1]. Scene recognition is widely used in many aspects, such as robotics path planning, video content analysis, content-based image retrieval, and video surveillance [2].

Compared with object recognition, scene recognition is more challenging because of ambiguity and variability in the content of scene images, which is further worsened by the variations in illumination and scale. Numerous efforts have been made to solve this kind of problem. Oliva and Torralba [3] proposed spatial envelope that represented the dominant spatial structure (naturalness, openness, roughness, expansion, ruggedness) of a scene, which achieved high accuracy in recognizing natural scenes. However, it performed bad about indoor scenes. Hoffman [4] put forward probabilistic latent semantic analysis (pLSA) model to perform probabilistic mixture decomposition. Bag of visual words (BoW) model [5, 6] becomes popular in recent years. BoW model represents an image as an unordered collection of local features, and has demonstrated impressive levels of performance [2]. But

the spatial information is neglected in BoW model. To improve the BoW model, Lazebnik *et al.* [7] incorporated spatial information by using spatial pyramid matching (SPM) scheme, and uses scale in variant feature transform (SIFT) [8] descriptor as the local feature. SIFT becomes the most popular descriptor in recent years [5, 7, 9, 10, 16, 19, 20, 21, 22, 23, 24, 25, 26]. Yang *et al.* [9] proposed linear SPM based on sparse coding (ScSPM) which developed an extension of the SPM method by generalizing vector quantization to sparse coding of SIFT descriptors, and followed by multi-scale spatial max pooling. ScSPM remarkably reduces the complexity of training and testing task. Gao *et al.* [19] proposed a Laplacian sparse coding method, which exploited the dependence among the local features to alleviate the sensitiveness of quantization. Gemert *et al.* [10] introduce visual word ambiguity to model a soft assignment instead of hard assignment, profiting in high-dimensional feature spaces and receive higher benefits when increasing the number of image categories.

Research on non-parametric nearest neighbor (NN) classification has also made progresses in past years. Boiman *et al.* [20] proposed a trivial NN-based classifier, which was called Naive-Bayes Nearest-Neighbor (NBNN). NBNN computes direct image-to-class distances without descriptor quantization. Wang *et al.* [21] learned metric for each class using Mahalanobis distance. Behmo *et al.* [22] relaxed the incremented assumption in NBNN and solve the parameter selection problem by hinge-loss minimization. Tuytelaars *et al.* [23] proposed the NBNN kernel which learned the classifier in a discriminative setting.

Although SIFT-based BoW model with SPM achieves remarkable performance, the computational complexity in both space and time is still a burden. Paris *et al.* [29, 30] combined Histogram of Local Binary Pattern with BoW, which outperforms SIFT-based methods. Recently, Wu and Rehg [1, 11] proposed spatial principal component analysis of census transform histograms (PACT), or Census Transform histogram (CENTRIST), which is an effective representation that fulfills the need for recognizing categories of places and scenes. CENTRIST captured local structures of an image by the Census Transform [13] and incorporated global structures with SPM. CENTRIST is superior to BoW model on scene recognition task for its simplicity and

\* Corresponding author, email: pguo@ieee.org

efficiency. Hu *et al.* [12, 28] utilize a multi-level kernel machine to alleviate the difference existing in various levels. Meng *et al.* [2] introduced local difference magnitude information as complement and built spatial Local Different Binary Pattern (LDBP) representation.

CENTRIST and LDBP have achieved excellent performances. However, both of them are based on the comparisons between center pixel and neighbors; the relationship among neighbors is ignored. Under a large amount of conditions, there exist different local structures that have the same LDBP code. In such cases, the different local patterns are not clearly represented and cannot be differentiated. Some important information with respect to edges and gradients is lost. This information is significant for describing the structure of scenes. Consequently, different patterns are assigned into the same category, which decreases the discriminative power. To address this problem, Local Neighbor Binary Pattern (LNBP) [27] was proposed as an extension of local binary pattern. LNBP is a complement of CENTRIST and LDBP for describing local structures, and therefore could improve scene recognition task. We therefore propose to combine LDBP with LNBP, and it can preserve the advantages - easy to implement, nearly no parameter to tune and fast to evaluate.

The rest of this paper is organized as follows. Section II briefly describes spatial PACT and spatial LDBP. In Section III we introduced LNBP, and our proposed image representation is presented. In section IV, Experimental results on common datasets are given. A conclusion of this paper is drawn in Section V.

## II. RELATED WORKS

### A. Census Transform and Spatial PACT

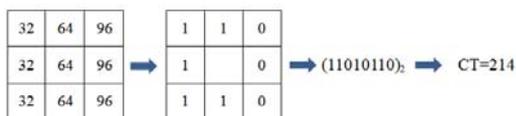


Figure 1. The Census Transform operation

Census Transform (CT) is a non-parametric local transform originally designed for establishing correspondence between local patches [13]. Census transform compares the intensity value of a pixel with its eight neighboring pixels, as illustrated in Fig. 1. If the center pixel is bigger than (or equal to) one of its neighbors, a bit 1 is set in the corresponding neighbor location. Otherwise a bit 0 is set, which only has a different bit order from the local binary pattern (LBP) code  $LBP_{8,1}$  [18]:

$$CT = \sum_{i=0}^{P-1} g(I_i, T) \cdot 2^i, \quad (1)$$

$$g(x, T) = \begin{cases} 1, & x \leq T \\ 0, & x > T \end{cases}, \quad T = I_c,$$

where  $P$  is the number of neighboring pixels, *i.e.* 8.

Census transform is robust to illumination changes, gamma variations, etc. As a visualization method, a census transformed image is created by replacing a pixel with its CT value. Shown by the example in Fig. 2, the census transform retains global structures of the picture (especially discontinuities) besides capturing the local structures.

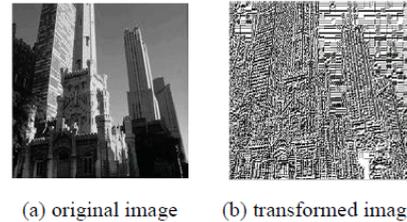


Figure 2. An example of census transformed image.

One important property of the transform is that CT values of neighboring pixels are highly correlated [1]. In the example of Fig. 3, the Census Transform for pixels valued 36 and 37 are depicted in right, and the two circled bits are both comparing the two center pixels (in different orders). Thus the two bits must be strict complement to each other if the two pixels are not equal. More generally, bit 5 of  $CT(x, y)$  and bit 4 of  $CT(x + 1, y)$  must be complementary to each other, if the pixels at  $(x, y)$  and  $(x+1, y)$  are not equal. Generally, there are eight such constraints between one pixel and its eight neighboring pixels.



Figure 3. Illustration of constraints between CT values of neighboring pixels.

Besides these deterministic constraints, there also exist indirect constraints. For example, in Fig. 3, the pixel valued 32 compares with both center pixels when computing their CT values (bit 2 of  $CT(x, y)$  and bit 1 of  $CT(x + 1, y)$ ). Depending on the comparison results between the center pixels, there are probabilistic relationships between these bits.

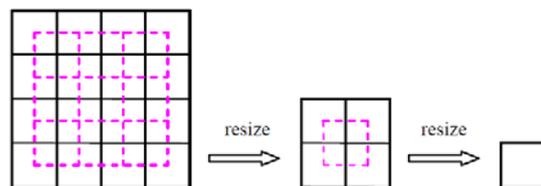


Figure 4. Illustration of the level 2, 1, and 0 spatial pyramid split of an image, from [1].

Wu *et al.* [1, 11] proposed Principal component Analysis of Census Transform histograms (PACT), that is, the principal component analysis (PCA) operation performs on the CT histograms, to remove these correlation effects, and to get a more compact representation. Because PACT can only encode global shape structure in a small image patch, in order to

capture the global structure of an image in larger scales, a spatial PACT representation based on the SPM scheme was proposed. A spatial pyramid, which divides an image into segments and concatenates correspondence results in these regions, encodes roughly spatial structure of an image and usually improves recognition rate. The level 2 split in a spatial pyramid divides the image into  $4 \times 4 = 16$  blocks. They also shift the division (dash line blocks) in order to avoid artifacts created by the non-overlapping division, which makes a total of 25 blocks in level 2. Similarly, level 1 and 0 have 5 and 1 block, respectively. The image is resized between different levels so that all blocks contain the same number of pixels. These blocks are shown in Fig. 4. PACT in all blocks is then concatenated to form an overall feature vector of 1240 dimensional.

**B. Local Difference Binary Pattern and Spatial LDBP**

Census transform concerns whether a center pixel is higher or lower than its neighboring pixels, resulting in some information loss of intensity contrast. It is not enough to discriminate different local structures using only census transform [2]. Meng *et al.* introduced local difference to better describe local structures.

Local difference is defined as the intensity difference between a center pixel and its neighboring pixels in a  $3 \times 3$  image patch. Given a center pixel  $I_c$  and its neighbors  $I_i$ ,  $i=0,1, \dots, 7$ . The local difference between  $I_c$  and  $I_i$  can be computed by  $d_i = I_c - I_i$ . Then the local difference  $d_i$  is decomposed into two components:

$$d_i = s_i \cdot m_i; s_i = \begin{cases} 1 & d_i \geq 0 \\ -1 & d_i < 0 \end{cases}; m_i = |d_i|, \quad (2)$$

where  $s_i$  is the sign and  $m_i$  is the magnitude of  $d_i$ . Obviously, the sign and magnitude components contain complementary information of original local difference.

The sign and magnitude components are both converted into binary codes. The positive and negative elements in sign component are coded as 1 and 0. The 8-bit code is converted into a base-10 number called Local difference Sign Binary Pattern (LSBP). LSBP is equivalent to Census Transform.

The Local difference Magnitude Binary Pattern (LMBP) is defined as follows:

$$LMBP = \sum_{i=0}^{P-1} g(m_i, T) \cdot 2^i, \quad (3)$$

$$g(x, T) = \begin{cases} 1, & x \geq T \\ 0, & x < T \end{cases}, \quad T = \frac{1}{NP} \sum_{j=1}^N \sum_{i=0}^{P-1} m_{ij},$$

where  $m_{ij}$  is  $m_i$  of the  $j$ th pixel,  $N$  is the number of pixels (excluding boundaries) in Image, and  $T$  is the mean  $m_i$  of the whole image. Finally, both LSBP and LMBP transform a  $3 \times 3$  image block into an integer in  $[0,$

255]. The coding process of LSBP and LMBP is shown in Fig. 5.

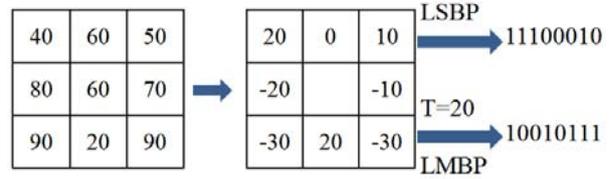


Figure 5. Coding process of LSBP and LMBP.

As shown in Fig. 6, the LMBP values of neighboring pixels are highly correlated. Similar to CT, Bit5 of LMBP at  $(x, y)$  and bit 4 of LMBP at  $(x+1, y)$  must be the same. There are eight such constraints between one pixel and its eight neighboring pixels. Applying the constraints to all pixels of an image, we can conclude that the number of pixels whose LMBP value's bit 5 is 1 must be equal to the number of pixels whose LMBP value's bit 4 is 1, and vice versa.

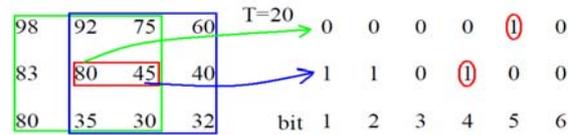


Figure 6. Illustration of constraints between LMBP values of neighboring pixels.

For visualization, a LMBP transformed image is created by replacing a pixel with its LMBP value. Shown by the example in Fig. 7, the LMBP transform also retains global structures of the picture.

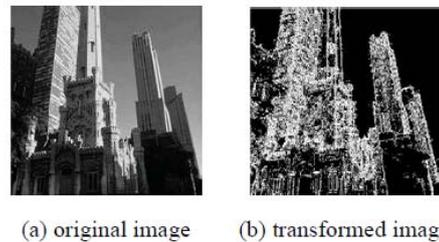


Figure 7. An example of LMBP transformed image

Because bins are strong correlated with each other in LSBP and LMBP, PCA is utilized to reduce the dimensionality. The LSBP and LMBP histograms perform PCA separately and then are concatenated to form the final feature vector, namely LDBP histogram. PACT only uses the sign component of LDBP.

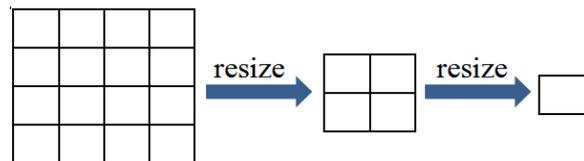


Figure 8. Illustration of the level 2, 1, and 0 spatial pyramid split of an image.(referred from [2])

The spatial pyramid scheme [2] is given in Fig. 8, the level 2 split in a spatial pyramid divides the image into 16 non-overlapping blocks. Similarly, level 1 has 4 blocks and level 0 has 1 block, respectively. The image

is resized between different levels so that all blocks contain the same number of pixels. In total, there are 21 blocks for each image. The vector representations in all blocks are concatenated to form an overall feature vector for each image as the global image representation. The dimension of final feature vector is 840.

### III. PROPOSED METHOD

#### A. Local Neighbor Binary Pattern

LDBP are based on the comparisons between center pixel and its neighboring pixels, both the sign part and magnitude part. The relationship between neighbor pixel and center pixel is well described. However, there is some information loss concerning the relationship of neighbor pixels because the relationship of neighbor pixels is neglected.

Under a large number of circumstances, there exist different local structures that have the same LDBP code. In such cases, the local pattern is not clearly represented and cannot be differentiated. According to our statistics from popular datasets, there are more than 15% pixels in scene images belonging to this category. These local structures may contain important information for scenes, for example, edges and gradients; it is also possible that some of them are smooth areas. Nevertheless, in LDBP, these distinct patterns are treated as the same one. LDBP has no discriminative ability for these local structures. Therefore, it is necessary to distinguish these different local structures that LDBP is not capable to differentiate with. A new descriptor is demanded to characterize such local patterns.

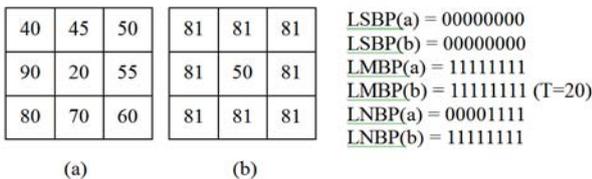


Figure 9. Example of complementary information from LNBP.

To address this problem, we propose an extension of local binary pattern, which only reflects local structure of neighboring pixels. The Local Neighbor Binary Pattern (LNBP) is defined as follows:

$$LNBP = \sum_{i=0}^{P-1} g(I_i, T) \cdot 2^i, \tag{4}$$

$$g(x, T) = \begin{cases} 1, & x \geq T \\ 0, & x < T \end{cases}, \quad T = \frac{1}{P} \sum_{i=0}^{P-1} I_i,$$

LNBP is also computed on a 3×3 local neighborhood. Note that LNBP is different with Multimodal Invariant Local Binary Pattern (MILBP) [31], Local Gradient Pattern (LGP) [32], Modified Census Transform (MCT) [14], and Improved LBP (ILBP) [15]. Both MILBP and LGP use the absolute value of difference, therefore we cannot know whether the intensity of one pixel is higher than other ones. However, LNBP compares neighbor

pixels with the mean of neighbors. Therefore we can know the relations of some neighbor pixels. This is the advantage of LNBP over MILBP or LGP. Compared with MCT and LBP, The center pixel is discarded to eliminate the influence of its intensity so that we can concentrate on the relationship of neighboring pixels. The threshold is set as the mean intensity value of eight neighbor pixels. Because the threshold is only related to neighboring pixels, the neighbor pattern can be described better. The local patch is transformed into an integer in [1, 255]. Note that 0 is not possible for that at least one neighbor pixel is larger than or equal to the mean value.

LNBP can provide useful complementary information for LSBP and LMBP. Fig. 9 shows an example which abounds in scene images. The patterns of the two example patches are different. However, they share the same LSBP and LMBP code. In other words, we cannot discriminate these patches only by LSBP and LMBP. With our LNBP code, the patches can be distinguished as different pattern. In the example we can know that the neighbors of block (b) are the same, while the neighbors of block (a) are different in intensities. LNBP can be used to describe the local structures about which

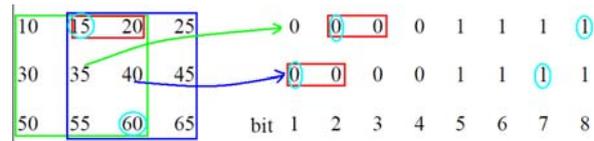


Figure 10. Illustration of constraints between LNBP values of neighboring pixels.

neighbors are larger than others. This also provides some information about gradient— how the intensity changed and the direction of gradient. LDBP may also provide this information for some local structures. But LDBP does not help much for the patches like those in Fig. 9. LNBP can always tell us such information. If the LNBP codes are all of 1, it tells us that this patch has neighbors having the same intensity. Therefore LNBP can provide useful complementary information for LDBP.

There are correlations between LNBP values. As shown in Fig. 10, when 15 and 20 are both smaller than the mean intensity in the neighboring patches, the bit 2 and bit 3 of the left patch are the same as well as the bit 1 and bit 2 in the right patch. Moreover, when the mean value is between 15 and 60, the bit 2 and bit 8 of the left block are identical to the bit 1 and bit 7 of the right block.

The transitive property of such constraints also makes them propagate to not only neighbor pixels, but also further ones. For example, in Fig. 10, the pixels valued 15 (coded as 0) and 55 (coded as 1) can be compared using various paths of comparisons. One path is 15 < 20 < 30 < 40 < 50 < 55, the other is 15 < 20 < 25 < 35 < 45 < 55. Similarly, although no deterministic comparisons can be deduced between some pixels (e.g. 55 and 60), probabilistic relationships can still be obtained. The propagated constraints make LNBP values and histograms implicitly contain information for describing global structures, just as CENTRIST.

To visualize the algorithm’s effect, the LNBP transformed images are created by replacing a pixel with its LNBP transformed value. The examples shown in Fig. 11 demonstrate that LNBP, as well as LSBP and LMBP, not only captures local structures, but also retains the global structural information, especially discontinuities.

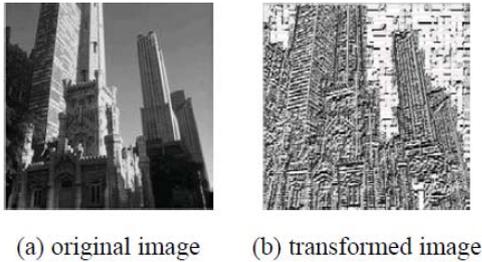


Figure 11. An example of LNBP transformed image.

**B. Spatial Image Representation by Combining LDBP with LNBP**

A histogram of LNBP for an image or image block is computed. We kept the bin of zero, whose value is always 0. The dimension of LNBP histogram is 256. For each pair of adjacent pixels, they share four neighbors. The bins of LNBP histograms are implicitly correlated with each other. PCA was used to reduce the dimensionality of LNBP histogram to 20. Then the LDBP histogram was concatenated with the compressed LNBP histogram, and the dimension of the final feature vector is 60.

In our experiments the performance of 40 eigenvectors (an average of 13 or 14 eigenvectors for LSBP, LMBP and LNBP) was also evaluated. According to [2], when the number of eigenvectors is smaller than 20, the performance drops dramatically. In such case, the recognition rate was almost the same as spatial LDBP. Compared with 20 eigenvectors for both LSBP and LMBP, when there are only 13 or 14 eigenvectors for LSBP, LMBP, and LNBP, the benefit provided by LNBP is counteracted by the information loss of dimension reduction.

The spatial pyramid matching scheme in Fig. 8 is adopted. The final feature vector for an image is 1260 dimension. We used support vector machine (SVM) for classification.

**IV. EXPERIMENTS**

In this section, our approach is evaluated on three benchmark datasets: (1) 8 scene categories dataset [3], (2) 15 class scene category [7], and (3) 8 class sports event [16]. In each dataset, the available data are randomly split into a training set and a testing set following the published protocols on these datasets. The random splitting is repeated 5 times, and the average accuracy and standard deviation is reported.

In the experiments, following the same experiment procedure of the CENTRIST [1], only the intensity values and ignore color information was used. We

normalized the LSBP, LMBP, and LNBP histograms and PCA eigenvectors such that they have zero mean and unit norm. LIBSVM [17] was utilized as the classifier and Radial Basis Function (RBF) kernel with recommended parameters  $(C, \gamma) = (8, 2^{-7})$  in [1] was adopted.

**A. The 8 Class Scene Category Dataset**



Figure 12. Sample images of 8 scenes. The categories are suburb, industrial, coast, forest, highway, inside city, mountain, open country, street, and tall building, respectively (from left to right, and from top to bottom).

TABLE I.  
RECOGNITION RATES ON THE 8 CALSS SCENE DATASET

Method	Rates(%)
LSBP [2]	75.53
LDBP [2]	79.18
LDBP+LNBP	81.36

The 8 class scene category dataset contains total 2688 images with 8 outdoor categories. Images are 256x256 in resolution, varying from 260 to 360 images in each category. It is a subset of the 15 scene category dataset. These categories are coast (360 images), forest (328 images), mountain (274 images), open country (410 images), highway (260 images), inside city (308 images), tall building (356 images), and street (292 images). Fig. 12 gives example images of the 8 categories. This dataset is used to investigate the usefulness of LNBP. Different schemes, include LSBP, LDBP, and LDBP

combined with LNBP are compared. No PCA operation or SPM scheme were used on this dataset.

In the experiments, 100 images are drawn in each category for training, and the remaining images for testing. The SVM with RBF kernel was utilized to classify the images. The recognition results are shown in Table I. We can see that LNBP provides useful information and improves the recognition.

B. The 15 Class Scene Category Dataset



Figure 13. Sample images of 15 scenes. The categories are bedroom, suburb, industrial, kitchen, living room, coast, forest, highway, inside city, mountain, open country, street, tall building, office and store, respectively (from left to right, and from top to bottom).

The 15 class scene category dataset contains total 4485 images with 15 categories. Images are about 300x250 in average resolution, varying from 210 to 410 images in each category. This dataset contains a wide range of scene categories in both indoor and outdoor environments (scene classes including office, store, coast, etc. Fig. 13 gives example images of all 15 categories). Also, this is one of the most complete scene category dataset used in the literature so far.

According to previous works [1, 2, 11], the first 100 images in each category are used to calculate the PCA eigenvectors. The recognition results are shown in Table II. From that table we can see that our method achieves the highest recognition rate.

TABLE II. RECOGNITION RATES ON THE 15 CALSS SCENE DATASET

Method	Feature Type	Rates(%)
SPM [7]	400 cluster centers	81.40 ± 0.50
ScSPM [8]	400 cluster centers	80.28 ± 0.93
Spatial PACT [1]	CENTRIST, 40 eigenvectors	83.88 ± 0.76
Spatial LDBP [2]	LDBP, 40 eigenvectors	83.58 ± 0.99
Our method	LDBP+LNBP, 60 eigenvectors	<b>84.09 ± 0.35</b>

Fig. 14 shows a confusion matrix from one run on this dataset using our approach. The biggest confusion happens between category pairs such as bedroom/living room, industrial/store, and coast/open country, which coincide well with the confusion distribution in [1, 2, 7, 11]. Our method achieves high recognition rates on forest, office, and suburb.

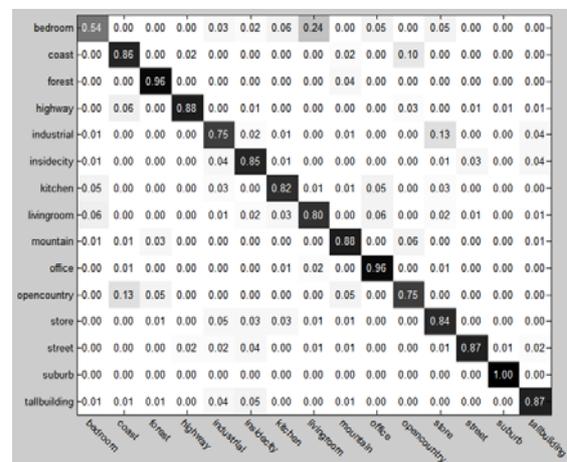


Figure 14. Confusion matrix from one run for 15 class scene category dataset recognition experiment.

TABLE III. RESULTS OF DIFFERENT FEATURE AND CLASSIFIER ON THE 15 CALSS SCENE DATASET

Feature Type	Classifier	Rates(%)
LDBP+LNBP, 60 eigenvectors	Non-linear	84.09
LDBP+LNBP, 60 eigenvectors	Linear	83.02
LDBP+LNBP, non-PCA	Linear	72.24

Linear SVM classifiers are also applied to the scene dataset to test the PCA compact feature and the non-PCA concatenated histograms. The results are shown in Table III. The compact feature achieves the accuracy of 83.02%, obviously higher than 81.8% of spatial LDBP [2]. The difference in performance of RBF kernels and linear kernels is quite small. However, the performance of original non-PCA concatenated histogram on linear classifier is very poor. Therefore it is the PCA operation that turns the histograms into compact features. PCA is necessary. Because of the fast testing speed of linear classifiers and small performance difference, linear SVM classifiers could be used to ensure real-time classification.

C. The 8 Class Event Dataset



Figure 15. Sample images of 8 sport events. The categories are badminton, bocce, croquet, polo, rock climbing, rowing, sailing, and snowboarding (from left to right, top to bottom).

The event dataset contains images of eight sports:

TABLE IV. RECOGNITION RATES ON THE 8 CLASS EVENT DATASET

Method	Feature Type	Rates(%)
Spatial PACT [1]	CENTRIST, 40 eigenvectors	78.25 ± 1.27
Spatial LDBP [2]	LDBP, 40 eigenvectors	82.96 ± 1.51
Our method	LDBP+LNBP, 60 eigenvectors	<b>83.66 ± 0.93</b>

badminton, bocce, croquet, polo, rock climbing, rowing, sailing, and snowboarding. The images have high resolution (from 800x600 to thousands of pixels per dimension). The number of images in each category varies from 137 to 250. We used this dataset for scene recognition purposes only. Fig. 15 shows the example images.

Also, following the previous works of spatial PACT and spatial LDBP, we randomly select 70 images per category for training, and 60 ones for testing. The

training images in each train/test split are used to compute the eigenvectors. The recognition results are shown in Table IV. Our method achieves best results outperforming spatial PACT and spatial LDBP.

We also compared the PCA feature with non-PCA concatenated histograms on linear classifier, and the results are similar to the experiment on the 15 scenes datasets, as is shown in Table V.

TABLE V. RESULTS OF DIFFERENT FEATURE AND CLASSIFIER ON THE 8 CLASS EVENT DATASET

Feature Type	Classifier	Rates(%)
LDBP+LNBP, 60 eigenvectors	Non-linear	83.66
LDBP+LNBP, 60 eigenvectors	Linear	81.88
LDBP+LNBP, non-PCA	Linear	68.81

Fig. 16 shows the confusion matrix of one run on this dataset. The biggest confusion happens between bocce and croquet, which is coincident with previous works. From the example images in Fig. 15 we can see that the two categories are very similar in human vision. High recognition rates are achieved for rock climbing, rowing and sailing categories.

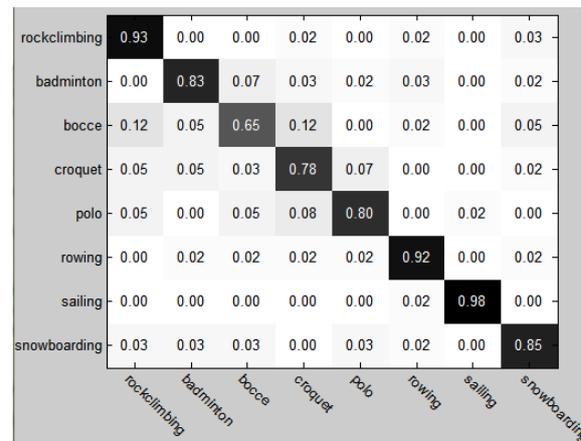


Figure 16. Confusion matrix from one run for 8 class event dataset recognition experiment .

V. CONCLUSION

In this paper, we propose to combine LNBP with LDBP to build an effective representation for scene images. LNBP provides extra complementary information with respect to local neighbor structures, which is neglected in LDBP. The new feature provides stronger discriminative ability for local structures in improving scene recognition. Experiments conducted on common benchmark datasets demonstrate that our proposed scheme outperforms spatial PACT or spatial LDBP on scene recognition task. Moreover, proposed method preserves the advantages of spatial PACT and spatial LDBP. It is easy to implement, has nearly no parameter to tune. In all the datasets we experimented with, the difference in recognition rates between these two kernel types are less than 2%. This indicates that

images from the same category are compact in the feature space. It works well on linear classifiers, thus is very fast for evaluation.

#### ACKNOWLEDGMENT

The research work described in this paper was fully supported by the grants from the National Natural Science Foundation of China (Project No. 90820010, 60911130513).

#### REFERENCES

- [1] J. Wu and J. M. Rehg, "CENTRIST: A Visual Descriptor for Scene Categorization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1489-1501, 2011.
- [2] X. Meng, Z. Wang and L. Wu, "Building Global Image Features for Scene Recognition," *Pattern Recognition*, vol. 45, pp. 373-380, 2012.
- [3] A. Oliva and A. Torralba, "Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope," *International Journal of Computer Vision*, vol. 42, no. 3, pp. 145-175, 2001.
- [4] T. Hofmann, "Unsupervised Learning by Probabilistic Latent Semantic Analysis," *Machine Learning*, vol. 41, pp. 177-196, 2001.
- [5] P. Quelhas, F. Monay, J. M. Odobez, D. Gatica-Perex, T. Tuytelaars and L. Van Gool, "Modeling scenes with local descriptors and latent aspects," in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, pp. 883-890, Oct. 2005.
- [6] D. Gokalp and S. Aksoy, "Scene classification using bag-of-regions representations," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.1-8, 2007.
- [7] S. Lazebnik, C. Schmid and J. Ponce "Beyond Bags Of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2169-2178, 2006.
- [8] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.
- [9] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1794-1801, 2009.
- [10] J. C. v. Gemert, C. J. Veenman, A. W. M. Smeulders, and J. M. Geusebroek, "Visual Word Ambiguity," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 7, pp. 1271-1283, 2010.
- [11] J. Wu and J. M. Rehg, "Place Instance and Category Recognition Using Spatial PACT," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.
- [12] J. Hu and P. Guo, "Multi-Level Kernel Machine for Scene Image Classification," in *Proceedings of 2011 International Conference on Computational Intelligence and Security*, pp. 1169-1173, 2011.
- [13] R. Zabih and J. Woodfill, "Non-Parametric Local Transforms for Computing Visual Correspondence," in *Proceedings of European Conference on Computer Vision (ECCV)*, pp. 151-158, 1994.
- [14] B. Fröba and Andreas Ernst, "Face Detection with The Modified Census Transform," in *Proceedings of the Sixth IEEE international conference on Automatic face and gesture recognition (FGC)*, 2004.
- [15] H. Jin, Q. Liu, H. Lu, and X. Tong, "Face Detection Using Improved LBP Under Bayesian Framework," in *Proceedings of the International Conference on Image and Graphics (ICIG)*, pp. 306-309, 2004.
- [16] L.-J. Li and L. Fei-Fei. "What, Where and Who? Classifying Events By Scene and Object Recognition," in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2007.
- [17] C.-C. Chang and C.-J. Lin, "LIBSVM: A Library for Support Vector Machines," 2001, Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [18] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971-987, 2002
- [19] S. Gao, I. Tsang, L. Chia, and P. Zhao, "Local Features Are Not Lonely – Laplacian Sparse Coding for Image Classification," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3555-3561, 2010.
- [20] O. Boiman, E. Shechtman, and Michal Irani, "In Defense of Nearest-Neighbor Based Image Classification," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1-8, 2010.
- [21] Z. Wang, Y. Hu, and L. Chia, "Image-to-Class Distance Metric Learning for Image Classification," in *Proceedings of European Conference on Computer Vision*, 2010.
- [22] R. Behmo, P. Marcobes, A. Dalalyan, and V. Prinet, "Towards Optimal Naive Bayes Nearest Neighbor", in *Proceedings of European Conference on Computer Vision*, 2010.
- [23] T. Tuytelaars, M. Fritz, K. Saenko, and T. Darrell, "The NBNN Kernel," in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2011
- [24] Y. Jia C. Huang, and T. Darrell, "Beyond Spatial Pyramids: Receptive Field Learning for Pooled Image Features," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.
- [25] A. Bosch, A. Zisserman, and X. Muoz, "Scene Classification Using A Hybrid Generative/Discriminative Approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 4, pp. 712-727, 2008.
- [26] J. J. Kivinen, E. B. Sudderth, and M. I. Jordan, "Learning Multiscale Representaiton of Natural Scenes Using Dirichlet Processes," in *Proceedings of IEEE Conference on Computer Vision (ICCV)*, 2007.
- [27] M. Song and P. Guo, "Scene Recognition via Combing Information of Neighbors," in *Proceedings of International Conference on Computational Intelligence and Security (CIS)*, pp. 345-349, 2012.
- [28] J. Hu, L. Wang, F. Duan, and P. Guo, "Adaptive Multi-level Kernel Machine for Scene Classification," *Mathematical Problems in Engineering*, Volume 2013, Article ID, 324945.
- [29] S. Paris, X. Halkias, H. Glotin, "Sparse Coding for Histograms of Local Binary Patterns Applied for Image Categorization: Toward A Bag-of-Scenes Analysis," in *Proceedings of International Conference on Pattern Recognition (ICPR)*, pp. 1-4, 2012.
- [30] S. Paris, X. Halkias, H. Glotin, "Efficient Bag of Scenes Analysis for Image Categorization," in *Proceedings of International Conference on Pattern Recognition Applications and Methods (ICPRAM)* pp. 10, 2013.
- [31] R. M. N. Sadat, S. W. Teng, G. Lu and S. F. Hasan, "Texture Classification Using Multimodal Invariant Local Binary Pattern," in *Workshop on the Applications of Computer Vision (WACV)*, 2011.
- [32] B. Jun, D. Kim, "Robust Face Detection Using Local Gradient Patterns and Evidence Accumulation," *Pattern Recognition*, vol.45, no.9, pp. 3304-3316, 2012.