# SCENE CLASSIFICATION USING SPATIAL RELATIONSHIP BETWEEN LOCAL POSTERIOR PROBABILITIES

Tetsu Matsukawa[†] and Takio Kurita[‡]

[†]*Department of Computer Science, University of Tsukuba, Tennodai 1-1-1 Tsukuba, Japan*

[‡]*National Institute of Advanced Industrial Science and Technology (AIST), Umezono 1-1-1 Tsukuba, Japan*

Keywords:     Scene classification, Higher order local autocorrelation features, Bag-of-features, Posterior probability image.

Abstract:     This paper presents scene classification methods using spatial relationship between local posterior probabilities of each category. Recently, the authors proposed the probability higher-order local autocorrelations (PHLAC) feature. This method uses autocorrelations of local posterior probabilities to capture spatial distributions of local posterior probabilities of a category. Although PHLAC achieves good recognition accuracies for scene classification, we can improve the performance further by using crosscorrelation between categories. We extend PHLAC features to crosscorrelations of posterior probabilities of other categories. Also, we introduce the subtraction operator for describing another spatial relationship of local posterior probabilities, and present vertical/horizontal mask patterns for the spatial layout of auto/crosscorrelations. Since the combination of category index is large, we compress the proposed features by two-dimensional principal component analysis. We confirmed the effectiveness of the proposed methods using Scene-15 dataset, and our method exhibited competitive performances to recent methods without using spatial grid informations and even using linear classifiers.

## 1   INTRODUCTION

Scene classification technologies have many possible applications such as content-based image retrieval (Vogel and Schiele, 2007) and also can be used as a context for object recognition (Torralba, 2003). Until now, many researchers have tackled this problem.

The most commonly used approach to scene classification is the approach that uses holistic or local statistics of local features. Especially, the spatial pyramid matching that uses bag-of-features on spatial grids showed excellent classification performances in scene classifications(Lazebnik et al., 2006). However, such spatial grided local statistics is weak for spatial misalignment. Another alternative approach is the spatial relationship of local features (Yao et al., 2009). If we use the spatial relationship histogram without spatial griding, this representation may be robust against spatial misalignment. In this paper, we realize the spatial relationship of local regions by using feature extractions on posterior probability images. A posterior probability image is the image in which each pixel shows the confidence of one of the given categories. Although the image feature that focused on posterior probability has not been a standard technique for scene classification so far, we focus on the feature of posterior probability images by two reasons as follows. a): One drawback of the spatial relationship of local features is the problem that possible combination of visual codebook is very large. Therefore, it was required to select distinguish codebook sets (e.g. A priori mining algorithm in (Yao et al., 2009)). b): Synonymy codebook(Zheng et al., 2008) that has similar posterior probabilities of category is existed, that may considered as redundant representations of local features. By using the relationship of local posterior probabilities, such problems can be handled.

Recently, as the features on posterior probability images, the authors proposed the probability higher-order local autocorrelations (PHLAC) feature(Matsukawa and Kurita, 2009). This method is designed by using higher order local autocorrelations (HLAC) features (Otsu and Kurita, 1988) of category posterior probability images. Although PHLAC can capture spatial and semantic informations of images, the autocorrelation was restricted to the posterior probabilities of a single class in the previous study.
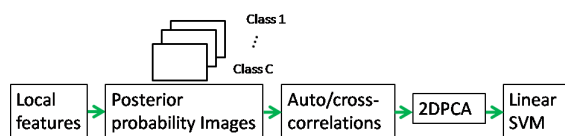
Figure 1: Proposed scene classification approach.

The crosscorrelation between posterior probability images of other categories might capture richer informations for scene classification. In this paper, we extend PHLAC features to categorical auto/crosscorrelations. Also, we introduce subtraction operator for describing another spatial relationship of local posterior probabilities. Since the combination of category index is large, we compress the proposed features by two-dimensional principal component analysis(2DPCA) (Yang et al., 2004). Finally, the proposed scene classification method becomes as shown in Figure 1. The proposed method is effective for scene classification problems even using linear classifiers.

## 2 RELATED WORK

There are some approaches for scene classification; classification with features on spatial domain (Renninger and Malik, 2004; Battiato et al., 2009; Szummer and Picard, 1998; Gorkani and Picard, 1994), and features on frequency domain(Oliva and Torralba, 2001; Torralba and Oliva, 2003; Farinella et al., 2008; Ladret and Gue'rin-Dugue', 2001). In scene classification in spatial domain, a scene is represented as different forms such as histogram of visual words(Lazebnik et al., 2006), textons(Battiato et al., 2009), semantic concepts(Vogel and Schiele, 2007), scene prototypes(Quattoni and Torralba, 2009) and so on. Our motivation is to improve classification accuracies of scene classification in spatial domain. To this end, we focus on features on posterior probability images (semantic information).

Rasiwasia et al. proposed a semantic feature representation by using the bag-of-features method based on the Gaussian mixture model (Rasiwasia and Vasconcelos, 2008). In their study, each feature vector indicated the probability of each class label, and they refer to this type of scene labeling as casual annotation. Using this feature, they could achieve high classification accuracies with low feature dimensions. Methods that provide posterior probabilities to a codebook have also been proposed (Shotton et al., 2008). However, these methods do not employ the spatial relationship of local regions. Recently, pixel(region) wise scene annotation methods by predetermined concepts

(e.g. sky, road, tree) are researched by many authors (Bosch et al., 2007). By combining their methods to our methods, total scene categories (e.g. suburb, coast, mountain) can be recognized. Vogel et al. classified local regions to semantic concepts by classifiers and classified total scene categories using histogram representations of local concepts (Vogel and Schiele, 2007). A similar method to PHLAC that uses local autocorrelations of similarity of category subspaces constructed by Kernel PCA was recently proposed(Hotta, 2009). However, this method also doesn't use crosscorrelation between different category subspaces.

## 3 PHLAC FEATURES

### 3.1 Posterior Probability Image

Let I be an image region, and $r = (x,y)^t$ be a position vector in I. The image patches whose center is $r_k$ are quantized to M codebooks $\{V_1,...,V_M\}$ by local feature extractions and the vector quantization algorithm $\text{VQ}(r_k) \in \{1,...,M\}$. These steps are the same as that of the standard bag-of-features method (Lazebnik et al., 2006). Posterior probability $P(c|V_m)$ of the category $c \in \{1,...,C\}$ is calculated to each codebook $V_m$ using image patches on the training images. Several forms of estimating the posterior probability can be used. In this study, the posterior probability is estimated by using Bayes' theorem as follows.

$$P(c|V_m) = \frac{P(V_m|c)P(c)}{P(V_m)} = \frac{P(V_m|c)P(c)}{\sum_{c=1}^{C} P(V_m|c)P(c)}, \quad (1)$$

where $P(c) = 1/C$, $P(V_m) =$ (# of $V_m$)/(# of all patches), $P(V_m|c) =$ (# of class c $\wedge$ $V_m$)/(# of class c patches). Here, $P(c)$ is a common constant and here is set to 1.

In this study, the grid sampling of local features (Lazebnik et al., 2006) is carried out at pixel interval of $p$ for simplicity. We denote the set of sample points as $I_p$ and the map of posterior probability of the codebook of each local region as a posterior probability image. Examples of posterior probability images are shown in Figure 2. White color represents the high probability. Note that the forest like regions appear in the vegetations of each image, and the flat regions such as sky have high probability of highway. In these way, each semantic region can be roughly characterized through class label.
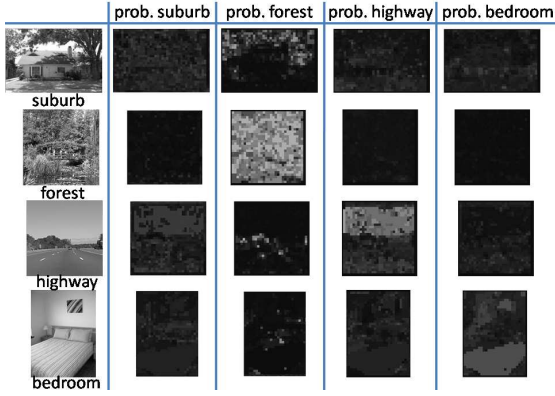
Figure 2: Posterior probability images.

Figure 3: Mask pattern for PHLAC. No.1 is 0th order, No.2-6 are 1st order, and No.7-35 are 2nd order. The numbers $\{1,2,3\}$ of the mask patterns show the frequency at which their pixel value is used for obtaining the product expressed in Eq. (2).

## 3.2 PHLAC

Autocorrelation is defined as the product of signal values from different points and represents the strong co-occurrence of these points. HLAC (Otsu and Kurita, 1988) has proposed for extracting spatial autocorrelations of intensity values. To capture the spatial autocorrelations of posterior probabilities, the feature called PHLAC is designed by HLAC features of posterior probability images. Namely, the Nth order PHLAC is defined as follows.

$$R(c,a_1,...,a_N) = \int_{I_p} P(c|V_{VQ(r)})P(c|V_{VQ(r+a_1)}) \cdots P(c|V_{VQ(r+a_N)})dr. \quad (2)$$

In practice, many forms of Eq. (2) can be obtained by varying the parameters N and $a_n$. These parameters are restricted to the following subset: $N \in \{0,1,2\}$ and $a_{nx}, a_{ny} \in \{\pm \Delta r \times p, 0\}$. In this case, HLAC feature can be calculated by sliding predetermined mask patterns. By eliminating duplicates that arise from shifts of center positions, the mask patterns of PHLAC can be represented as shown in Figure 3.

**Larger Mask Patterns.** Larger mask patterns are obtained by varying the spatial interval $\Delta r$. By calculating the autocorrelations in local regions, PHLAC becomes robust against small spatial difference and noise. Namely, $P(c|V_{VQ(r+a_n)})$ in E.q.(2) is replaced to the local averaged posterior probability $L_a(P(c|V_{VQ(r+a_n)}))$. Local averaging is calculated in the rectangle region in which $(x-\frac{a_{nx}}{2}, y-\frac{a_{ny}}{2})$ is the upper left corner, and $(x+\frac{a_{nx}}{2}, y+\frac{a_{ny}}{2})$ is the lower left corner. The example of a larger mask pattern is shown in Figure 4.

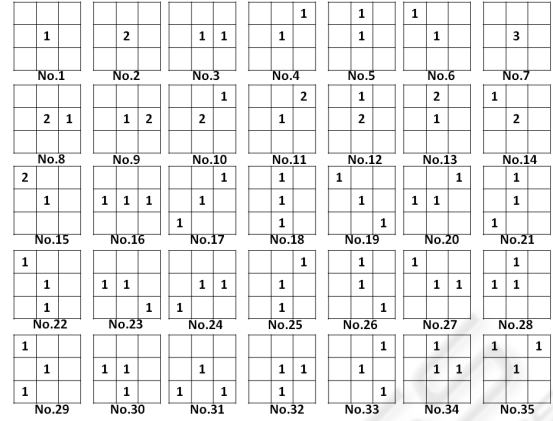In previous paper, PHLAC feature was calculated by using a single spatial interval. Recently, HLAC feature was extended to 8th order and the multi-scale spatial interval(Toyoda and Hasegawa, 2007). It was not obvious that how many order of autocorrelations and the multi-scale spatial interval are effective for PHLAC. We confirmed that the multi-scale spatial interval is effective for scene classification, and the order of autocorrelation is sufficient upto 2nd order.

**Effectiveness of PHLAC.** PHLAC is an extension of the bag-of-features approach. It was shown that 0th order PHLAC has the almost same property of bag-of-features, when linear classifiers are used. Additionally, higher order features of PHLAC have spatial distribution informations of each posterior probability image. Therefore, PHLAC is possible to achieve higher classification performances than bag-of-features.

Furthermore, the desirable property of PHLAC is its shift invariance. PHLAC uses spatial informations by the meaning of relative positions of local regions that are robust to spatial misalignment in shift than histogram in spatial grids. Thus, the integration by spatial autocorrelations could be better approach to integrate the local posterior probabilities.

The autocorrelation of PHLAC is calculated on each posterior probability image of category. Thus, the dimension of PHLAC is (# of mask patterns $\times$ C) $\times$ # of spatial intervals. However, the crosscorrelation between other categories could contains more richer informations for classification.
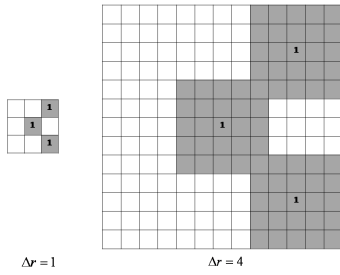
Figure 4: Larger mask pattern. Autocorrelation of larger regions is realized by varying the parameter of spatial interval $\Delta r$. The local averaged areas are indicated in gray area.

# 4 SPATIAL RELATIONSHIP BETWEEN LOCAL POSTERIOR PROBABILITIES

## 4.1 CIPHLAC

To capture more richer information between posterior probability images, such as a road like region is left to a forest like region, we extend the autocorrelation of PHLAC to auto/crosscorrelations between posterior probabilities of different categories. The difference between the autocorrelation and crosscorrelation is shown in Figure 5. We call this method as category index probability higher order autocorrelations (CIPHLAC). The N-th order CIPHLAC is defined as follows.

$$R(c_0,..,c_N a_1,..,a_N) =$$
$$\int_{I_p} P(c_0|V_{VQ(r)}) P(c_1|V_{VQ(r+a_1)})$$
$$\cdots P(c_N|V_{VQ(r+a_N)}) dr. \quad (3)$$

Similar with the case of PHLAC, the parameters N and $a_n$ are restricted to the following subset: $N \in \{0,1,2\}$ and $a_{nx}, a_{ny} \in \{\pm\Delta r \times p, 0\}$. Also, the auto/crosscorrelations are calculated by regions and the multi-scale spatial interval is used. To avoid the increase of dimensions, we restrict the combination of category index for 2nd order as follows: $(c_0, c_1, c_2) = (c_0, c_1, c_1)$. The calculation of CIPHLAC is realized by extending PHLAC mask patterns(Figure 6). Note that this mask pattern contains both autocorrelations and crosscorrelations. The dimension of CIPHLAC becomes ( C + # of 1st and 2nd order mask patterns $\times$ C$^2$ )$\times$ # of spatial intervals.
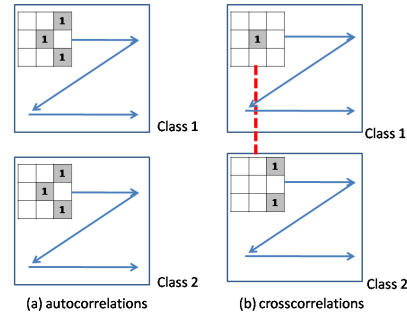


Figure 5: Comparison of autocorrelations and crosscorrelations. (a) autocorrelation calculates correlation among posterior probabilities of each category. (b) crosscorrelaiton calculates correlation among posterior probabilities of different categories.

## 4.2 CIPLAS

CIPHLAC uses only the products between local posterior probabilities of different categories. Another relationship operation may capture another feature between posterior probability images. Here, we propose a subtraction operation of posterior probabilities. In this paper, only 1st order subtraction is considered. We define this method as category index probability local autosubtraction (CIPLAS). The method construct the sum of subtraction values by distinguishing the positive and negative values. Namely, the definition of CIPLAS is given as follows.

$$R^+(c_0,c_1 a_1) =$$
$$\int_{I_p} \Psi^+ \left( P(c_0|V_{VQ(r)}) - P(c_1|V_{VQ(r+a_1)}) \right) dr,$$
$$\Psi^+(x) = \begin{cases} 0 \, (if \quad x \leq 0), \\ x \, (if \quad x > 0). \end{cases}$$
$$R^-(c_0,c_1 a_1) =$$
$$\int_{I_p} \Psi^- \left( P(c_0|V_{VQ(r)}) - P(c_1|V_{VQ(r+a_1)}) \right) dr,$$
$$\Psi^-(x) = \begin{cases} |x| \, (if \quad x \leq 0), \\ 0 \, (if \quad x > 0). \end{cases}$$
$$(4)$$

Similar with the case of PHLAC, the parameters N and $a_n$ are restricted to the following subset: $N \in \{0,1,2\}$ and $a_{nx}, a_{ny} \in \{\pm\Delta r \times p, 0\}$. Also, the auto/crosssubtractions are calculated by regions, and the multi-scale spatial interval is used. In this paper, we use the 1st order mask patterns (No.2-6) in PHLAC mask patterns. Thus, the dimension of CIPLAS is ($2\times$ number of 1st order mask patterns $\times$ C$^2$)$\times$# of spatial intervals. Because auto/crosscorrelation values becomes high if
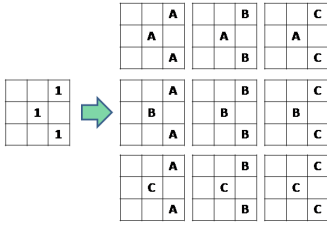
Figure 6: Mask patterns for CIPHLAC that restricted center and the other category index. (A, B, C) shows a category index.

the two reference posterior probabilities are high, and auto/crosssubtraction values becomes high if these are different, CIPLAS may be used with CIPHLAC complementary.

## 4.3 Vertical/Horizontal Mask Patterns

The original mask patterns for PHLAC calculate autocorrelation of square block regions. In scene classification, it is also effective the relationship of vertical or horizontal regions by neglecting y or x axis information of the images. In this paper, we consider auto/crosscorrelations of vertical and horizontal regions as shown in Figure 7. Each posterior probability is averaged as in the upper of Figure 7 and auto/crosscorrelations are calculated each local averaged regions by sliding mask pattern described in the bottom of Figure 7. These mask patterns can be used for PHLAC, CIPHLAC, and CIPLAS. In the case of CIPLAS, No.1, 2, 3, 7, 8, 9 of the mask patterns are used, since CIPLAS is 1st order. Thus, the number of the additional mask patterns is only 6. However, these mask patterns were significantly effective for CIPLAS.
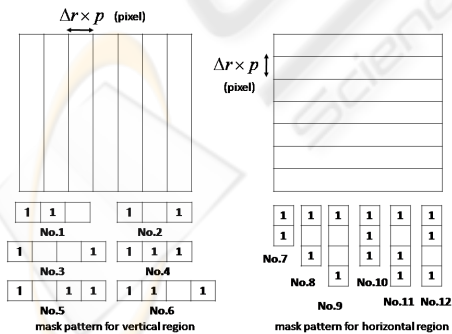


Figure 7: Vertical and horizontal mask patterns.

## 4.4 Feature Compression using 2DPCA

Because the dimensions of 2nd order CIPHLAC and CIPLAS are large, we compress the features vectors. The most standard technique for dimension

compression is principal component analysis(PCA). However, in the high dimensional and small sample data, the accurate covariance matrix for PCA can not be calculated. For two-dimensional data compression, two-dimensional principal component analysis(2DPCA)(Yang et al., 2004) has been proposed to slove this problem. Because the CIPHLAC and CIPLAS is viewed as two-dimensional data (combination of category index $\times$ mask patterns), so we compress the feature vector by 2DPCA.

Let $\mathbf{A}$ denotes an m $\times$ n feature vector, and $\mathbf{X}$ denotes an n-dimensional unitary column vector. Here, m is # of combinations of category index, and n is # of mask patterns $\times$ # of spatial intervals. In 2DPCA, the covariance matrix of PCA is replaced to the following image covariance matrix $G_t$,

$$G_t = \frac{1}{M} \sum_{j=1}^{J} (\mathbf{A_j} - \overline{\mathbf{A}})^t (\mathbf{A_j} - \overline{\mathbf{A}}), \qquad (5)$$

where $\overline{\mathbf{A}}$ is the average of all training samples $\mathbf{A_j}$, (j=1,2,...J). The optimal projection vectors, $\mathbf{X_1},...,\mathbf{X_d}$ are obtained as the orthonormal eigenvectors of $\mathbf{G_t}$ corresponding to the first d largest eigenvalues. The projected feature vectors, $\mathbf{Y_1},...,\mathbf{Y_d}$, are obtained by using the optimal projection vectors, $\mathbf{X_1},...,\mathbf{X_d}$ as following equations.

$$\mathbf{Y_k} = \mathbf{A X_k}, k = 1,2,...,d. \qquad (6)$$

The obtained principal component vectors are used to form an m $\times$ d matrix $\mathbf{B} = [\mathbf{Y_1}, \cdots, \mathbf{Y_d}]$. We apply second 2DPCA to $\mathbf{B^t}$ for reduction both redundancy of combination of category index and mask patterns. Both 1st and 2nd order vectors are compressed by 2DPCA simultaneously and the compressed vectors are concatenated to 0th order vectors.

## 5 EXPERIMENT

We performed experiments on Scene-15 dataset (Lazebnik et al., 2006). The Scene-15 dataset consists of 4485 images spread over 15 categories. The fifteen categories contain 200 to 400 images each and range from natural scene like mountains and forest to man-made environments like kitchens and office. We selected 100 random images per categories as a training set and the remaining images as the test set. Some examples of dataset images are shown in Figure 8.

To obtain reliable results, we repeated the experiments 10 times. Ten random subsets were selected from the data to create 10 pairs of training and test data. For each of these pairs a codebook was created by using k-means clustering on training
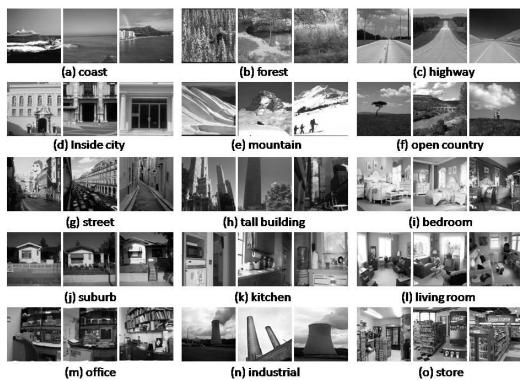
Figure 8: Example of Scene-15 dataset. Eight of these (a-h) were originally collected in (Oliva and Torralba, 2001), five (i-m) in (FeiFei and Perona, 2005), and two (n-o) in (Lazebnik et al., 2006).



Figure 9: Recognition rates of PHLAC per number of auto-correlation order.



Figure 10: Recognition rates of PHLAC per number of spatial interval.

set. For classification, a linear SVM was used by one-against-all. As implementation of SVM, we used LIBLINEAR(Fan et al., 2008). Five-fold cross-validation on the training set was carried out to tune parameters of SVM. The classification rate we report is the average of the per-class recognition rates which in turn are averaged over the 10 random test sets. As local features, we used a gradient local auto-correlation(CLAC) descriptor (Kobayashi and Otsu, 2008) sampled on a regular grid. Because GLAC can extract richer information than SIFT descriptor and produce better performance. GLAC descriptor used in this paper is 256-dimensional co-occurrence histogram of gradient direction that contains 4 types of local autocorrelation patterns. We calculated the feature values from a $16 \times 16$ pixel patch sampled every 8 pixels (i.e. $p = 8$), and histogram of each autocorrelation pattern is L2-Hys normalized. In the codebook creation process, all features sampled every 24 pixel on all training images were used for k-means clustering. The codebook size k is set to 400. As normalization method, we used L2-norm normalization per autocorrelation order for PHLAC, CIPHLAC, and CIPLAS. When using dimensional compression, 2DPCA is applied after this normalization and the compressed feature vector were L2-norm normalized using all compressed dimension.

**Autocorrelation-order.** First, we compare the auto-correlation order of PHLAC. The autocorrelation order is changed from 0th to 8th order by using mask patterns described in (Toyoda and Hasegawa, 2007). The numbers of mask patterns for 3rd - 8th order PHLAC are $\{153, 215, 269, 297, 305, 306\}$ respectively. In this comparison, single spatial interval $\Delta r \in \{4, 8\}$ is used and feature vectors are not compressed. The result is shown in Figure 9. The recognition rates
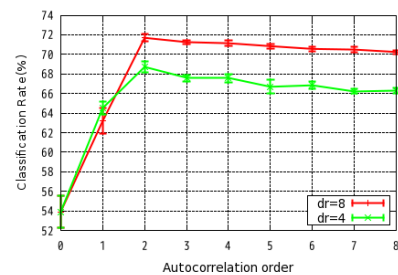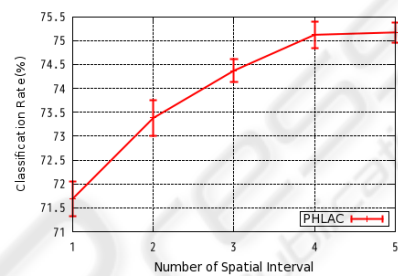
becomes better as to increase the autocorrelation order upto 2nd. Therefore, the practical restriction of autocorrelation order $N \in \{0, 1, 2\}$ was reasonable.

**Number of Spatial Interval.** Next, the number of spatial interval is changed in 2nd order PHLAC. The spatial interval is selected in the best combination of $\Delta r \in \{1, 2, 4, 8, 12\}$ by the first spilt classification rates. In this comparison, feature vectors are not compressed. The recognition rates per number of spatial interval are shown in Figure 10. It can be confirmed that the recognition rates becomes increase as to increase the number of spatial interval.

**Effectiveness of Relationship Between Posterior Probability Images.** The comparison between PHLAC, CIPHLAC and CIPLAS is shown in Table 1. In this comparison, spatial interval $\Delta r = 8$ is used and feature vectors are not compressed. It is shown that crosscorrelations between posterior probability images improved classification performance of PHLAC. For 1st order, the CIPHLAC had 8.89 % better classification performance than PHLAC and 1.19 % for 2nd order. The recognition rates of CIPLAS were also better than PHLAC for 1st order. The combination of CIPHLAC and CIPLAS slightly improved the recognition rates of CIPHLAC.

**Effectiveness of Vertical/horizontal Mask Patterns.** The effectiveness of vertical/horizontal mask pat-

Table 1: Recognition rates of different relevant operations (single spatial interval).

| features | 1st order | 2nd order |
|---|---|---|
| PHLAC | 63.19(± 1.29) | 71.69(± 0.37) |
| CIPHLAC | 72.08(± 0.51) | 73.67(± 0.24) |
| CIPLAS | 70.83(± 0.63) | - |
| CIPHLAC + CIPLAS | **72.45(± 0.60)** | **74.26(± 0.24)** |

Table 2: Recognition rates of vertical/horizontal mask patterns. (a) standard mask patterns. (b) vertical/horizontal mask patterns. (c) combination of (a) and (b).

| | PHLAC | CIPHLAC | CIPLAS |
|---|---|---|---|
| (a) | 71.69(± 0.37) | 73.67(± 0.24) | 70.83(± 0.63) |
| (b) | 71.77(± 0.30) | 74.57(± 0.29) | **76.46(± 0.23)** |
| (c) | 74.23(± 0.20) | 75.61(± 0.21) | **76.83(± 0.32)** |

terns are confirmed. In this comparison, spatial interval $\Delta r = 8$ is used and feature vectors are not compressed. The experimental results are shown in Table 2. By using only vertical/horizontal mask patterns, the recognition rates of all methods outperformed the original mask patterns. By using both vertical/horizontal mask patterns and original mask patterns, the recognition rates were improved. The CIPLAS with vertical/horizontal mask patterns exhibited the best performance.

**Comparison with State of the Art.** We used five scales of spatial interval $\Delta r = (1,2,4,8,12)$ for both original and vertical/horizontal mask patterns. We compressed feature vectors by 2DPCA. We applied 2DPCA to the standard mask patterns and vertical/horizontal(VH) mask patterns separability, and concatenated these compressed vectors. The rank of 2DPCA is experimentally set using first spilt data. The ranks (rank of category index combination, rank of mask pattern) used in this experiment for each method are: (50, 80) for CIPHLAC, (50, 40) for CIPHLAC(VH mask patterns), (60, 40) for CIPLAS, (80, 20) for CIPLAS(VH mask patterns). The classification results of our proposed methods are shown in upper of Table 4. CIPLAS with vertical/horizontal mask patterns exhibited the best performance in proposed methods. The confusion matrix of CIPLAS with vertical/horizontal mask patterns is shown in Figure 11. For the baseline of our proposed methods, we evaluated the two spatial gridded features by using the same feature and codebook to ours. These are:
(a) **SPM(linear):** the histogram of visual codebook was created in the spatial pyramid grids as in the spatial pyramid matching (Lazebnik et al., 2006).
(b) **SP-PHLAC(0th):** the 0th order PHLAC (i.e. sum of posterior probabilities) was created in the spatial pyramid grids as in the spatial pyramid matching.
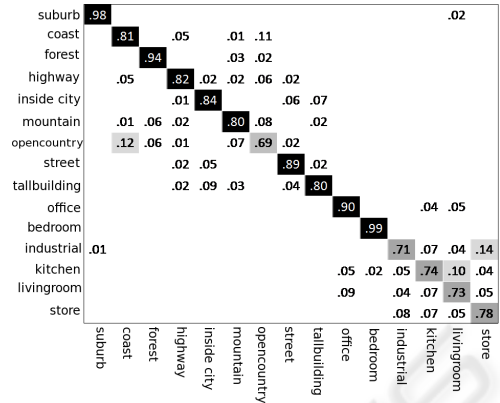


Figure 11: Confusion matrix of CIPLAS with VH.

Table 3: Average required time per image (msec).

| BOF | PHLAC | CIPHLAC | CIPLAS |
|---|---|---|---|
| 241.427 | 325.613 | 861.7 | 408.373 |

These method was classified by linear SVM to compare the goodness of feature representations. The results of baseline methods are shown in the middle of Table 4. Our methods significantly outperformed these baseline methods. The comparison with the state-of-the-art methods in Scene-15 dataset is shown in the bottom of Table 4. Our method has the better classification performance than spatial pyramid matching (Lazebnik et al., 2006) without using spatial grid informations and kernel methods. Furthermore, our method exhibited the best performance in the method that uses linear classifiers.

**Computational Cost.** Table 3 shows the average computation time of Scene-15 dataset in the feature calculation process. One core of the quad core CPU (Xeon 2.66 GHz), and C++ implementation were used. In table 3, PHLAC, CIPHLAC, and CIPLAS were calculated in 5 spatial intervals, and BOF is the standard bag-of-features without spatial griding. Since PHLAC is an extension of bag-of-features, the method requires more computational costs than bag-of-features. However, the method enables us to achieve high accuracies with linear classifiers that can calculate much faster than kernel methods. Although CIPHLAC requires larger computational times than PHLAC because the number of mask patterns is large, CIPHLAC produces better accuracies than PHLAC.

# 6 CONCLUSIONS

We have proposed scene classification methods using spatial relationship between local posterior probabilities. The autocorrelations of PHLAC feature are

Table 4: Comparison with previous methods.

| algorithm | spatial | classifier | recognition rate |
|---|---|---|---|
| CIPHLAC(without VH) | relative | linear | 80.37(± 0.37) |
| CIPLAS(without VH) | relative | linear | 77.46(± 0.33) |
| CIPHLAC(with VH) | relative | linear | 80.65(± 0.57) |
| CIPLAS(with VH) | relative | linear | **82.63(± 0.25)** |
| ALL | relative | linear | 81.64(± 0.53) |
| SPM(linear)(Baseline) | grid | linear | 72.60(± 0.27) |
| SP-PHLAC(0th)(Basel.) | grid | linear | 68.29(± 0.22) |
| (Bosch et al., 2008) | grid | kernel | **83.7** |
| (Wu and Rehg, 2008) | grid | kernel | **83.3(± 0.5)** |
| (Lazebnik et al., 2006) | grid | kernel | 81.4(± 0.5) |
| (Yang et al., 2009) | grid | linear | 80.28(± 0.93) |
| (Zheng et al., 2009) | grid | kernel | 74.0 |

extended to auto/crosscorrelations between posterior probability images. The autosubtraction operator for describing another spatial relationship between posterior probability images, and vertical/horizontal mask patterns for spatial layout of auto/crosscorrelations are also proposed. Since the combination of category index is large, the features are compressed by 2DPCA. Experiments using Scene-15 dataset have demonstrated that the crosscorrelations between posterior probabilities improves classification performances of PHLAC, and auto/crosssubtraction with vertical/horizontal mask patterns indicated the best performance in our methods. The classification performances of our methods in Scene-15 dataset were competitive to the recent proposed methods without using spatial grid informations and even using linear classifiers.

# REFERENCES

Battiato, S., Farinella, G., Gallo, G., and Ravi, D. (2009). Spatial hierarchy of textons distributions for scene classification. In *MMM*, pages 333–343.

Bosch, A., Munoz, X., and Freixenet, J. (2007). Segmentation and description of natural outdoor scenes. *Image and Vision Computing*, 25:727–740.

Bosch, A., Zisserman, A., and Munoz, X. (2008). Scene classification using a hybrid generative/discriminative approach. *IEEE Trans. on PAMI*, 30(4):712–727.

Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., and Lin, C.-J. (2008). Liblinear: A library for large linear classification. *JMLR*, 9:1871–1874.

Farinella, G., Battiato, S., Gallo, G., and Cipolla, R. (2008). Natural versus artificial scene classification by ordering discrete fourier power spectra. In *SSPR&SPR*.

FeiFei, L. and Perona, P. (2005). A bayesian hierarchical model for learning natural scene categories. In *CVPR*.

Gorkani, M. and Picard, R. (1994). Texture orientation for sorting phots at a glance. In *ICPR*.

Hotta, K. (2009). Scene classification based on local auto-correlation of similarities with subspaces. In *ICIP*.

Kobayashi, T. and Otsu, N. (2008). Image feature extraction using gradient local auto-correlations. In *ECCV*.

Ladret, P. and Gue'rin-Dugue', A. (2001). Categorisation and retrieval of scene photographs from a jpeg compressed database. *Pattern Analysis & Applications*, 4:185 – 199.

Lazebnik, S., Schmid, C., and Ponece, J. (2006). Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *CVPR*.

Matsukawa, T. and Kurita, T. (2009). Image classification using probability higher-order local auto-correlations. In *ACCV*.

Oliva, A. and Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *IJCV*, 42(3):145–175.

Otsu, N. and Kurita, T. (1988). A new scheme for practical flexible and intelligent vision systems. In *IAPR Workshop on Computer Vision*, pages 431–435.

Quattoni, A. and Torralba, A. (2009). Recognizing indoor scenes. In *CVPR*, pages 413–420.

Rasiwasia, N. and Vasconcelos, N. (2008). Scene classification with low-dimensional semantic spaces and weak supervision. In *CVPR*, pages 1 – 8.

Renninger, L. and Malik, J. (2004). When is scene identification just texture recognition? *Vision Research*, 44(19):2301–2311.

Shotton, J., Johnson, M., and Cipolla, R. (2008). Semantic texton forests for image categorization and segmentation. In *CVPR*, pages 1 – 8.

Szummer, M. and Picard, R. (1998). Indoor-outdoor image classification. In *IEEE Intl. Workshop on Content-Based Access of Image and Video Databases*.

Torralba, A. (2003). Contextual priming for object detection. *IJCV*, 53:169 – 191.

Torralba, A. and Oliva, A. (2003). Statistics of natural image categories. *Network: Computing in Nueral Systems*, 14:391 – 412.

Toyoda, T. and Hasegawa, O. (2007). Extension of higher order local autocorrelation features. *Pattern Recognition*, 40:1466–1477.

Vogel, J. and Schiele, B. (2007). Semantic modeling of natural scenes for content-based image retrieval. *IJCV*, 72(2):133–157.

Wu, J. and Rehg, J. (2008). Where am i: Place instance and category recognition using spatial pact. In *CVPR*.

Yang, J., Yu, K., Gong, Y., and Huang, T. (2009). Linear spatial pyramid matching using sparse coding for image classification. In *CVPR*.

Yang, J., Zhang, D., Frangi, A., and Yang, J. (2004). Two-dimensional pca: a new approach to appearance-based face representation and recognition. *IEEE Trans. on PAMI*, 26(1):131–137.

Yao, B., Niebles, J., and Fei-Fei, L. (2009). Mining discriminative adjectives and prepositions for natural scene recognition. In *The joint VCL-ViSU 2009 workshop*.

Zheng, Y., Lu, H., Jin, C., and Xue, X. (2009). Incorporating spatial correlogram into bag-of-features model for scene categorization. In *ACCV*.

Zheng, Y.-T., Zhao, M., Neo, S.-Y., Chua, T.-S., and Tian, Q. (2008). Visual synset: towards a higher-level visual representation. In *CVPR*, pages 1 – 8.