

# Comparison of Combining Methods of Correlation Kernels in kPCA and kCCA for Texture Classification with Kansei Information

Yo Horikawa<sup>1</sup> and Yujiro Ohnishi<sup>2</sup>

<sup>1</sup> Faculty of Engineering, Kagawa University, Takamatsu, 761-0396 Japan

<sup>2</sup> Ryobi Systems Corporation, Okayama, 700-8504 Japan

horikawa@eng.kagawa-u.ac.jp

<http://www.eng.kagawa-u.ac.jp/~horikawa/>

**Abstract.** The authors consider combining correlations of different orders in kernel principal component analysis (kPCA) and kernel canonical correlation analysis (kCCA) with the correlation kernels. We apply combining methods, e.g., the sums of the correlation kernels, Cartesian spaces of the principal components or the canonical variates and the voting of kPCAs and kCCAs output and compare their performance in the classification of texture images. Further, we apply Kansei information on the images obtained through questionnaires to the public to kCCA and evaluate its effectiveness.

**Keywords:** kernel method, principal component analysis, canonical correlation analysis, correlation kernel, combining classifiers, Kansei information, texture classification.

## 1 Introduction

Kernel principal component analysis (kPCA) [1] and kernel canonical correlation analysis (kCCA) [2], [3], [4] are kernelized versions of PCA and CCA in multivariate statistical analysis. In PCA, the linear projections which allow to reconstruct original feature vectors are obtained with minimal quadratic errors. It is used to reduce the dimensionality of the original data retaining most existing structure in the data. In CCA, linear transformations that yield maximum correlation between two kinds of features vectors of objects, e.g., their images and sounds, are obtained. It is also applied to dimensionality reduction or feature extraction. In the kernel methods, the inner products of the feature vectors are replaced to nonlinear kernel functions [5], [6]. Nonlinear mappings of the feature vectors to high-dimensional spaces are then performed in implicit manners. Then nonlinear characteristics of the original data can be extracted with them.

Correlation kernels were recently proposed as kernel functions [7], [8]. They are inner products of the autocorrelation functions of the feature vectors. The idea was shown about forty years ago and their characteristics are that higher-order correlation kernels are effectively calculated [9]. They are suitable to image data, which have strong spatial correlations and support vector machines (SVM), kPCA and kCCA with the correlation kernels were applied to invariant texture classification [10], [11], [12].

In this study we consider combining correlation kernels of different orders in kPCA and kCCA in the classification of texture images. Combining classifiers have been of wide interest in pattern recognition [13], [14] and they can show higher classification performance than any single classifiers. Some combining methods: use of the sums of the correlation kernels as kernel functions, use of Cartesian spaces of the principal components or the canonical variates as feature vectors and the voting of output of the classifiers with kPCA or kCCA are employed and compared in their classification performance. Further, we try to use Kansei information, the perceptual and cognitive ability to feel objects, e.g., impressions to images. Kansei information was applied to image retrieval systems with CCA [15] and it can be adopted in kCCA as the second feature vectors.

In Sect. 2, related theoretical background: kPCA, kCCA, correlation kernels, combining methods and Kansei information are mentioned. In Sect. 3, Method and results of texture classification experiment with kPCA and kCCA are shown. Discussions are given in Sect. 4.

## 2 Theoretical Background

### 2.1 Kernel Principal Component Analysis (kPCA)

The feature vectors  $\mathbf{x}_i$  ( $1 \leq i \leq n$ ) of sample objects are transformed to  $\boldsymbol{\varphi}(\mathbf{x}_i)$  in another feature spaces with an implicit nonlinear mapping  $\mathbf{h}$ . We assume that the mean of the transformed features are zero, i.e.,  $\sum_{i=1}^n \mathbf{h}(\mathbf{x}_i) = \mathbf{0}$ , for simplicity. The mean centering can be done in the calculation of kernel functions [1]. The kernel version of PCA is PCA for  $\mathbf{h}(\mathbf{x}_i)$  and the principal components are obtained through the eigenproblem

$$\boldsymbol{\Phi} \mathbf{v} = \lambda \mathbf{v} \quad (1)$$

where  $\boldsymbol{\Phi}$  is a kernel matrix and its elements are calculated with a kernel function  $\Phi_{ij} = \mathbf{h}(\mathbf{x}_i) \cdot \mathbf{h}(\mathbf{x}_j) = \varphi(\mathbf{x}_i, \mathbf{x}_j)$ . Let  $\mathbf{v}_r = (v_{r1}, \dots, v_{rn})^T$  ( $1 \leq r \leq R$  ( $\leq n$ )) be the eigenvectors in the non-increasing order of the corresponding non-zero eigenvalues  $\lambda_r$ , which are normalized as  $\lambda_r \mathbf{v}_r^T \mathbf{v}_r = 1$ . The  $r$ th principal component  $u_r$  for a new data  $\mathbf{x}$  is then obtained by

$$u_r = \sum_{i=1}^n v_{ri} \mathbf{h}(\mathbf{x}_i) \cdot \mathbf{h}(\mathbf{x}) = \sum_{i=1}^n v_{ri} \varphi(\mathbf{x}_i, \mathbf{x}) \quad (2)$$

Classification methods, e.g., the nearest-neighbor method, the discriminant analysis and SVMs can be applied in the principal component space  $(u_1, \dots, u_R)$ .

### 2.2 Kernel Canonical Correlation Analysis (kCCA)

The kernel version of CCA is as follows [2], [3], [4]. Let  $(\mathbf{x}_i, \mathbf{y}_i)$ , ( $1 \leq i \leq n$ ) be pairs of the feature vectors of  $n$  sample objects, which describe different aspects of the objects, e.g., sounds and images. Define kernel matrices  $\boldsymbol{\Phi}$  and  $\boldsymbol{\Theta}$  by  $\Phi_{ij} = \varphi(\mathbf{x}_i, \mathbf{x}_j)$  and  $\Theta_{ij} = \theta(\mathbf{y}_i, \mathbf{y}_j)$ , ( $1 \leq i, j \leq n$ ), which correspond to the inner products of implicit functions of  $\mathbf{x}$  and  $\mathbf{y}$ , respectively. Then we obtain the eigenvectors  $(\mathbf{f}^T, \mathbf{g}^T)^T$  of the generalized eigenproblem:

$$\begin{pmatrix} \mathbf{0} & \Phi\Theta \\ \Theta\Phi & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{f} \\ \mathbf{g} \end{pmatrix} = \lambda \begin{pmatrix} \Phi^2 + \gamma_x \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \Theta^2 + \gamma_y \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{f} \\ \mathbf{g} \end{pmatrix} \quad (3)$$

Small multiples of the identity matrix  $\gamma_x \mathbf{I}$  and  $\gamma_y \mathbf{I}$  are added for the regularization. The canonical variates  $u$  and  $v$  of  $(\mathbf{x}, \mathbf{y})$  of a object are linear projections of the implicit functions of  $\mathbf{x}$  and  $\mathbf{y}$  which maximize correlation between them. They are obtained with the eigen vectors  $\mathbf{f}$  and  $\mathbf{g}$  by

$$\begin{aligned} u &= \sum_{i=1}^n f_i \phi(\mathbf{x}_i, \mathbf{x}) \\ v &= \sum_{i=1}^n g_i \theta(\mathbf{y}_i, \mathbf{y}) \end{aligned} \quad (4)$$

An indicator vector is used as the second feature vector  $\mathbf{y}$  for classification problems [2]. When an object  $\mathbf{x}$  is categorized into one of  $C$  classes, the indicator vector corresponding to  $\mathbf{x}$  is defined by

$$\mathbf{y} = (y_1, \dots, y_C)^T \quad (5)$$

$$\begin{aligned} y_c &= 1 \quad \text{if } \mathbf{x} \text{ belongs to class } c \\ y_c &= 0 \quad \text{otherwise} \quad (1 \leq c \leq C) \end{aligned}$$

Then the linear inner product  $\mathbf{y}_i^T \mathbf{y}_j$  is used as the kernel function  $\theta(\mathbf{y}_i, \mathbf{y}_j)$ . The canonical variates  $u_i$  ( $1 \leq i \leq C-1$ ) are obtained corresponding to non-zero eigenvalues of Eq. (3). This is equivalent to Fisher's discriminant analysis in two class problems. Standard classification methods are also applied in the canonical variate space.

### 2.3 Correlation Kernels and Their Modification

The autocorrelation of the original feature vector  $\mathbf{x}$  is used in the correlation kernels [7], [8]. In the following, we consider 2-dimensional image data  $x(l, m)$ , ( $1 \leq l \leq L$ ,  $1 \leq m \leq M$ ) as the feature vector  $\mathbf{x}$ . The  $k$ th-order autocorrelation  $r_x(l_1, l_2, \dots, l_{k-1}, m_1, m_2, \dots, m_{k-1})$  of  $x(l, m)$  is defined by

$$\begin{aligned} & r_x(l_1, l_2, \dots, l_{k-1}, m_1, m_2, \dots, m_{k-1}) \\ &= \sum_l \sum_m x(l, m) x(l+l_1, m+m_1) \dots x(l+l_{k-1}, m+m_{k-1}) \end{aligned} \quad (6)$$

The inner product of the autocorrelations  $r_{xi}$  and  $r_{xj}$  of image data  $x_i(l, m)$  and  $x_j(l, m)$  is calculated by the sum of the  $k$ th power of the cross-correlation  $cc_{xi, xj}(l_1, m_1)$  of the image data [9]

$$r_{xi} \cdot r_{xj}(k) = \sum_{l=0}^{L-1} \sum_{m=0}^{M-1} \{cc_{xi, xj}(l_1, m_1)\}^k / (LM) \quad (7)$$

$$cc_{xi, xj}(l_1, m_1) = \sum_{l=1}^{L-l_1} \sum_{m=1}^{M-m_1} x_i(l, m) x_j(l+l_1, m+m_1) / (LM) \quad (8)$$

Computational costs are reduced in the practical order even for high-order  $k$  of the correlation and large data size  $L$  and  $M$  since the calculation of the explicit values of the autocorrelations are avoided. Equation (7) is employed as the  $k$ th-order correlation kernel function  $k(\mathbf{x}_i, \mathbf{x}_j)$  and thus  $\Phi$  in Eq. (3).

Since the performance of the correlation kernels of odd or higher-orders is degraded, the following modified versions of the correlation kernels have been proposed [12] and are used in this study.

$$L_p \text{ norm kernel (P)} \\ r_{xi} \cdot r_{xj} = \text{sgn}(cc_{xi, xj}(l_1, m_1)) |\sum_{l_1, m_1} \{cc_{xi, xj}(l_1, m_1)\}^k|^{1/k} \quad (9)$$

$$\text{Absolute correlation kernel (A)} \\ r_{xi} \cdot r_{xj} = \sum_{l_1, m_1} |cc_{xi, xj}(l_1, m_1)|^k \quad (10)$$

$$\text{Absolute } L_p \text{ norm kernel (AP)} \\ r_{xi} \cdot r_{xj} = |\sum_{l_1, m_1} \{cc_{xi, xj}(l_1, m_1)\}^k|^{1/k} \quad (11)$$

$$\text{Absolute } L_p \text{ norm absolute kernel (APA)} \\ r_{xi} \cdot r_{xj} = |\sum_{l_1, m_1} |cc_{xi, xj}(l_1, m_1)|^k|^{1/k} \quad (12)$$

$$\text{Max norm kernel (Max)} \\ r_{xi} \cdot r_{xj} = \max_{l_1, m_1} cc_{xi, xj}(l_1, m_1) \quad (13)$$

$$\text{Max norm absolute kernel (MaxA)} \\ r_{xi} \cdot r_{xj} = \max_{l_1, m_1} |cc_{xi, xj}(l_1, m_1)| \quad (14)$$

The  $L_p$  norm kernel (P) and the absolute correlation kernel (A) take the  $k$ th roots and absolute values, respectively, of the original ones. The max norm kernel (MAX) is regarded as the  $L_p$  norm kernel in the limit of  $k \rightarrow \infty$ . The absolute  $L_p$  norm kernel (AP), the absolute  $L_p$  norm absolute kernel (APA) and the Max norm absolute kernel (MaxA) are combinations of them.

## 2.4 Combining Correlation Kernels

The following combining methods in three levels are employed for combining autocorrelation features of different orders.

First is combination in a kernel level. The sums of the correlation kernels of different orders are used as kernel functions, which are used in [7], [8].

$$\varphi(\mathbf{x}_i, \mathbf{x}_j) = \sum_{k=1}^{km} r_{xi} \cdot r_{xj}(k) \quad (15)$$

Second is combination in a feature level. Cartesian spaces of the principal components in kPCA or the canonical variates in kCCA with the correlation kernels of different orders are used as combined feature spaces:  $(u_1, \dots, u_{n_c}, u'_1, \dots, u'_{n'_c})$  ( $0 \leq n_c, n'_c \leq C-1$ ) obtained from two sets of the canonical variates  $(u_1, \dots, u_{C-1})$  and  $(u', \dots, u'_{C-1})$ , for instance. Third is combination of classifiers output level. Classification methods are applied to the principal component spaces or the canonical variate spaces with different correlation kernels and their output (classification results) are combined. In this study, the simple nearest-neighbor classifier and the majority vote output are used as the classification method and combining classifiers, respectively.

## 2.5 Kansei Information

Kansei information can be obtained from human experts or from the public. A standard way is that we prepare impression words for objects, e.g., {beautiful, dark, clear} for images and ask the persons to choose the impression words which they think most matched to the objects. We then use the vectors the elements of which are the numbers of the votes of the impression words to the objects as Kansei information. Questionnaire systems can be employed to collect them.

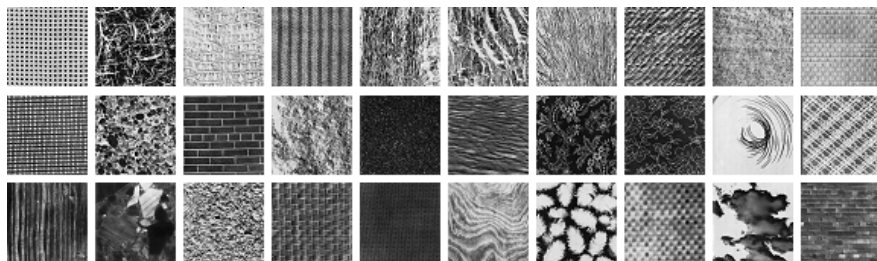
The obtained vectors, which we call the impression vectors, are used as the second feature vectors  $y$  instead of the indicator vectors in Eq. (5) in kCCA. The canonical variates obtained with them retain Kansei information and then consist of the features differ from those without them. These features can contribute to increases in classification performance.

## 3 Texture Classification Experiment

Classification experiment with 30 texture images arbitrarily taken from the Brodatz album [16] were done using kPCA and kCCA with combining correlation kernels of different orders as well as Kansei information.

### 3.1 Collection of Kansei Information

We made a questionnaire system on Web to collect Kansei information from Japanese undergraduate student in our faculty. Java Server Pages (JSP) was used for the client-server system. The Web page shows 30 texture images in the Brodatz album in Fig. 1 and asks people to choose one image which matches each Japanese impression word. Twenty impression words for images [17] are used: delicate, beautiful, bold, sharp, decorative, fine, sophisticated, simple, soft, deep, impressive, quiet, elegant, chic, natural, hard, grave, silent, solid, rural. We sent requests for the questionnaire by e-mail to about 200 students and obtained answers from 50 students.



**Fig. 1.** Texture images from the Brodatz album

Kansei information of each texture image is a 20-dimensional vector the element of which is the number of the vote of the impression word to the image (the impression vector). For instance, the impression vector for the top-left texture image (Brodatz D101), top-left in Fig. 1, was

$$\mathbf{y}_1 = (1, 1, 0, 0, 7, 1, 1, 25, 1, 0, 7, 2, 1, 0, 1, 0, 0, 3, 0, 2) \quad (16)$$

which indicates that many answerers thought the image simple. We use the impression vectors as the second feature vectors  $\mathbf{y}$  in kCCA.

### 3.2 Method

Thirty 8bit image data of 640×640 pixels in the Brodatz album are obtained from AMOVIP-DB [18]. Ten subimages of 10×10 pixels are taken from each original image without overlap, 300 images in total, are used as sample data and one hundred subimages for each, 3000 images in total, are used as test data as well.

As the second feature vector  $\mathbf{y}$  in kCCA, we use the indicator vectors (Eq. (5)) or the impression vectors, in which Kansei information is taken into account. The values of regularization parameters  $\gamma_x$  and  $\gamma_y$  in kCCA are set to be  $0.1n$ . The principal components with 299 non-zero elements in kPCA and the canonical variates with 29 elements in kCCA with the indicator vectors and 19 elements in kCCA with the impression vectors are calculated with the sample data. In kPCA, the first 50 elements of the principal components are used for classification.

A simple nearest neighbor classifier (1-NN) in the feature space (the principal component spaces, the canonical variate spaces and their combination) in each kPCA and kCCA is used for the classification of the test data. For the combining methods, the sums of the correlation kernels of different orders in a kernel level, Cartesian spaces of the principal components and the canonical variates with the correlation kernels of different orders in a feature level, and the majority vote of the 1-NN classifiers output with kPCA and kCCA with the correlation kernels are employed as explained in Sect. 2.4. Further, the impression vectors (Kansei information) as well as the indicator vectors are adopted for the second feature vectors  $\mathbf{y}$  in kCCA.

To express the set  $(M, \phi)$  of the kernel method and the kernel function  $\phi$ , the following symbols are used.

kPCA: P, kCCA with the indicator vector: I, kCCA with the impression vector K

the  $k$ th-order correlation kernel:  $C_k$ , the  $k$ th-order  $L_p$  norm kernel:  $P_k$ , the max norm kernel: Max, etc., the symbols in which are the same as Eqs. (9)-(14), for the kernel function  $\phi$

For instance, (P, C2) indicates kPCA with the 2nd-order correlation kernel.

### 3.3 Results

Figure 2 shows the correct classification rates (CCRs) of the single classifiers. The correct classification rates are calculated with the first  $r$  principal components ( $u_1, \dots, u_r$ ) ( $1 \leq r \leq 50$ ) for all  $r$  in kPCA and with the first  $i$  canonical variates ( $u_1, \dots, u_i$ ) ( $1 \leq i \leq 29$  or 19) for all  $i$  in kCCA. The maximum values of CCRs are usually obtained with parts of the principal components or the canonical variates. The highest CCRs are obtained with the max norm kernel (Max) in all cases and their values are: 0.2237 (P,  $r = 50$ ), 0.2250 (P (max),  $r = 32$ ) in kPCA; 0.2347 (I,  $i = 29$ ) with the indicator vector (I), in kCCA; 0.2237 (K,  $i = 19$ ), 0.2250 (K (max),  $i = 15$ ) with the impression

vector (K). In most cases kCCAs with the indicator vector (I) show the highest CCRs, as is expected since kCCA is superior to kPCA and the indicator vector is most suitable for classification tasks.

Figure 3 shows CCRs with the combined classifiers, the sum of the kernels (a), Cartesian spaces of the principal components and the canonical variates (b), and the voting of the 1-NN classifiers output (c). In each case, the kernels or the classifiers are added in descending order of CCRs of the single classifiers and CCRs with all elements of the principal components and the canonical variates. The highest CCRs are obtained with the combination of small numbers (1 – 14) of the classifiers, though the maximum 96 classifiers can be combined. Table 1 shows the highest CCRs and the combination of the classifiers. The maximum value 0.2560 is obtained with Cartesian subspace of canonical variates of the three correlation kernels ((I, Max), (I, AP7), (I, APA3)), and the second value 0.2553 is with Cartesian space of the six correlation kernels ((I, Max), (K, Max), (I, AP7), (I, APA3), (I, AP9), (I, A5)), both in kCCA. While the optimal combination consists of the kernels with the indicate vectors, Kansei information contributes to the second one. In the combination with the voting, the highest CCR is obtained with the combination including both kPCA and kCCA.

Since the combining with Cartesian space performs well, CCRs of all combination of two and three kernels are calculated and the optimal combination of the correlation kernels with Cartesian spaces is shown in Table 2. Using the all elements of the canonical variates, CCR increases to 0.2567 with 2 classifiers ((I, Max), (I, A3)) and 0.2684 with 3 classifiers ((I, Max), (I, AP7), (I, A7)). Using the optimal elements of the canonical variates in descending order of the corresponding eigenvalues, CCR increases to 0.2623 with 2 classifiers ((I, Max), (K, P4)) and 0.2787 with 3 classifiers ((I, Max), (I, A5), (K, C3)). They exceed the maximum CCR 0.2560 in Table 1. Note that the combination with the classifiers with low CCR in themselves, e.g., (I, C6), (I, C8), shown in Fig. 2, shows high CCRs, though the best single classifiers (I, Max) is included in most cases. The second-best single classifier (P, Max) does not appear in Table 2.

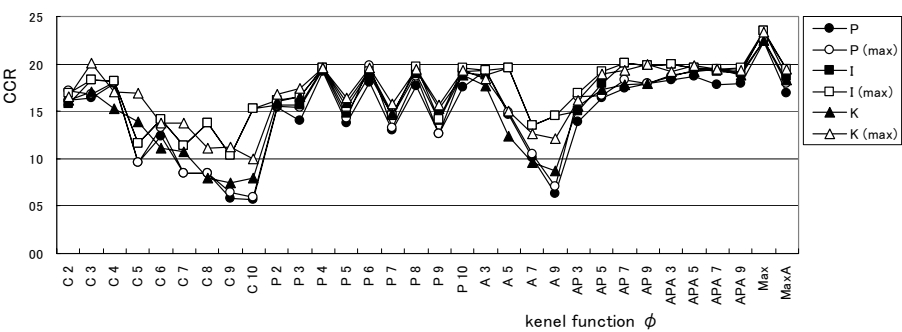


Fig. 2. Correct classification rates (CCRs) of the single classifiers

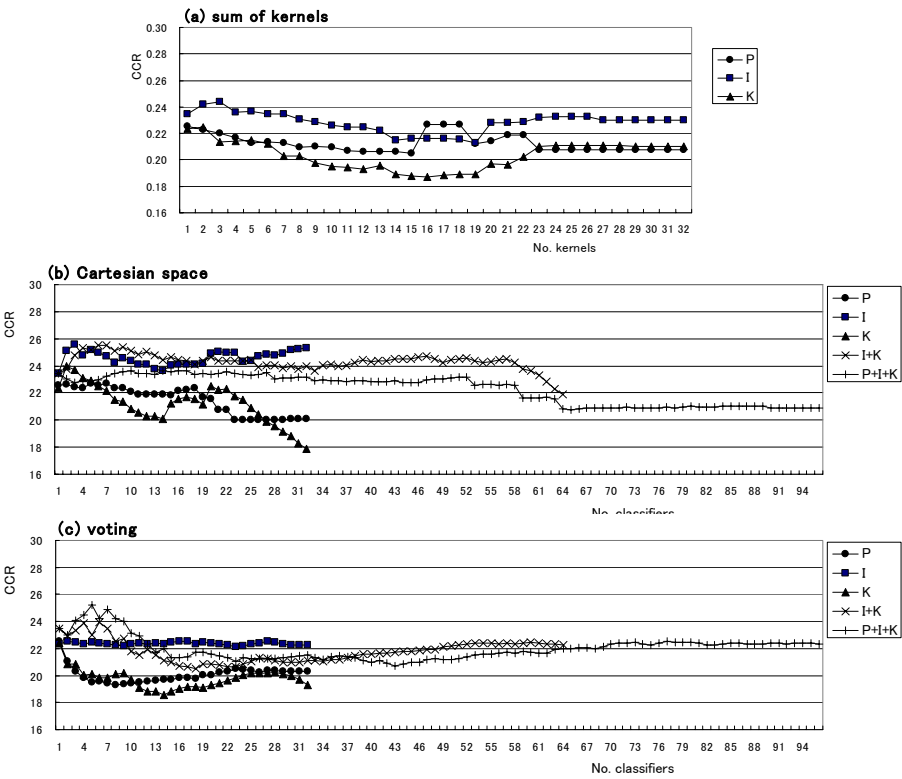


Fig. 3. CCRs of combined classifiers. The sum of the correlation kernels (a), Cartesian space of the feature vectors (b), the voting of the classifiers output (c).

Table 1. Highest CCRs of the combining classifiers in Fig3

Highest CCR (classifiers)			
	(a) Sum	(b) Cartesian	(c) Voting
P	0.2333 (P, Max)	0.2267 (P, Max), (P, P4), (P, APA7), (P, P6), (P, P8)	0.2253 (P, Max)
I	0.2440 (I, Max), (I, AP7), (I, APA3)	0.2560 (I, Max), (I, AP7), (I, APA3)	0.2347 (I, Max)
K	0.2247 (K, Max), (K, A3)	0.2397 (K, Max), (K, A3)	0.2237 (K, Max)
I+K		0.2553 (I, Max), (K, Max), (I, AP7), (I, APA3), (I, AP9), (I, A5)	0.2390 (I, Max), (K, Max), (I, AP7), (I, APA3)
P+I+K		0.2367 (I, Max), (P, Max), (K, Max), (I, AP7), (I, APA3), (I, AP9), (I, A5), (I, APA5), (P, P4), (I, P4), (P, APA7), (P, P6), (P, P8), (K, A3)	0.2521 (I, Max), (P, Max), (K, Max), (I, AP7), (I, APA3)

Table 2. Highest CCRs of the combining classifiers in Cartesian spaces of the feature vectors

Highest CCRs with Cartesian space					
2 classifiers			3 classifiers		
all elements	optimal elements		all elements	optimal elements	
0.2567 (I, Max), (I, A3)	0.2623 (I, Max) 5, (K, P4) 6		0.2683 (I, Max), (I, AP7), (I, A7)	0.2787 (I, Max) 29, (I, A5) 29, (K, C3) 9	
0.2560 (I, A7), (K, Max)	0.2617 (I, Max) 21, (I, C6) 29		0.2677 (I, Max), (I, AP7), (I, C8)	0.2767 (I, Max) 29, (I, A7) 29, (K, APA7) 6	
0.2557 (I, Max), (I, A7)	0.2610 (I, Max) 8, (I, A9) 29		0.2643 (I, Max), (I, P4), (I, A7)	0.2753 (I, Max) 12, (I, C6) 29, (I, APA3) 6	
0.2550 (I, Max), (I, C6)	0.2603 (I, Max) 18, (I, A7) 29		0.2643 (I, Max), (I, AP7), (I, A9)	0.2753 (I, Max) 29, (I, A7) 29, (K, P6) 4	
0.2550 (I, Max), (I, A9)	0.2600 (I, Max) 8, (K, C2) 29		0.2640 (I, Max), (I, C8), (K, P4)	0.2750 (I, Max) 29, (I, A7) 29, (K, P4) 5	

## 4 Discussion

The combination of the correlation kernels of different orders and Kansei information were employed in kPCA and kCCA, and their performance is compared in the experiment of texture classification. The combining with Cartesian space of the feature elements (principal components and the canonical variates) obtained with different kernels in kPCA and kCCA performed better than the use of the sum of the correlation kernels and the voting of the output of the 1-NN classifiers with multiple kPCAs and kCCAs.

As can be seen from Fig. 2, the highest CCRs were obtained with the combination of only a few classifiers. This is ascribed to the correlations between the feature elements of the correlation kernels, and then not so many classifiers can contribute to increases in CCR. However, after the CCRs once drop to the minimum values, they increase again as the numbers of the classifiers increase in most cases. This implies that the classifiers with low CCRs in themselves have a potential ability to improve classifiers performance through combining. In fact, high CCRs were obtained with the combination of the classifiers of the highest CCR and rather low CCR in Table 2. Choosing the optimal combination of the classifiers as well as the feature elements with brute-force search is impractical as their number increases. Even the calculation of CCRs for the all combination of the feature elements of three classifiers in 96 classifiers was a formidable task. It is expected that applying feature selection methods to the feature elements obtained with multiple kPCAs and kCCAs works and it is a future problem.

## References

1. Schölkopf, B. et al.: Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10, 1299–1319 (1998)
2. Kuss, M., Graepel, T.: The geometry of kernel canonical correlation analysis, Technical Report 108, Max Plank Institute for Biological Cybernetics (2002)
3. Melzer, T. et al.: Appearance models based on kernel canonical correlation analysis. *Pattern Recognition*, 36, 1961–1971 (2003)
4. Hardoon, D.R. et al.: Canonical correlation analysis; an overview with application to learning methods, Technical Report CSD-TR-03-02, Dept. of Computer Science, University of London (2003)
5. Ruiz, A., López-de-Teruel, P.E.: Nonlinear kernel-based statistical pattern analysis. *IEEE Trans. Neural Networks* 12, 16–32 (2001)
6. Müller, K.-R. et al.: An introduction to kernel-based learning algorithms. *IEEE Trans. Neural Networks* 12, 181–201 (2001)
7. Popovici, V., Thiran, J.-P.: Higher order autocorrelations for pattern classification, In: *Proc. IEEE 2001 International Conference on Image Processing (ICIP2001)*, pp. 724–727 (2001)
8. Popovici, V., Thiran, J.-P.: Pattern recognition using higher-order local autocorrelation coefficients. *Pattern Recognition Letters* 25, 1107–1113 (2004)
9. McLaughlin, J.A., Raviv, J.: Nth-order autocorrelations in pattern recognition. *Information and Control* 12, 121–142 (1968)

10. Horikawa, Y.: Comparison of support vector machines with autocorrelation kernels for invariant texture classification, In: Proc. 17th International Conference on Pattern Recognition (ICPR 2004), vol.1, 3P.Moiii-4 (2004)
11. Horikawa, Y.: Use of autocorrelation kernels in kernel canonical correlation analysis for texture classification. In: Pal, N.R., Kasabov, N., Mudi, R.K., Pal, S., Parui, S.K. (eds.) ICONIP 2004. LNCS, vol. 3316, pp. 1235–1240. Springer, Heidelberg (2004)
12. Horikawa, Y.: Modification of correlation kernels in SVM, KPCA and KCCA in texture classification, In: Proc. 2005 International Joint Conference on Neural Networks (IJCNN 2005), P2-A2fac, 2042, pp. 2006–2011 (2005)
13. Kittler, J.K. et al.: On combining classifiers. *IEEE Trans. Patt. Anal. Machine Intell.* 20, 226–239 (1998)
14. Jain, A.K. et al.: Statistical pattern recognition: A review. *IEEE Trans. Patt. Anal. Machine Intell.* 22, 4–37 (2000)
15. Kurita, T., Kato, T.: Learning of personal visual impression for image database systems, In: Proc. the Second International Conference on Document Analysis and Recognition, pp. 547–552 (1993)
16. Brodatz, P.: Textures: A Photographic Album for Artists and Designers. Dover, New York (1966)
17. Chijjiwa, H.: Color Science (in Japanese), Fukumura, Japan, (J.Suzuki's Website, <http://www.teu.ac.jp/chiit/~jsuzuki/kanseiwr.html>) (1983)
18. AMOVIP-DB, <http://www.iv.optica.csic.es/projects/database.html>