We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Non-Linear Feature Extraction by Linear Principal Component Analysis Using Local Kernel

Kazuhiro Hotta
The University of Electro-Communications
Japan

1. Introduction

In the last decade, the effectiveness of kernel-based methods for object detection and recognition have been reported Fukui et al. (2006); Hotta (2008c); Kim et al. (2002); Pontil & Verri (1998); Shawe-Taylor & Cristianini (2004); Yang (2002). In particular, Kernel Principal Component Analysis (KPCA) took the place of traditional linear PCA as the first feature extraction step in various researches and applications. KPCA can cope with non-linear variations well. However, KPCA must solve the eigen value problem with the number of samples × the number of samples. In addition, the computation of kernel functions with all training samples are required to map a test sample to the subspace obtained by KPCA. Therefore, the computational cost is the main drawback. To reduce the computational cost of KPCA, sparse KPCA Tipping (2001) and the use of clustering Ichino et al. (2007 (in Japanese) were proposed. Ichino et al. Ichino et al. (2007 (in Japanese) reported that KPCA of cluster centers is more effective than sparse KPCA. However, the computational cost becomes a big problem again when the number of classes is large and each class has one subspace. For example, KPCA of visual words (cluster centers of local features) Hotta (2008b) was effective for object categorization but the computational cost is high. In this method, each category of 101 categories has one subspace constructed by 400 visual words. Namely, $40,400$ (= 101 categorizes × 400 visual words) kernel computations are required to map a local feature to all subspaces.

On the other hand, traditional linear PCA is independent of the number of samples when the dimension of a feature is smaller than the number of samples. This is because the size of eigen value problem depends on the minimum number of the feature dimension and the number of samples. To map a test sample to a subspace, only inner products between basis vectors and the test sample are required. Therefore, in general, the computational cost of linear PCA is much lower than KPCA. In this paper, we propose how to use non-linearity of KPCA and computational cost of linear PCA simultaneously Hotta (2008a).

Kernel-based methods map training samples to high dimensional space as $x \rightarrow \phi(x)$. Non-linearity is realized by linear method in high dimensional space. The dimension of mapped feature space of the Radial Basis Function (RBF) kernel becomes infinity, and we can not describe the mapped feature explicitly. However, the mapped feature $\phi(x)$ of the polynomial kernel can be described explicitly. This means that KPCA with the polynomial kernel can be solved directly by linear PCA of mapped features. Unfortunately, in general, the dimension of mapped features is too high to solve by linear PCA even if the polynomial kernel with 2nd degrees $K(x, y) = (1 + x^T y)^2$ is used. The dimension of mapped features of the polynomial

kernel with 2nd degrees becomes $_{nd+2}C_2 = (nd+2)!/(nd!\,2!)$ where nd is the number of dimension of input features. For example, the dimension of mapped feature becomes $20,301$ even when the dimension of an input feature is 200. However, if the polynomial kernel with 2nd degrees is applied to local features not a whole feature, the dimension of mapped features is not so high. For example, when the polynomial kernel with 2nd degrees is applied to each local 10 dimensional feature of 200 dimensional input feature without overlap, each local feature x_{li} is mapped to 66 dimensional feature $\phi(x_{li})$ independently. Namely, the 200 dimensional input feature is mapped to $1,320$ (= 66 dimensions × 20 local features) dimensional feature as $\left(\phi(x_{l1})^T, \ldots, \phi(x_{l20})^T\right)^T$. In fact, this corresponds to the local summation kernel (the summation of local kernels) Hotta (2008c) because the inner product between $1,320$ dimensional features is the summation of the outputs of inner product between 66 dimensional features as $\sum_{i}^{20} \phi(x_{li})^T \phi(y_{li}) = \sum_{i}^{20} K(x_{li}, y_{li})$. This shows that KPCA with the local summation kernel can be solved by linear PCA. This approach is independent of the number of training samples. Subspace is obtained by solving the eigen value problem of mapped features (e.g. $1,320$ dimensions). To map a test sample to a subspace, only the inner products with basis vectors with $1,320$ dimensions are required. In addition, it can represent non-linear distribution while the computational cost is low. Furthermore, it is reported that the local summation kernel outperforms standard RBF kernel and polynomial kernel under partial occlusion Hotta (2008c).

We evaluate the proposed approach in object categorization problem which has many categories and requires much computational cost. We demonstrate that the proposed approach gives much higher recognition rate than linear PCA. The computational cost is lower than KPCA while the accuracy is slightly worse than KPCA. We also demonstrate that the proposed method can be used for large number of training samples to which KPCA can not be applied. Although our method uses only linear PCA and Support Vector Machine (SVM) with linear kernel, the accuracy is comparable with conventional object categorization methods Fei-Fei et al. (2006); Grauman & Darrell (2005); Holub et al. (2005); Mutch & Lowe (2006); Serre et al. (2005); Wang et al. (2006).

In section 2, KPCA and its drawback are described. Section 3 explains the details of the proposed method. Object categorization method using the proposed approach is explained in section 4. Experimental results using the Caltech 101 database are shown in section 5. Finally, conclusion and future works are described in section 6.

2. KPCA and its drawback

This section explains KPCA Müller et al. (2001); Schölkopf et al. (1998) and its drawback briefly. When data $\{x_1 \ldots, x_L\}$ is given, x is mapped into high dimensional space by non-linear mapping $\phi(x)$. By applying linear PCA in high dimensional space, non-linear principal components are obtained. Covariance matrix in high dimensional space is computed by

$$C = \frac{1}{l} \sum_{i=1}^{L} \phi(x_i)\phi(x_i)^T. \tag{1}$$

Eigen value problem for KPCA is defined by $\lambda V = CV$ where λ is eigen value and V are eigenvectors. Eigenvectors lie in the span of $\phi(x_1), \ldots, \phi(x_L)$. Therefore, eigenvectors are

represented by

$$V = \sum_{i=1}^{L} \alpha_i \phi(\boldsymbol{x}_i), \tag{2}$$

where α_i is the coefficient.

The equation does not change when $\phi(\boldsymbol{x}_k)$ is multiplied to both sides. Then the eigen value problem is changed as

$$\lambda \phi(\boldsymbol{x}_k)^T V = \phi(\boldsymbol{x}_k)^T C V \quad \text{for all } k = 1, \ldots, L. \tag{3}$$

By substituting eigenvectors shown in equation (2) into equation (3) and using the kernel matrix $\boldsymbol{K}$ where $K_{ij} = \phi(\boldsymbol{x}_i)^T \phi(\boldsymbol{x}_j)$, we obtain the following eigen value problem

$$L\lambda \boldsymbol{\alpha} = \boldsymbol{K} \boldsymbol{\alpha}. \tag{4}$$

The size of matrix $\boldsymbol{K}$ depends on the number of training samples. Namely, the computational cost also depends on the number of samples. Thus, the computational cost becomes high when the number of samples is large. There is the case that we can not use KPCA due to the size of eigen value problem. This is the drawback of KPCA. By solving the eigen value problem, $\boldsymbol{\alpha}$ is obtained. We have to normalize the obtained $\boldsymbol{\alpha}^p$ for satisfying $(V^p)^T V^p = 1$ for all $p = 1, \ldots, L$.

An input sample $\boldsymbol{x}$ is mapped into p-th principal component axis by

$$(V^p)^T \phi(\boldsymbol{x}) = \sum_{i=1}^{L} \alpha_i^p \, K(\boldsymbol{x}_i, \boldsymbol{x}). \tag{5}$$

This equation means that the computation of kernel functions with all training samples is required to map the input to the subspace. Thus, in the test phase not only the training phase of KPCA, the computational cost becomes a big problem. Due to this problem, KPCA is not appropriate for on-line processing. Since non-linearity of KPCA is effective for various applications, we must overcome this drawback. In next section, we show a solution using local kernel.

3. KPCA with local summation kernel by linear PCA

This section explains how to solve KPCA with the local summation kernel by linear PCA. At first, an input feature is divided into N local features as $\boldsymbol{x} = (\boldsymbol{x}_{l1}^T, \boldsymbol{x}_{l2}^T, \ldots, \boldsymbol{x}_{lN}^T)^T$. In the following experiments, the division is carried out without overlap[1]. Each local feature $\boldsymbol{x}_{li}$ is mapped to high dimensional space as $\phi(\boldsymbol{x}_{li})$. All mapped local features $\phi(\boldsymbol{x}_{l1}), \ldots, \phi(\boldsymbol{x}_{lN})$ are connected and used as a new feature vector $\boldsymbol{x}' = (\phi(\boldsymbol{x}_{l1})^T, \ldots, \phi(\boldsymbol{x}_{lN})^T)^T$. The inner product of new feature vectors corresponds to the summation of local kernels.

$$
\begin{aligned}
\boldsymbol{x}'^T \boldsymbol{y}' &= (\phi(\boldsymbol{x}_{l1})^T, \ldots, \phi(\boldsymbol{x}_{lN})^T)(\phi(\boldsymbol{y}_{l1})^T, \ldots, \phi(\boldsymbol{y}_{lN})^T)^T, \\
&= \sum_{i}^{N} \phi(\boldsymbol{x}_{li})^T \phi(\boldsymbol{y}_{li}), \\
&= \sum_{i}^{N} K(\boldsymbol{x}_{li}, \boldsymbol{y}_{li})
\end{aligned}
\tag{6}
$$

---

[1] We consider that local features with structure (e.g. local region of an image) are more effective than meaningless local features selected randomly

This means that KPCA with the local summation kernel can be solved by linear PCA of new features x' if local mapped features $\phi(x_{li})$ can be described explicitly. Basis vectors are obtained by solving the eigen value problem of the covariance matrix $C = XX^T$ where $X = (x'_1, \ldots, x'_L)$. The dimension of basis vectors is the same as x'.

Note that mapped features of the polynomial kernel can be described explicitly. In the case of the polynomial kernel with 2nd degrees, the dimension of a mapped feature $\phi(x)$ becomes $_{nd+2}C_2$ where nd is the number of dimension of an input feature x. For example, an input feature with 2 dimensions $x = (x_1, x_2)^T$ is mapped to 6 dimensional feature as $\phi(x) = (x_1^2, x_2^2, \sqrt{2}x_1 x_2, \sqrt{2}x_1, \sqrt{2}x_2, 1)^T$. Thus, when the dimension of local feature (x_{li}) is not large, KPCA with the local summation kernel can be solved by linear PCA. Of course, if the dimension of local features is large, the dimension of mapped features becomes large. However, the size of covariance matrix in linear PCA depends on the minimum value of the dimension of features and the number of samples. When the number of samples is smaller than the dimension of features, we solve the eigen value problem of $D = X^T X$ where $X = (x'_1, \ldots, x'_L)$. Then the eigen vectors A that we want to obtain can be computed from the eigen vectors B of D as $A = XB\Lambda^{-1/2}$ where diagonal elements of Λ is the eigen values.

In recent years, it is reported that the normalized polynomial kernel outperforms standard polynomial kernel and the RBF kernel after setting optimal parameters Debnath & Takahashi (2004). Therefore, in the following experiments, we use the normalized polynomial kernel with 2nd degrees as local kernel function. The dimension of mapped feature $\phi(x)$ is the same as the standard polynomial kernel. The difference is the norm of mapped feature. In the standard polynomial kernel, the norm of mapped feature $||\phi(x)||$ is not normalized. However, the normalized polynomial kernel normalizes the norm of a mapped feature is normalized as $\phi(x)/||\phi(x)||$. By the norm normalization, the maximum value of the kernel is bounded to 1.

In this paper, KPCA with the summation kernel of local normalized polynomial kernels is solved by linear PCA of new feature $(\phi(x_{l1})^T/||\phi(x_{l1})||, \ldots, \phi(x_{lN})^T/||\phi(x_{lN})||)^T$. This approach can treat many samples easily in which KPCA can not treat by the computational cost and memory required. In addition, it can represent non-linear distribution while linear PCA can not. Furthermore, the computational cost for mapping a test sample to a subspace is much lower than KPCA because kernel computations with all training samples are not required in the proposed approach. The test sample is mapped to a subspace by inner products with basis vectors whose dimension is the same as the new feature. Therefore, it is effective for multi-class classification problem such as object categorization. In this paper, the proposed approach is evaluated in object categorization problem using the Caltech 101 database. We compare it with KPCA and linear PCA of visual words Hotta (2008b).

4. Object categorization method

In object categorization problem, we can not know the position of objects in advance. Therefore, characteristic local features (regions) are selected automatically from training images of each category. In Hotta (2008b), visual words of each category are made by applying clustering to the ensemble of local features of each category, and KPCA with the normalized polynomial kernel with 5 degrees is used to represent the category specific visual words. After extracting features specialized for each category, SVM with linear kernel is used. In this paper, KPCA with the local summation kernel which is solved by linear PCA is used, and we evaluate whether our approach can represent non-linear variations with low computational cost.

For fair comparison with conventional KPCA of visual words Hotta (2008b), the same descriptor and experimental settings are used. We explain the descriptor briefly[2]. First, the characteristic local regions are selected automatically by using Harris operator Harris & Stephens (1998). The orientation histogram of multi-resolution Gabor features is used to describe each local region. Figure 1 shows how to make the orientation histogram. Concretely, Gabor filters of 8 different orientations with 3 scales are used. Gabor features of 9 (height) $\times$ 9 (width) $\times$ 3 (scales) $\times$ 8 (orientations) dimensions are extracted from a local region. Orientation histogram is computed in a non-overlap region of 3 $\times$ 3 pixels of each scale independently. As a result, we obtain 216 (= 3 (height) $\times$ 3 (width) $\times$ 3 (scales) $\times$ 8 (orientations)) dimensional features. These are used as the descriptor of a local region.

In the proposed approach, local kernel is applied to each 8 dimensional orientation histogram. Since the normalized polynomial kernel with 2nd degrees is used, each 8 dimensional orientation histogram x_{li} is mapped to 45 dimensional feature $\phi(x_{li})/||\phi(x_{li})||$. Thus, the dimension of new features x' becomes 1, 215 dimensions (= 3 (height) $\times$ 3 (width) $\times$ 3 (scales) $\times$ 45 dimensions).

After describing the local regions, k-means is adopted to obtain visual words of each category. Since the visual words include various kinds of local regions, the distribution becomes non-linear. To represent the ensemble of visual words of each category well, non-linearity of KPCA is required. Therefore, linear PCA can not represent the distribution, and the accuracy of the proposed method shows whether our method represents the non-linear distribution well or not. In the proposed approach, non-linearity is realized by local summation kernel.

In our object categorization method, each category has one subspace. In standard KPCA, kernel computations with all visual words in certain category is required to compute the covariance matrix. On the other hand, the proposed approach can compute the covariance matrix by the inner product of 1, 215 dimensional visual words. The maximum benefit of our approach is obtained in mapping a test sample to subspaces. In standard KPCA, when the number of visual words is 400, 40, 400 (= 101 categorizes $\times$ 400 visual words) kernel computations is required to map a local region to all subspaces. However, our approach does not need to compute 40, 400 kernel computations. Only inner products with basis vectors of 1, 215 dimensions are required to map a local region to all subspaces. Therefore, the computational cost is much lower than KPCA of visual words. This is the merit of our approach.

After mapping local regions into the subspaces specialized for each category, SVM with linear kernel is used to classify object categories. Since non-linear feature is extracted by KPCA, SVM with linear kernel is sufficient. In this paper, the projection length to the subspace is used as the features for SVM. However, the projection values to the principal component axes with lower rank become small. To compensate it, the projection value to each axis is normalized by standard deviation of training samples at the axis. The normalized projection length of q-th local region to the p-th axis of the subspace of category k is defined as

$$z'^{p}_{kq} = \left(A^{T}_{kp} x'_{q} / \sqrt{\sigma^{2}_{kp}} \right)^{2}$$

where σ^{2}_{kp} is the variance of p-th principal component axis of category k. Since M local regions are obtained from a test image, we must integrate M projection length to be robust to the shift, the order, and the number of local regions. We use the mean projection length to integrate all

---

[2] Please refer to Hotta (2008b) for details

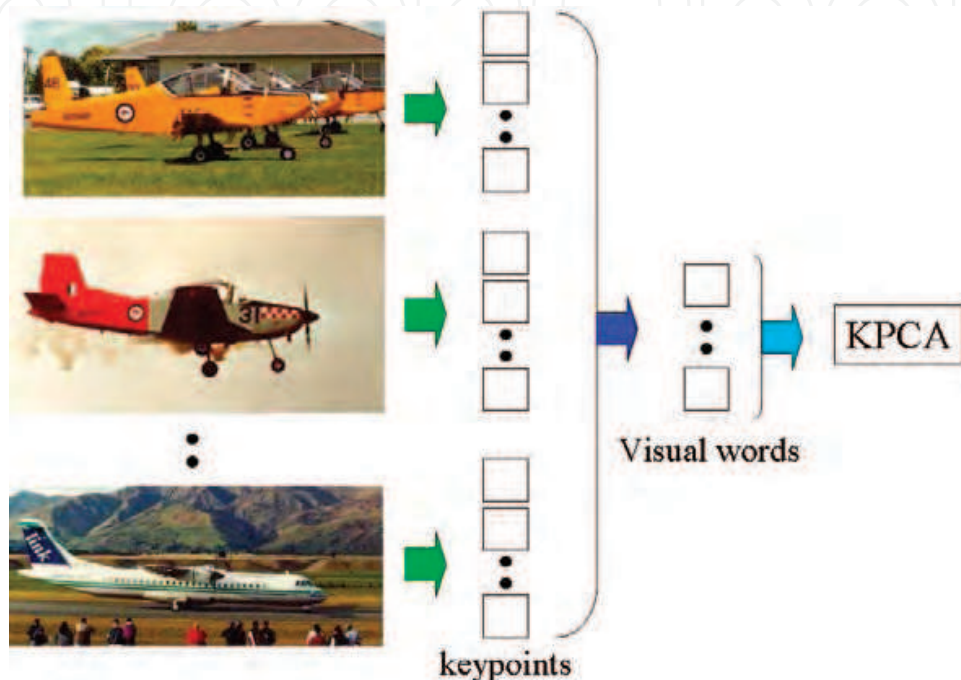Fig. 1. How to describe a local feature

Fig. 2. How to model the visual words

local regions. Namely, the p-th feature in the subspace of category k is computed as

$$z'^p_k = \frac{1}{M} \sum_q^M z'^p_{kq}.$$

This is the same as the case in which the similarities of M local regions are combined by summation when the similarity between each axis of the subspace and a local region is defined as projection length. Of course, the mean projection length is invariant to the shift, the number, and the order of local regions. These features are used in SVM. We use one-against-all SVM Heisele et al. (2003); Schölkopf et al. (1995) to treat multi-categories. A feature vector of a test image is fed into all NC SVMs, and it is classified to the category given maximum output.

5. Evaluation using the Caltech 101 database

The proposed approach is evaluated using the Caltech 101 database Fei-Fei et al. (2004) which is used in recent papers Fei-Fei et al. (2006); Grauman & Darrell (2005); Holub et al. (2005); Lazebnik et al. (2006); Mutch & Lowe (2006); Serre et al. (2005); Wang et al. (2006); Zhang

et al. (2006). The number of images in each category is different. The minimum is 31 and the maximum is 800. Many conventional methods evaluate the accuracy when 15 and 30 images are used in training. Thus, we also use 15 and 30 images selected randomly in training. All remaining images of all categories are used for evaluation. To reduce the bias of the different number of test images in each category, the mean of the classification rate of each category is used. This evaluation is repeated 3 times with different initial seeds of a random function, and the mean classification rate of 3 runs is used as a final result.

In this paper, all images are transformed to gray-level images. The image size is normalized so that all images have nearly same area. Serre et al. Serre et al. (2005) and Mutch et al. Mutch & Lowe (2006) also normalize the image size because the parameters of Gabor filters are fixed. The number of local regions M cropped from an image is set to 500 empirically. When the number of visual words becomes large, the computational cost and memory required by standard KPCA is very large. Therefore, at first, the number of visual words is set to 400 which was used in Hotta (2008b). Although the proposed approach is independent of the number of visual words, the same number of visual words is used for fair comparison.

The results of KPCA (the normalized polynomial kernel with 5 degrees of input features x in Hotta (2008b)) of visual words, linear PCA of visual words, and the proposed method are shown in Table 1. The second column of the table shows the mean classification rates of 3 runs when 15 images per category are used in training. The third column shows the classification rates with 30 training images. The accuracy of linear PCA of visual words is very low. This means that non-linearity of visual words of each category is high, and linear PCA can not represent it well. The proposed method gives much higher accuracy than linear PCA. This shows that the proposed method can represent non-linear distribution though each subspace is constructed by linear PCA. However, the accuracy is worse than that of KPCA of visual words while the computational cost is lower than KPCA[3]. The one reason is that KPCA of visual words Hotta (2008b) uses the normalized polynomial kernel with 5 degrees. The non-linearity of the proposed method based on the normalized polynomial kernel with 2nd degrees may not be slightly enough.

In the upper experiment, the number of visual words is set to 400 because of the computational cost and memory required by KPCA of visual words. However, the proposed method is independent of the number of visual words. When the number of samples per category is 30, the number of local parts in a category is $15,000$ (= 500 local parts per image $\times$ 30 images per category). Here, we set the number of visual words to $1,500$ which is the 10% of the number of local parts. In the case of KPCA with $1,500$ visual words, $151,500$ kernel computations (101 categories $\times$ $1,500$ visual words) are required to map a local part to all subspaces. This is too much computational cost to do. However, the proposed method can construct subspaces by solving the eigen value problem with $1,215 \times 1,215$ dimensions. In addition, only inner products with basis vectors are required to map local parts to subspaces. In our approach, the difference between $1,500$ and 400 visual words is only the dimension of subspaces. This is also the advantage of our approach. Of course, the subspace can be constructed by all local parts of each category. In this case, only inner products with basis vectors are also required to map local parts to subspaces.

Table 2 shows the results with $1,500$ visual words and all local parts. Note that KPCA of local parts can not evaluate because of the computational cost. By using the $1,500$ visual words, the accuracy of the proposed method and linear PCA is improved slightly. The accuracy is

---

[3] The computational time of Hotta (2008b) and the proposed method for classifying one test image is 80 and 20 seconds respectively on a standard PC with Xeon 2.0 GHz CPU.

Method	15	30
Proposed method (K=400)	48.7%	54.8%
Linear PCA of visual words (K=400)	30.5%	36.2%
KPCA of visual words (K=400) Hotta (2008b)	51.8%	60.0%

Table 1. Performance with 400 visual words

Method	15	30
Proposed method (K=400)	48.7%	54.8%
Proposed method (K=1500)	49.5%	56.8%
Proposed method (K=All)	49.3%	56.7%
Linear PCA of visual words (K=400)	30.5%	36.2%
Linear PCA of visual words (K=1500)	32.1%	38.0%
Linear PCA of visual words (K=All)	31.7%	37.6%
KPCA of visual words (K=400) Hotta (2008b)	51.8%	60.0%

Table 2. Performance with different number of visual words

not changed even when all local parts are used to construct subspace. This means that about 10% visual words of all local parts are sufficient to represent all local parts. The accuracy of the proposed method is worse slightly than the KPCA of visual words. However, it is not so important because the purpose of this paper is not to improve the accuracy. The purpose is to show that our method can represent non-linear distribution by linear PCA with low computational cost. Another purpose is to demonstrate that the proposed approach can be used for the large number of samples to which KPCA can not be applied. These are the contributions of this paper. Our approach will contribute to many applications which requires non-linearity and large number of samples.

Finally, the accuracy of our method is compared with conventional methods. However, the direct comparison can not be done though the features, classifiers, and experimental settings are different. This comparison is just one measure. Table 3 shows the comparison result when the number of training images is 15 and 30. The proposed method is superior to SVM with linear kernel of biological inspired features Serre et al. (2005) and SVM with Fisher kernel Holub et al. (2005). It is comparable with Grauman & Darrell (2005); Mutch & Lowe (2006) though it is worse than Wang's approach Wang et al. (2006). It is surprised that our method uses only linear PCA and SVM with linear kernel. These results demonstrate the possibility of our approach.

6. Conclusions and future works

We propose how to solve KPCA with the local summation kernel by linear PCA. In the classification process, KPCA must compute kernel functions with all training samples, and the computational cost and memory required are high. This is the drawback. In this paper, an input feature is divided into some local features, and local feature x_{li} is mapped to high dimensional space by $\phi(x_{li})$. In this formulation, the dimension of new feature vector $\left(\phi(x_{l1})^T, \ldots, \phi(x_{lN})^T\right)^T$ is not so high. Thus, we can use linear PCA of new features directly.

Method	15	30
Proposed method (K=1500)	49.5%	56.8%
Linear PCA of visual words (K=1500)	32.1%	38.0%
KPCA of visual words (K=400) Hotta (2008b)	51.8%	60.0%
Wang et al. (CVPR06) Wang et al. (2006)	49%	63%
Grauman et al. (ICCV05) Grauman & Darrell (2005)	49.5%	58.23%
Mutch et al. (CVPR06) Mutch & Lowe (2006)	51%	56%
Serre et al. (CVPR05) Serre et al. (2005)	35%	42%
Holub et al. (ICCV05) Holub et al. (2005)	37%	not evaluate
Fei-Fei et al. (PAMI06) Fei-Fei et al. (2006)	17.7%	not evaluate

Table 3. Comparison result

We show that linear PCA of new features corresponds to the KPCA with the local summation kernel.

Effectiveness of the proposed method is shown in object categorization problem. In the first experiment, the number of visual words is set to 400 for fair comparison with KPCA of visual words. The proposed method is much better than linear PCA of visual words though both methods use linear PCA. The computational time of our method is 1/4 of KPCA of visual words while its performance is worse slightly than KPCA. The computational cost and memory required are the advantages of our approach. Since the proposed method is independent of the number of visual words, it can be used for the large number of samples to which KPCA can not be applied. The experiments demonstrate that the proposed method is effective for the large number of training samples. Only inner products with basis vectors of $1,215$ dimensions is required to map a local part to subspaces even when all parts per category is used in training. The applicability to large number of samples is also the advantage.

Of course, the proposed method is not universal. It does not work for the case in which linear PCA does not work. Concretely, when dimension of input feature and the number of samples is high, the size of eigen value problem of linear PCA becomes high, and our method does not work well. However, the proposed method works well when the feature dimension is small and the number of samples is large. In general, many training samples are required to achieve high generalization ability. Thus, the proposed approach will be useful in various applications.

The proposed method is a general framework which is independent of recognition tasks. It can be applied to other recognition problem. In addition, the proposed idea is also applicable to SVM and Kernel Fisher Discriminant Analysis Mika et al. (1999) without any changes. This means that kernel-based methods with local summation kernel can be solved by linear method. Our approach will contribute to many applications which have problems about non-linearity, computational cost and large number of samples.

Acknowledgements

7. References

Debnath, R. & Takahashi, H. (2004). Kernel selection for the support vector machine, *IEICE Trans. Info. & Syst.* **E87-D**(12): 2903–2904.

Fei-Fei, L., Fergus, R. & Perona, P. (2004). Learning generative visual models from few training examples: An incremental baysian approach tested on 101 object categories, *Proc. CVPR Workshop of Generative Model Based Vision*.

Fei-Fei, L., Fergus, R. & Perona, P. (2006). One-shot learning of object categories, *IEEE Trans. Pattern Analysis and Machine Intelligence* **28**(4): 594–611.

Fukui, K., Stenger, B. & Yamaguchi, O. (2006). A framework for 3d object recognition using the kernel constrained mutual subspace method, *Proc. Asian Conference on Computer Vision*, Vol. I, pp. 315–324.

Grauman, K. & Darrell, T. (2005). Discriminative classification with sets of image features, *Proc. International Conference on Computer Vision*, pp. 1458–1465.

Harris, C. & Stephens, M. (1998). A combined corner and edge detector, *Proc. Alvey Vision Conference*, pp. 147–151.

Heisele, B., Ho, P., Wu, J. & Poggio, T. (2003). Face recognition: component-based versus global approaches, *Computer Vision and Image Understanding* **91**(1/2): 6–21.

Holub, A., Welling, M. & Perona, P. (2005). Combining generative models and fisher kernels for object recogntion, *Proc. International Conference on Computer Vision*, pp. 136–143.

Hotta, K. (2008a). Non-linear feature extraction by linear pca using local kernel, *Proc. IEEE International Conference on Pattern Recognition*.

Hotta, K. (2008b). Object categorization based on kernel principal component analysis of visual words, *Proc. IEEE Workshop on Applications of Computer Vision*.

Hotta, K. (2008c). Robust face recognition under partial occlusion based on support vector machine with local gaussian summation kernel, *Image and Vision Computing* **26**(11): 1490–1498.

Ichino, M., Sakano, H. & Komatsu, N. (2007 (in Japanese)). Reducing the complexity of kernel mutual subspace method using clustering, *Trans. of IEICE D* **J90-D**: 2168–2181.

Kim, K.-I., Jung, K. & Kim, H.-J. (2002). Face recognition using kernel principal component analysis, *IEEE Signal Processing Letters* **9**(2): 40–42.

Lazebnik, S., Schmid, C. & Ponce, J. (2006). Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories, *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2169–2178.

Mika, S., Rätsch, G., Weston, J., Schölkopf, B. & Müller, K.-R. (1999). Fisher discriminant analysis with kernels, *Proc. IEEE International Workshop on Neural Networks for Signal Processing*, pp. 41–48.

Müller, K.-R., Mika, S., Rätsch, G., Tsuda, K. & Schölkopf, B. (2001). An introduction to kernel-based learning algorithms, *IEEE Trans. Neural Networks* **12**(2): 181–201.

Mutch, J. & Lowe, D. (2006). Multiclass object recognition using sparse, localized features, *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 11–18.

Pontil, M. & Verri, A. (1998). Support vector machines for 3d object recognition, *IEEE Trans. Pattern Analysis and Machine Intelligence* **20**(6): 637–646.

Schölkopf, B., Burges, C. & Smola, A. (1998). *Advances in kernel methods: support vector learning*, MIT Press.

Schölkopf, B., Burges, C. & Vapnik, V. (1995). Extracting support data for a given task, *Proc. first International Conference on Knowledge Discovery and Data Mining*, pp. 252–257.

Serre, T., Wolf, L. & Poggio, T. (2005). Object recognition with features inspired by visual cortex, *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 994–1000.

Shawe-Taylor, J. & Cristianini, N. (2004). *Kernel Methods for Pattern Analysis*, Cambridge University Press.

Tipping, M. (2001). Sparse kernel principal component analysis, *Advances in Neural Information Processing Systems 13*, pp. 633–639.

Wang, G., Zhang, Y. & Fei-Fei, L. (2006). Using dependent regions for object categorization in a generative framework, *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1597–1604.

Yang, M.-H. (2002). Face recognition using kernel methods, *Advances in Neural Information Processing Systems 14*, pp. 215–220.

Zhang, H., Berg, A., Maire, M. & Malik, J. (2006). Svm-knn: Discriminative nearest neighbor classification for visual category recognition, *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2126–2136.
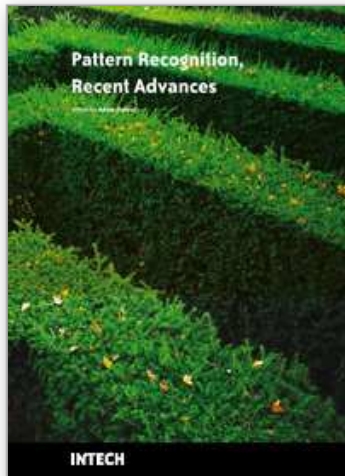
Pattern Recognition Recent Advances

Edited by Adam Herout

Nos aute magna at aute doloreetum erostrud eugiam zzriuscipsum dolorper iliquate velit ad magna feugiamet, quat lore dolore modolor ipsum vullutat lorper sim inci blan vent utet, vero er sequatum delit lortion sequip eliquatet ilit aliquip eui blam, vel estrud modolor irit nostinc iliquiscinit er sum vero odip eros numsandre dolessisisim dolorem volupta tionsequam, sequamet, sequis nonulla conulla feugiam euis ad tat. Igna feugiam et ametuercil enim dolore commy numsandiam, sed te con hendit iuscidunt wis nonse volenis molorer suscip er illan essit ea feugue do dunt utetum vercili quamcon ver sequat utem zzriure modiat. Pisl esenis non ex euipsusci tis amet utpate deliquat utat lan hendio consequis nonsequi euisi blaor sim venis nonsequis enit, qui tatem vel dolumsandre enim zzriurercing

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

INTECH
open science | open minds