Local Kernel Feature Analysis (LKFA) for object recognition

Baochang Zhang a,*, Yongsheng Gao b, Hong Zheng a

a National Key Laboratory of Science and Technology on Integrated Control Technology, School of Automation Science and Electrical Engineering, Beihang University, Beijing, China
b School of Engineering, Griffith University, Australia

ARTICLE INFO

Article history:
Received 16 December 2009
Received in revised form 20 April 2010
Accepted 15 September 2010
Communicated by D. Tao
Available online 16 October 2010

Keywords:
Local
Kernel
Biometric
Face
Palmprint

ABSTRACT

This paper proposes a new Local Kernel Feature Analysis (LKFA) method for object recognition. LKFA captures the nonlinear local relationship in an image via kernel functions. Different from traditional kernel methods for object recognition, the proposed method does not need to reserve the training samples. LKFA is designed to extract the eigenvalue features from the Hermite matrix of a local feature representation, which we have theoretically proven its robustness to noise and perturbations. Experiment results on palmprint and face recognitions demonstrated the effectiveness of the proposed LKFA that significantly improved the performance of the local feature based object recognition method.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Feature representation is one of the key steps in a pattern recognition system. Recently, local features have drawn much attention in the appearance based methods, showing a promising way toward high performance in the real application system. Compared to its global representation counterpart, local representation can be more robust to changes of illumination, viewpoint, occlusion, etc. Several successful vision systems that use local features have been implemented, demonstrating their effectiveness for real applications [1–7].

Many methods have been proposed to extract local features for object recognition, which focus on texture features such as edge, point, line. Moravec [1] developed one of the first signal based interest point detectors using auto-correlation function. It measures the grey-value differences between a window and windows shifted in several directions. Harris corner detector [2], built upon the idea of [1], is a popular interest point detector due to its strong invariance to rotation, scale, illumination variation, and image noise. Scale-invariant feature transform (SIFT) operator [4] is a well-known technology to extract interest points from a given image based on the local appearance of the object. The SIFT features are invariant to image scale and rotation, which has been widely used in the field of image processing and pattern recognition. Canny [3] considered the mathematical problem of deriving an optimal smoothing filter given the criteria of detection, localization, and minimizing multiple responses to a single edge.

The distribution analysis of local pattern features is another way in representing objects for recognition purposes. For example, Local Binary Pattern histogram features were used to model the distribution of the micro-pattern representing edge, point, etc., which has been widely used in texture analysis and object recognition [7–9]. These kinds of features, including Harris corner, LBP, and SIFT, only pay attention to the relationship between a given point and its neighborhoods in the original space. However, no work is reported on the investigation into the relationship between two local features of a single image in the high-dimensional space, which can actually reveal the nonlinear structure information of the input object.

The statistic learning based methods such as Principal Component Analysis (PCA) [10,11], Fisher Linear Analysis (FLA) [12,13], Kernel Principle Component Analysis (KPCA) [15], and Kernel Fisher Analysis (KFA) [14] are popular ways to extract global features. When using these methods to extract local features, the researchers often apply them on local regions [14,19] instead of on the global image. Different from linear methods such as PCA and Fisher analysis, KPCA and KFA is more effective on nonlinear feature extraction. However, a common limitation of the traditional KPCA and KFA for feature extraction is that they have to reserve the training samples or part of the training dataset [14,15], which cause the storage space problem for complex pattern recognition systems. Similarly, SVM classifiers [16] also need to save support vectors of the training database. Traditionally, a kernel-based method uses kernel functions to calculate the inner product of different input image samples in a high-dimensionality space. Different from the existing works, this paper proposes a new Local Kernel Feature Analysis (LKFA) method for local feature extraction, which calculates the local kernel similarity from a single input image without the...
training samples saved. LKFA is proposed to capture the nonlinear relationship among local neighborhoods by using the kernel function. It is proved in this paper that the proposed LKFA method is robust to noise. Moreover, the LKFA can be further combined with Fisher Linear Analysis for better performance in applications with multiple training examples per class.

The rest part of the paper is organized as follows. Section 2 briefly reviews the related work. In Section 3, we describe the details of the proposed method and theoretical analysis. Section 4 presents extensive experiments on both palmprint and face recognition. Conclusions are drawn in Section 5.

2. Related work

Kernel function is successfully used on SVM, KPCA, and KFA which are nonlinear extensions to the linear methods. Except for directly using the classic kernel functions such as Gaussian, polynomial, and RBF kernels, many researchers focus on designing new types of kernels for improved performances in various applications. For example, inspired by similarity measure, histogram intersection (HI) [17] and Gaussian weighted chi-square kernel (GWchi) [18] are designed as new kernel functions for vision applications. For example, inspired by similarity measure, histogram intersection (HI) [17] and Gaussian weighted chi-square kernel (GWchi) [18] are designed as new kernel functions for vision applications. Therefore, we have

\[ K\text{GWchi}(\mathbf{x}_1, \mathbf{x}_2) = \exp(-r^2 \text{GWchi}(\mathbf{x}_1, \mathbf{x}_2)) \]

where \( \text{GWchi}(\mathbf{x}_1, \mathbf{x}_2) \) is the Chi-square statistic, and \( r \) is a constant. \( K\text{GWchi}(\mathbf{x}_1, \mathbf{x}_2) \) had been proved to be positively definite in [17, 18]. Eq. (4) is equivalent to performing inner product of two local features in the high-dimensional space. It is obvious that LKF is a semi-positive-definite matrix, which is also a Hermite matrix as can be seen from its definition in Eq. (3). Compared to KPCA and KFA, we know that LKFA merely extracts features from itself; therefore, avoiding the storage problem as no training sample needs to be saved.

The LKF matrix is a generic nonlinear feature; however, it is symmetric which contains redundant information. To obtain a compact and robust feature representation from LKF, we calculate its eigenvalue vector, as the final extracted feature which is proven below to be robust to a small degree of noise such as the Gaussian noise:

Let \( A \) be a LKF matrix defined in the above section. \( E \) be a symmetric noise matrix, and \( A + E \) be the LKF matrix corrupted with noise. It should be noted that \( E \) is symmetric because the noise added on a sub-region feature is used to calculate LKF resulting in symmetric matrix as shown in Eq. (3). Therefore, \( A, E, \) and \( A + E \) are all Hermite matrices. Their eigenvalue vectors of \( A, E, \) and \( A + E \) are \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n, \quad 0 \geq \varepsilon_2 \geq \cdots \geq \varepsilon_n, \) and \( 0 \geq \mu_1 \geq \mu_2 \geq \cdots \geq \mu_n \), respectively.

For the Hermite matrix \( H \), we can have the Rayleigh quotient as

\[ R_H(x) = \frac{x^T H x}{x^T x} \]

Then we can define the eigenvalue \( \lambda_k \) as follows:

\[ \lambda_k = \max_{c_k} \min_{x \in c_k, x \neq 0} R_H(x) \]

where \( c_k \) can be any \( k \)-dimensional subspace. Thus we have

\[ \mu_k = \max_{c_k} \min_{x \in c_k, x \neq 0} R_{A + E}(x) = \max_{c_k} \min_{x \in c_k, x \neq 0} \left( \frac{x^T A x}{x^T x} + \frac{x^T E x}{x^T x} \right) \]

As the eigenvalue is in the decreasing order, we have

\[ \varepsilon_1 \geq \mu_1 \geq \cdots \geq \mu_n \]

From Eqs. (8), (9), and 11, we have

\[ \mu_k \leq \lambda_k - \varepsilon_k \]

We also know that

\[ A = (A + E) - E \]

where \( A + E \) and \( E \) are both Hermite matrices, we can get

\[ \lambda_k = \mu_k - \varepsilon_k \]

Therefore, we have

\[ \lambda_k \geq \lambda_k - \varepsilon_k \]
From Eqs. (12) (15), we can know that
\[ \lambda_k - \varepsilon_0 \leq \mu_k \leq \lambda_k + \varepsilon_1 \]  
(16)

The trace of $E^\top E$ is defined as
\[ \text{tr}(E^\top E) = \sum_{i=1}^{n} c_i^2 \]  
(17)
where $\varepsilon < \text{tr}(E^\top E)$, and $\text{tr}(E^\top E)$ is in general small; therefore, the eigenvalue is robust to noise.

We choose a sample (see Fig. 1a) from the FRGC database to show how LKFA can eliminate the noise. The image shown in Fig. 1c is obtained by adding the Gaussian noise with 0.0001 and 20/255 (255 x 255) as mean and variance on the original image. In LKFA, the first eigenvalue feature is extracted and rescaled to a 256 grey-scale image (see Fig. 1c) for visualization purpose. The signal to noise ratio (SNR) is $1.4151$ for Fig. 1a and $1.3179$ for Fig. 1c, which show that LKFA is effective on eliminating the noise.

In the classification procedure, assume $v^1$ and $v^2$, as the eigenvalue vectors generated from two LKFA matrices of images $P_1$ and $P_2$, the cosine similarity rule is used to calculate the similarity as
\[
\cos\theta = \frac{v^1 \cdot v^2}{\|v^1\| \|v^2\|} 
\]  
(18)

For a complex object recognition problem such as face recognition, it is always beneficial to further exploit the subspace method, i.e. Fisher Linear Analysis, to extract the discriminant features from LKFA. If $W^f$ is the transformation matrix calculated from the FLA method, we can get the extracted features as $W^f v^1, W^f v^2$, where $v^1$ and $v^2$ are calculated by the proposed LKFA method. In the classification procedure, we use the same cosine similarity rule as shown in Eq. 18.

4. Experiments

In the experiment, we exploit Gabor wavelet to enhance the performance of object recognition. The Gabor wavelets (kernels, filters) are defined as follows [13,20,21]:
\[
\psi_{u,v}(z) = \frac{\left| k_{u,v} \right|^2}{\sigma^2} e^{-\left| k_{u,v} \right|^2 / 2\sigma^2} e^{-a \sigma^2 / 2} 
\]  
(19)
where $z = \left( \begin{array}{c} x \\ y \end{array} \right)$, $k_{u,v} = \left( \begin{array}{c} k_{x} \\ k_{y} \end{array} \right) = \left( \begin{array}{c} k_u \cos \phi_u \\ k_v \sin \phi_u \end{array} \right)$, $k_v = \pi / 2^{\nu / 2}$, $\phi_u = u \pi / 2$, $\nu = 0, \ldots, v_{\text{max}} - 1$, $u = 0, \ldots, u_{\text{max}} - 1$, $\nu$ is the frequency, $u$ is the orientation, with $v_{\text{max}} = 5, u_{\text{max}} = 8$, and $\sigma = 2\pi$ for a given pixel in the feature space, around which we have 8 neighbors. For example, the 8 neighbors of $Z_0$ are $Z_1$ to $Z_8$ as shown in Fig. 2(a).

Therefore, the size of the LKFA matrix at the position $Z_0$ is $9 \times 9$. In the experiment, the final chosen number of the eigenvalues is $9 - 1 = 8$. However, for the marginal position like $Z_1, Z_3, Z_5, Z_7$, the size of LKFA matrix is $4 \times 4$, and then the size of the final feature is $4 - 1 = 3$. The size of the LKFA matrix for other marginal pixels as $Z_2, Z_4, Z_6, Z_8$ is $6 \times 6$, and then the final extracted feature size is 5. It should be noted that LKFA is calculated from Gabor magnitude.

4.1. Experiment on Polyu Palmprint database

We first do experiment on the Polyu Palmprint database [22], which is captured under different illumination conditions with deformation. The database has 600 images from 100 people, with 6 images per person. The gallery database contains 100 images with 1 image per person, and the remaining 500 images are used as the probe database. In this experiment, the images in the database are normalized into the size of $128 \times 128$. We first test the proposed approach on grey-level images, while the baseline algorithm is the nearest neighborhood classifier with the cosine similarity as the distance measure, which achieve 75.8% recognition rate. The proposed method is also based on the original image with each pixel grey-value as the local feature as shown in Fig. 3. In this
situation, the dimension of the LKF matrix for each pixel is $9 \times 9$, from which we just calculate a 1D eigenvalue vector as the final feature. We achieve 77.7% and 81% recognition rates with polynomial kernel (parameter is 2) and Gaussian kernel, respectively. The performance of LKFA with Gaussian kernel is the best, which shows that the nonlinear Local Kernel Feature is more effective for Palmprint recognition.

We perform the experiment based on the Gabor magnitude LBP feature, the proposed LKFA method using LBP as the local feature achieves a better performance than LBP when the number of histogram bins is changed from 8 to 256 as shown in Fig. 3. LKFA with Gaussian kernel achieves the best result, which shows that the nonlinear relationship among local regions is useful for object recognition. LBP performs better when using a large number of histogram bins. It is interesting to note that even using a smaller number of histogram bins (e.g., 8 bins), LKFA still outperformed LBP with much larger number of histogram bins (e.g., 256 bins).

4.2. Comparisons based on the FRGC Version 1

To verify the performance of the proposed method on face recognition, we conducted experiments on the well-known FRGC database using Experiment #4 protocol. Some samples are given as in Fig. 5. As shown in [23], the Experiment #4 is designed for indoor controlled still images versus uncontrolled still images, which is the most challenging FRGC experiment. In the FRGC Version 1, the training set contains 366 images, the target (gallery) set contains 943 controlled images, and the query (probe) set has 943 uncontrolled images. For Experiment #4 of the FRGC Version 2, the training set contains 12,776 images, the target set includes 16,028 controlled images, and the query set has 8014 uncontrolled images [23]. In both experiments, face images are normalized and cropped to the size of $128 \times 144$ images, which are divided into $16 \times 24$-sized sub-regions. Then we construct an ensemble of $8 \times 6$ classifiers based on the sum rule, and each one of which is applied on one sub-region with the size of $16 \times 24$. To further reserve more spatial information, we divide each sub-region (classifier) into smaller regions with the size of $4 \times 6$, as shown in Fig. 2(b). In each region, we extract LBP histogram feature from the Gabor magnitude for LKFA, and the bin number is set as 8.

The experiment is first conducted on Experiment #4 of the FRGC Version 1 by comparing LBP+Fisher, LBP+KFA, and LKFA+Fisher. In this experiment, the mean sample for each class is calculated in the target set. It should be noted that the LBP+KFA method directly uses the LBP feature with the GWChi kernel. As shown in Fig. 4, the LKFA+Fisher method achieved a much better result than LBP+Fisher which confirms that the proposed method is effective toward high performance in the real application, as the FRGC database is designed for both controlled and uncontrolled conditions. Compared to the Efficient Kernel Fisher method [24], the proposed method achieved a higher performance as shown in Table 1.

4.3. Comparisons based on the FRGC Version 2

In this section, Experiment #4 on the FRGC Version 2 is used to evaluate the performance of the face recognition system for face verification. Besides the LBP+Fisher method, we also compared our results with reported results of some state-of-the-art methods. From Table 2, it can be seen that the proposed method achieved a better performance than the original LBP based method, because LKFA can use the histogram similarity measure and capture the nonlinear structure of the input object. The LBP+Fisher method cannot effectively use the histogram measure, which may cause the decrease of the recognition performance [25]. The proposed method also achieves a much better result than some state-of-the-art methods, which shows that LKFA is also an effective way to enhance the face recognition performance.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Recognition rates with FAR=0.1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline [23]</td>
<td>0.12</td>
</tr>
<tr>
<td>Result in [15]</td>
<td>0.76</td>
</tr>
<tr>
<td>Result in [26]</td>
<td>0.784</td>
</tr>
<tr>
<td>Result in [24]</td>
<td>0.743</td>
</tr>
<tr>
<td>LBP+Fisher</td>
<td>0.78</td>
</tr>
<tr>
<td>LKFA+Fisher</td>
<td>0.83</td>
</tr>
</tbody>
</table>

Fig. 4. Comparative results on the Palmprint database.

Fig. 5. Samples of FRGC database.
5. Conclusions and future work

This paper proposes a new method, named Local Kernel Feature Analysis, for object recognition. Different from the traditional kernel-based method using kernel functions to calculate the inner product of different input samples in the high-dimensionality space, this paper exploits it for the feature extraction from a single input image to capture the relationship among local neighborhood. We also prove in theory that the eigenvalue vector feature is stable, which is extracted as the final feature for object representation. Comparative experiments on both Palmprint and face recognition show that the proposed method achieves a better performance than the original LBP-based method.

The kernel function affects the final performance of object recognition, we will focus on this topic to design new kernel for a better performance. The future work will also focus on the application of the proposed method on other object recognition.

Acknowledgement

The work was supported by the Natural Science Foundation of China (NSFC) under Contract No. No.60903065, and No.61039003. This work is also supported by a grant from the Ph.D. Programs Foundation of Ministry of Education of China (No. 20091102120001) and the Fundamental Research Funds for the Central Universities.

References