Non-Linear Principal Component Embedding for Face Recognition

Eimad Eldin Abd Al Ali Abusham and Wong Eng Kiong
Faculty of Information Science and Technology, Multimedia University, Melaka Campus,
Jalan Ayer Keroh Lama, 75450 Bukit Beruang, Melaka, Malaysia

Abstract: A new face recognition method, based on the local non-linear mapping, is proposed in this study.
Face images are typically acquired in frontal views and often illuminated by a frontal light source.
Unfortunately, recognition performance is found to significantly degrade when the face recognition systems
are presented with patterns that go beyond these controlled conditions. Face images acquired under
uncontrolled conditions have been proven to be highly complex and are non-linear in nature; thus, the linear
methods fail to capture the non-linear nature of the variations. The proposed method in this study is known
as the Non-linear Principal Component Embedding (NPCE) which is aimed to solve the limitation of both linear
and non-linear methods by extracting discriminant linear features from highly non-linear features; the method
can be viewed as a linear approximation which preserves the local configurations of the nearest neighbours.
The NPCE automatically learns the local neighbourhood characteristic and discovers the compact linear
subspace which optimally preserves the intrinsic manifold structure; a principal component is then carried out
onto low dimensional embedding with reference to the variance of the data. To validate the proposed method,
Carnegie Mellon University Pose, Illumination and Expression (CMU-Pose) database was used. Experiments
conducted in this research revealed the efficiency of the proposed method in face recognition as follows: (1)
extract discriminant linear features from highly non-linear features based on the local mapping and (2) Runtime
speed is improved as face feature values are reduced in the embedding space. The proposed method achieves
a better recognition performance in the comparison with both the linear and non-linear methods.

Key words: Face recognition, biometrics, manifold learning, principal component, feature extraction

INTRODUCTION

The state-of-the-art face recognition systems are found to yield satisfactory performance under controlled
conditions, i.e., where face images are typically acquired in frontal views and often illuminated by the frontal light
source. Unfortunately, recognition performance is found to significantly degrade when the face recognition
systems are presented with patterns which go beyond these controlled conditions. Some examples of
unconstrained conditions include illumination, pose variations, etc. In particular, variations in face images
have been proven to be highly complex and non-linear in nature. Linear subspace analysis has been extensively
applied to face recognition. A successful face recognition methodology is largely dependent on particular choice
of features used by the classifier. Although, linear methods are easy to understand and are very simple to
implement, the linearity assumption does not hold in many real-world scenarios. A disadvantage of the linear
techniques is that they fail to capture the characteristics of the non-linear appearance manifold. This is due to the
fact that the linear methods extract features only from the input space without considering the non-linear
information between the components of the input data. However, non-linear mapping can often be approximated
using a linear mapping in a local region. This has motivated the design of the non-linear mapping methods
in this study. The history of the non-linear mapping is long; it can be traced back to non-linear mapping
(Sammon, 1969). Over time, different techniques have been proposed such as the projection pursuit (Friedman
and Tukey, 1974), the projection pursuit regression (Friedman and Stuetzle, 1981), self-organizing maps
or SOM, principal curve and its extensions (Hastie and Stuetzle, 1989; Kegl et al., 2000; Smola et al., 2001;
Tibshirani, 1992), auto-encoder neural networks (Baldi and Hornik, 1989; DeMers and Cottrell, 1993) and
generative topographic maps or GTM (Bishop et al., 1998). A comparison of some of these methods can be
found by Mao and Jain (1995). Recently, a new line of the non-linear mapping algorithms was proposed based on
the notion of manifold learning. Given a data set which is assumed to be lying approximately on the manifold in a
high dimensional space, dimensionality reduction can be achieved by constructing a mapping which respects

Corresponding Author: Eimad Eldin Abd Al Ali Abusham, Faculty of Information Science and Technology, Multimedia University,
Melaka Campus, Jalan Ayer Keroh Lama, 75450 Bukit Beruang, Melaka, Malaysia

2625
certain properties of the manifold. Manifold learning has been demonstrated in different applications including face pose detection (Hadid et al., 2002; Li et al., 2001), high dimensional data discriminant analysis (Bouveyron et al., 2007), face recognition (Yang, 2002; Zhang and Wang, 2004), analysis of facial expressions (Chang et al., 2004; Elgammal and Lee, 2004), human motion data interpretation (Jenkins and Mataric, 2004), gait analysis (Elgammal and Lee, 2004a, b), visualization of fibre traces (Brun et al., 2003), wood texture analysis (Niskanen and Silvn, 2003) and kernel fractional l-step discriminant analysis (KFDA) for the non-linear feature extraction and dimensionality reduction by Guang et al. (2006). Recently, Li et al. (2008) proposed the non-linear DCT discriminant feature which analyzes the non-linear discriminabilities of the DCT frequency bands and selects appropriate bands. Nevertheless, these methods still lack discriminant features representation, based on the local structure of data which is very important for recognition when variations of face images are present. Therefore, the aim of this study to devise local non-linear discriminant feature representations which are reliable and have more discriminative power face recognition.

MATERIALS AND METHODS

Pre-processing: Face pre-processing and normalization is a significant part of the face recognition systems. Changes in lighting conditions have been found to dramatically decrease the performance of face recognition. Therefore, all images have been pre-processed to obtain a representation of the face which is invariant to illumination, while keeping the information necessary to allow a discriminative recognition of the subjects. Gaussian kernel has been used to estimate the local mean and standard deviation of images to correct non-uniform illumination. The local normalization is computed as follows:

\[
g(x, y) = \frac{f(x, y) - m_f(x, y)}{\sigma_f(x, y)}
\]

where, \(f(x, y)\) is the original image, while \(m\) is an estimation of the local mean of \(f\) and \(\sigma\) is an estimation of the local SD. Figure 1 below illustrates the block diagram of the developed method.

The NPCE algorithm: This method finds reconstruction weight by capturing the intrinsic geometry of the neighbourhood. The NPCE creates a locally linear mapping from the high dimensional coordinates to the low dimensional embedding as shown in Fig. 2.

Compute the average weight which represents every face data by its neighbours.

\[
\phi(w) = \sum_{i} w_i x_i || x_i \in X \in \mathbb{R}^n
\]

where, \(x_i\) refers the ith unknown sample and \(x_i\) is the corresponding training sample according to the K-values (the nearest neighbours).

Compute the low-dimensional embedding \(D\), the following cost function is minimized:

\[
\phi(D) = \sum_{i} || D_i - \sum_{j} w_j D_j ||^2
\]

where, \(N\) is the number of training and \(K\) is the number of the nearest neighbours.

Then, the principal component of the training is calculated as follows:

\[
\Phi = x - \bar{x}
\]

\[
C = \frac{1}{M} \sum_{i=1}^{M} \Phi_i \Phi_i^T
\]

Fig. 1: Block diagram of the NPCE
where, $\bar{x}$ is the mean and $C$ is the covariance matrix; 
\{P_1, P_2, ..., P_d\} are the eigenvectors of $C$. The eigenvectors 
then play a role which projects a vector in the low-
dimensional face subspace into discriminatory feature 
space that can be formulated as follows:

$$Q = D^T P = Q e R^T, \; D e R^d$$ (5)

Once, the weighted values of each neighbour sample of 
the unknown sample are obtained, the mapping formulate 
can be seen as follows:

$$q_i = \sum_{i \in N} W_{ij} q_i$$ (6)

where, $q_i$ is the closely training sample and the neighbour 
indices are the same as that of the sample in the original 
high dimensional space and $q_i$ is the corresponding one 
of the unknown samples in the discriminant space.

RESULTS AND DISCUSSION

CMU-PIE database: This is one of the largest datasets 
developed to investigate the affect of pose, illumination 
and expression. It contains images of 68 people, each 
under 13 different poses, 43 different illumination 
conditions and 4 different expressions (Sim et al., 
2002). In the experiments conducted in this study, 6 out 
of 13 poses were selected for each person. Out of 43 
ilumination configurations, 21 were selected to typically 
span the set of variations and these covered the left to the 
right profile.

Non-linear Principal Component Embedding (NPCE): In 
this set of experiment, the Principal Component Analysis 
(PCA) and Linear Discriminant Analysis (LDA) are two 
powerful tools used for dimensionality reduction and feature 
extraction in most of pattern recognition applications; these tools were used to assess the 
efficiency of the method proposed in this study. Figure 3 
shows that the dimensions used for testing the NPCE 
range (between 10 and 150) and the proposed method was 
found to significantly outperform the PCA and LDA. 
More experiments were conducted on the reduced 
selected dimensions (65, 75, 90 and 110) to assess 
the performance of the NPCE. For this, good recognition rates 
were obtained; the recognition rates of 49.66, 55.9 and 
82.87% were obtained by the PCA, LDA and NPCE, with 
feature dimensions of 110, 67 and 100, respectively. As 
for the LDA, the maximum feature dimension cannot be more 
than 67, which is C-1 (number of classes-1).

Table 1 shows the average recognition error rates 
across ten tests and four dimensions (65, 75, 90 and 110). 
From these results, the NPCE was found to achieve the 
lowest error rate, as compared to the standard linear 
methods of PCA and LDA.

Figure 4 shows the results when the NPCE was used, 
as compared to the KPCA and LDA (Jian et al., 2005), as 
well as the Generalized Discriminant Analysis (GDA) 
(Baudat and Anouar, 2000). The method was shown to 
achieve 82.87% accuracy and had significantly 
outperformed the KPCA plus LDA and GDA; the later 
methods achieved the maximum accuracy of 77.22 and 
79.92%, respectively.

The proposed method was developed to learn 
embedding in the non-linear manifold based on the k-nearest neighbour method and preserve the local 
geometry of the original high-dimensional data in a low-
dimensional space as good as possible. In addition to 
these, the NPCE was found to minimize the reconstruction 
error of the neighbour weights for every data point in the 
low-dimensional space. The training sets are projected 
into the intrinsic low-dimensional space to improve their 
classification ability and runtime speed, while the principal 
components are projected into the low-dimensional 
embedding, with reference to the variance of the data, as 
given in Eq. 5. As a result, the maximum feature dimension

---

Table 1: The average error rates (%) of the PCA, LDA, and NPCE across ten tests and four dimensions

<table>
<thead>
<tr>
<th>Methods</th>
<th>Average error rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>69.75</td>
</tr>
<tr>
<td>LDA</td>
<td>48.15</td>
</tr>
<tr>
<td>NPCE</td>
<td>17.13</td>
</tr>
</tbody>
</table>
Fig. 4: The recognition rate of the KPCA plus LDA, GDA and NPCE

can be more than C-1 (number of classes-1). Therefore, this is considered as a solution to Small Sample Size (SSS) problem, where the size of sample is always smaller than the dimension of sample. In addition, the performance of the proposed method is compared with several different state-of-the-art non-linear methods. Based on the results presented in Table 1 and Fig. 4, the feature representations are proven to have more discriminative power, while the NPCE achieves a better recognition performance as compared to the linear and non-linear methods.

CONCLUSION

A new Non-linear Principal Component Embedding (NPCE) for face recognition has been introduced and proposed in this research. The proposed method is based on the local non-linear discriminant representation, which is particularly robust against the SSS problem as compared to the traditional one used in LDA. NPCE utilize a novel discriminant principal component to estimate the face feature values in the reduced embedding space. At the same time, the proposed methods have been found to perform an implicit reduction over the whole set of features, as shown by the results derived from the experiments. Therefore, the researchers regard this as significant due to the fact that the runtime speed is as important as the actual recognition rate, i.e., if only a subset of the features is used. The experiments conducted in this study clearly reveal that the proposed method is superior to the state-of-the-art methods. Thus, the future study will concentrate on achieving continuous improvement for the devised method and extending it so as to incorporate more local features of the subjects.

ACKNOWLEDGMENT

This study is conducted as a part of my Ph.D at Multimedia University from 2004 to 2007. The author is thankful to the Multimedia University for providing the research facilities and financial support which enable us to carry out this research successfully.

REFERENCES


