Abstract — This paper introduces the establishment of PolyU Near-Infrared Face Database (PolyU-NIRFD) for face recognition. The PolyU-NIRFD contains images from 350 subjects, each contributing about 100 samples with variations of pose, expression, focus, scale and time, etc. In total, 35,000 samples were collected in the database. The PolyU-NIRFD provides a platform for researchers to develop and evaluate various near-infrared face recognition techniques under large scale, controlled and uncontrolled conditions. Finally, we provide three protocols to evaluate the baseline face recognition methods, including Gabor based Eigenface, Fisherface and LBP (local binary pattern) on the PolyU-NIRFD database.

Index Terms — Near-infrared face recognition, face database, feature extraction
1. INTRODUCTION

Face recognition (FR) is a promising technology for automated personal authentication and it has a great potential in applications of public security, video surveillance, access control and forensics, etc. [R. Chellappa et al. 1995, W. Zhao et al. 2003]. Meanwhile, FR is one of the most active topics in the field of computer vision, and several large-scale face databases [W. Gao et al. 2008, Enrique et al. 2003, T. Sim et al. 2003, K. Messer, et al. 1999, A. M. Martinez and R. Benavente 1998, A. S. Georghiades et al. 2001, K. C. Lee et al. 2005] are publicly available for evaluating and comparing various FR methods. Generally speaking, FR in visible spectrum has been mostly studied because it is convenient to implement in various environment and has a wide range of applications [R. Chellappa et al. 1995, W. Zhao et al. 2003, Turk and Pentland. 1991, Belhumeur et al. 1997]. Many FR algorithms have been proposed [R. Chellappa et al. 1995, W. Zhao et al. 2003], and the large-scale face databases play an important role in evaluating and developing FR algorithms.

Face recognition technology (FERET) [P.J. Phillips et al. 2000], face recognition vendor test (FRVT) [P. J. Phillips et al. 2002], and face recognition grand challenge (FRGC) [P. J. Phillips et al. 2005] have pioneered both evaluation protocols and database construction. FRGC is more challenging than FERET and FRVT, as it contains more uncontrolled variations and 3D images in its database. For example, in the most challenging set of the FRGC v2 (Exp#4), the training set contains 10776 images from 222 subjects, while the query and target sets contain 8,014 and 16,028 images respectively. The 3D training set of FRGC consists of controlled and uncontrolled still images from 943 subject sessions, while the validation partition of FRGC contains images from 466 subjects collected in 4007 subject sessions. Other publicly available face databases include the CAS-PYAL [W. Gao et al. 2008], BANCA [Enrique et al. 2003], CMU PIE [T. Sim et al. 2003], XM2VTSDB [K. Messer, et al. 1999], AR [A. M. Martinez and R. Benavente 1998], Yale [Georghiades et al. 2001, Lee et al. 2005], etc. These face databases in visible spectrum provide a good evaluation platform for various FR techniques, and in return they greatly facilitate the development of new FR methods.

In visual face recognition, the performance suffers from the lighting variations. To solve this problem,
traditional methods are mostly based on Lambertian model [T. Sim et al. 2003], which is too simply to describe
the real face surface under various illuminations. Recently, active near-infrared (NIR) FR was proposed to deal
with the illumination variations in different environments, and NIR based FR has shown promising performance
in real application, which uses different imaging sensors in invisible spectral bands to reduce the affection of
ambient light. X. Zou et al. proposed to use active NIR light to localize face areas in the images and then
recognize faces [X. Zou et al. 2005]. S.Z. Li et al. extracted the Local Binary Pattern (LBP) feature and used
Fisher analysis for NIR based FR, and they developed a complete NIR FR system which can perform face
detection, eye localization and face identification [Stan Li et al. 2007]. Pan et al. [Z.H. Pan et al. 2003] proposed
an NIR-based FR system which captures face images in wavelength of 0.7um-1.0um. Some other works using
NIR images for FR can be found in [J. Dowdall et al. 2003, A. S. Georgiades et al. 2001]. Zou et al. [X. Zou et
al. 2005] have shown that FR in NIR band has better performance than that in the visible band, and this was
also validated in S.Z. Li’s work [Stan Li et al. 2007]. However, so far there is not a large scale NIR face
database which is publicly available. There is a high demand to construct an open NIR FR database, on which
the researchers can test and compare their algorithms. In this paper, we will introduce such a database we
constructed in the Hong Kong Polytechnic University, and name the database as PolyU Near-infrared Face
Database (PolyU-NIRFD).

The face images in PolyU-NIRFD were collected from 350 subjects, each subject providing about 100
samples. The sample images involve various variations of expression, pose, scale, focus and time, etc. To
evaluate the performance of different FR methods on the PolyU-NIRFD, we provide three test protocols,
including the partition strategy of the training, gallery and probe sets and the baseline evaluation schemes. The
baseline algorithms we used for comparison are Eigenface, Fisherface, Local Binary Pattern (LBP) [T. Ojala et
al. 2002] and their Gabor filtering enhanced versions, which are well-known and representative methods in the
field of FR. Considering that the Gabor filtering can improve significantly the FR accuracy (e.g. Gabor
Fisherface and Gabor LBP are among the best FR algorithms), we also examine its performance on PolyU-
NIRFD.
The rest of the paper is organized as follows. Section II introduces the developed NIR face acquisition system and the establishment of PolyU-NIRFD. In Section III, the baseline algorithms are briefly described. Section IV presents extensive experiments using three protocols to evaluate the performance of various FR methods, including Gabor-Eigenface, Gabor-Fisherface, Gabor-LBP, on the PolyU-NIRFD. Conclusions are drawn in Section V.

2. POLYU-NIRFD CONSTRUCTION

Different from the FR in visible band, which can simply use a common camera to capture face images, FR in NIR band needs some additional hardware and special system design for image acquisition. This section describes the NIR image acquisition system and the collection of the PolyU-NIRFD database.

2.1. Near-Infrared Face Image Acquisition

The hardware of our NIR face image acquisition system consists of a camera, an LED (light emitting diode) light source, a filter, a frame grabber card and a computer. A snapshot of the constructed imaging system is shown in Fig. 1. The camera used is a JAI camera, which is sensitive to NIR band. The active light source is in the NIR spectrum between 780nm - 1,100 nm and it is mounted on the camera. The peak wavelength is 850nm, and it lies in the invisible and reflective light range of the electromagnetic spectrum. An NIR LED array is used as the active radiation sources, and it is strong enough for indoor use. The LEDs are arranged in a circle and they are mounted on the camera to make the illumination on the face is as homogeneous as possible. The strength of the total LED lighting is adjusted to ensure a good quality of the NIR face images when the camera-face distance is between 80cm-120cm, which is convenient for the users. When mounted on the camera, the LEDs are approximately coaxial to the imaging direction and thus provide the best possible straight frontal lighting. Although NIR is invisible to the naked eyes, many CCD cameras have sufficient response to the NIR spectrum. The filter we used in the device is used to cut off the visible light, whose spectrum is shorter (780nm,
visible light). For the convenience of data collection, we put the imaging device into a black box of 19cm width, 19cm long, and 20cm high, as shown in Figure 1.

Fig. 1: The NIR face image acquisition device. ‘A’ is the NIR LED light source, and ‘B’ is the NIR sensitive camera with a NIR filter.

2.2. PolyU-NIRFD Construction

By using the self-designed data acquisition device described in Section II-A, we collected NIR face images from 350 subjects. During the recording, the subject was first asked to sit in front of the camera, and the normal frontal face images of him/her were collected. Then the subject was asked to make expression and pose changes and the corresponding images were collected. To collect face images with scale variations, we asked the subjects to move near to or away from the camera in a certain range. At last, to collect face images with time variations, samples from 15 subjects were collected at two different times with an interval of more than two months. In each recording, we collected about 100 images from each subject, and in total 35,000 images were collected in the PolyU-NIRFD database. The sample images in the PolyU-NIRFD are labeled as ‘NN_xxxxx_S_D_****’, where “NN” represents the prefix of the label string, ‘S’ represents the Gender information, ‘xxxxxx’ indicates the ID serial number of the subject, ‘D’ denotes the place where the image was captured, and ‘****’ is the index of the face image. For example, “NN_200001_F_B_024” means that the 24th image is from a Female subject collected in Beihang University. Figure 2 shows some face images of a subject
with variations of expression, pose and scale. Figure 3 shows the images of some subjects which were taken in
different times.

![Sample NIR face images of a subject.](image1.png)

(a) Normal face image; and images with (b) expression variation; (c) pose variation and (d) scaling variation.

Fig. 3: Sample NIR face images captured in more than two months.

To evaluate the performance of different methods on the PolyU-NIRFD, we design three types of experiments, each of which contains a training set, a target (Gallery) set and a query (Probe) set. In Exp#1, the
used images include frontal face images as well as images with expression variations, scale changes (include
blurring), and time difference, etc. In this experiment, the sizes of the training set, target set and query set are 419, 574 and 2763, respectively. In Exp#2, we add more faces captured in uncontrolled conditions to make the test more challenging. The sizes of the training set, target set and query set are 1876, 1159, and 4747 respectively. In Exp#3, we focus on the images with high pose variations and exclude the images with expression, scale and time variations. The sizes of the training set, target set and query set are 578, 951, and 3648, respectively. In next section, we will present baseline face recognition algorithm, and then in Section IV the three types of experiments will be conducted by different methods.

3. BASELINE ALGORITHMS

We choose both grey-level image and Gabor based Eigenface, Fisherface, and the LBP method as the baseline algorithms.

3.1 Gabor Wavelet

In image processing and object recognition, Gabor features are widely used for image feature descriptors extracted by a set of Gabor wavelets (kernels) which model the receptive field profiles of cortical simple cells. They can capture salient visual properties in an image, such as spatial characteristics, because the kernels can selectively enhance features in certain scales and orientations. The Gabor wavelets (kernels, filters) can be defined as follows:

\[ \psi_{u,v}(z) = \frac{\|k_{u,v}\|^2}{\sigma^2} e^{-\|k_{u,v}\|\|z\|^2/2\sigma^2} \left[ e^{k_{u,v}z} - e^{-\sigma^2/2} \right], \]  

(1)

where \( z = \begin{pmatrix} x \\ y \end{pmatrix} \), \( k_{u,v} = \begin{pmatrix} k_x \\ k_y \end{pmatrix} = \begin{pmatrix} \cos \phi_u \\ \sin \phi_u \end{pmatrix} \), \( k_x = \pi / 2^{v/2} \), \( \phi_u = u \pi / 8 \), \( v = 0, ..., v_{\max} - 1 \), \( u = 0, ..., u_{\max} - 1 \), \( v \) is the frequency, \( u \) is the orientation, \( v_{\max} = 5 \), \( u_{\max} = 8 \) and \( \sigma = 2\pi \). In this paper, only magnitude part of the Gabor feature is used to enhance the performance of Eigenface, Fisherface, and LBP.
3.2 Baseline Algorithm

Principal Components Analysis (PCA) is commonly used for dimensionality reduction in signal processing, also known as Eigenface in face recognition. Eigenface serves as the baseline method due to its easy implementation and reasonable performance. PCA chooses projection directions that maximize the total scatter across all images of all faces in the training set, and the scatter matrix is calculated as:

$$ S = \sum_{i=1}^{C} p(x_i)(x_i - \bar{x})(x_i - \bar{x})^T $$

(2)

Linear Discriminant Analysis (LDA) is a widely used method for feature extraction and dimensionality reduction in face recognition, which is the core part of the well-known Fisherface method. LDA tries to find the discriminative project direction in which training samples belonging to different classes are separated. Mathematically, it calculates the projection matrix in a way that the ratio of the determinant of the between-class scatter matrix of the projected samples and the within-class scatter matrix of the projected samples is maximized. The between-class scatter matrix $S_b$ and within-class scatter matrix $S_w$ are defined as follows:

$$ S_b = \sum_{i=1}^{C} p(\sigma_i)(u_i - u)(u_i - u)^T, $$

(3)

$$ S_w = \sum_{i=1}^{C} p(\sigma_i)E\{(x_i - u_i)(x_i - u_i)^T|\sigma_i\}, $$

(4)

where $u_i = (1/n)\sum_{j=1}^{n} x_{ij}$ denotes the sample mean of class $i$, $u$ is the mean of all training images, and $p(\sigma_i)$ is the prior probability.

In this paper, we also test the LBP based face recognition on the PolyU-NIRFD database. Derived from a general definition of texture in a local neighborhood, LBP is defined as a gray-scale invariant texture measure and is a useful tool to model texture images. LBP later has shown excellent performance in many comparative studies, in terms of both speed and discrimination performance. The original LBP operator labels the pixels of an image by thresholding the $3\times3$ neighborhood of each pixel with the value of the central pixel and concatenating the results to form a number. The thresholding function $f()$ for the basic LBP can be formally
represented as:

\[
f(I(Z_i), I(Z_0)) = \begin{cases} 
  0, & \text{if } I(Z_i) - I(Z_0) \leq \text{threshold} \\
  1, & \text{if } I(Z_i) - I(Z_0) > \text{threshold}
\end{cases}
\]  

(5)

where \( Z_i, i = 1, \ldots, 8 \) is an 8-neighborhood point around \( Z_0 \) as shown in the central part in Fig. 4. A LBP can also be considered as the concatenation of the binary gradient directions, and is called a micro-pattern. Fig. 5 shows an example of obtaining a LBP micro-pattern when the threshold is set to zero. The histograms of these micro-patterns contain information of the distribution of the edges, spots, and other local features in an image.

![Fig. 4. An example of 8-neighborhood around \( Z_0 \).](image)

![Fig. 5. An example of obtaining the LBP micro-pattern for the region in the black square.](image)

### 4. EXPERIMENTS

The main objectives of the experiments in this section are to evaluate the performance of well-known face recognition algorithms on the PolyU-NIRDF and investigate the strengths and weaknesses of the proposed method and baseline algorithms. We choose the Eigenface (i.e. PCA) [Turk and Pentland. 1991], Fisherface (i.e. LDA) [Belhumeur et al. 1997], LBP [T. Ahonen et al. 2006], Gabor-PCA, Gabor-LDA and Gabor-LBP as the baseline algorithms. The Gabor-LDA and Gabor-LBP are among the best FR methods and they are benchmarks to evaluate the performance of FR techniques.
4.1 Experiment 1

In Exp#1, we set up a subset from the whole PolyU-NIRFD database. In this subset, the training set contains 419 frontal images from 138 subjects, while the gallery set and probe set have 574 and 2763 images respectively. No images in the probe and gallery sets are contained in the training set. The facial portion of each original image is automatically cropped according to the location of the eyes. The cropped face is then normalized to $64 \times 64$ pixels. The eight methods are then applied to this dataset to evaluate their performance.

For subspace based methods, the distance used in the nearest neighbor classifier is the cosine of the angle between two feature vectors. For LBP histogram features, we use the histogram intersection similarity measure. The sub-region size and the number of histogram bins for LBP are $8 \times 8$ and 32. Since the Gabor filtering is performed on five scales and eight orientations, we need to downsample the response image to reduce data amount. The downsampling factor is set to 4. Therefore, for Gabor-PCA and Gabor-LDA, the input signal size is $64 \times 64 \times 5 \times 8 \div 4^2 = 10240$.

The FR results by the eight methods are illustrated in Figures 7 and 8. For subspace based methods PCA, LDA, Gabor-PCA and Gabor-LDA, the curves of recognition rate versus feature vector dimensionality are plotted in Figure 7. We see that the curves for PCA and LDA are flat when the dimension of feature vector changes from 60 to 120. In this experiment PCA gets similar performance to LDA because the number of training samples for each class is limited. From Figures 7 and 8 we can clearly see that using Gabor features can improve greatly the performance of all the four methods. For example, the recognition rate of Gabor-LDA is $10\%$ higher than that of LDA. Comparing Figure 7 with Figure 8, it can be seen that the LBP method achieve better performance than PCA and LDA.
4.2 Experiment 2

In Exp#2, we extracted from the whole database a much bigger subset than in Exp#1. In this subset, the training set contains 1876 frontal images of 150 subjects, while the gallery and probe sets have 1159 and 4747 images, respectively. The recognition results of PCA, LDA, and LBP are illustrated in Figure 9 and Figure 10. Different from that in Exp#1, LDA achieves much better performance than PCA when using a small number of features. Similar to those in Exp#1, Gabor based methods get much better performances than their original image based
counterparts. We can also observe that the best result of Gabor-LDA is 92.1%, which is not as good as in exp#1. This is because the dataset in Exp#2 is more challenging by involving more variations of pose, expression and scale than that in Exp#1. We can see that LBP has much better performance than PCA and LDA, which validates again that LBP is effective way to model infrared face images.

![Fig. 9: Recognition results by PCA, LDA, Gabor-PCA and Gabor-LDA in Exp#2.](image)

![Fig. 10: Recognition results by LBP, Gabor-LBP in Exp#2.](image)
4.3 Experiment 3

Exp#3 is designed to evaluate the performance of the algorithms on the large variations of pose, expression, illumination and scale, etc. In this subset, training set contains 578 images from 188 subjects, while gallery and probe sets have 951 and 3648 images individually. The results are shown in Figures 11 and 12.

Fig. 11: Recognition results by PCA, LDA, Gabor-PCA and Gabor-LDA in Exp#3.

Fig. 12: Recognition results by LBP, Gabor-LBP in Exp#3.
5. CONCLUSIONS

We introduced in this paper the establishment of PolyU near infrared face database (PolyU-NIRFD), which is one of the largest NIR face databases so far. The main characteristics of the PolyU-NIRFD lie in two aspects. First is its large scale. It consists of 35,000 images from 350 subjects. Second is its diversity of variations. It includes variations of pose, expression, illumination, scale, blurring and the combination of them. Comparative study of baseline algorithms was performed on the PolyU-NIRFD to verify its effectiveness. In the future we will acquire a larger database under both controlled and uncontrolled environment, and perform more experiments to investigate more effective NIR FR algorithms.

6. OBTAINING THE POLYU-NIRFD

The PolyU-NIRFD will be publically available soon. The information about how to obtain a copy of the database will be found on the website (http://www.comp.polyu.edu.hk/~biometrics/polyudb_face.htm).

ACKNOWLEDGMENT

The work is supported by the GRF grant of HKSAR, the central fund from Hong Kong Polytechnic University, the Natural Science Foundation of China (NSFC) under Contract No. 60620160097, and the National High-Tech Research and Development Plan of China (863) under Contract No. 2006AA01Z193 and 2007AA01Z195.

REFERENCES


