Face Recognition from Low Resolution to High Resolution

Prof. Kenneth K.M. Lam
Joey Pong, Dong Li, Huiling Zhou
Centre for Signal Processing
Department of Electronic & Information Engineering
The Hong Kong Polytechnic University
Email: enkmlam@polyu.edu.hk

Innovative Technology Series

Outline

- Introduction
- Low-resolution face recognition
  - Gabor features
  - Feature super-resolution
  - Canonical Correlation Analysis (CCA)
- Standard-resolution face recognition
  - Multi-resolution feature-fusion (MFF)
- High-resolution face recognition
  - Pore-scale features
- Conclusions
Introduction

- Face images captured may have a wide range of resolutions: from low resolution to high resolution.
- For face recognition, image features are first extracted and then matched to those features in a gallery set.
- The recognition performance depends on the quantity and quality of the features extracted from face images.
- Low-resolution face recognition
  - The resolution is usually lower than $24\times24$ pixels, super-resolution techniques are needed.

Evaluation of image resolution and super-resolution on face recognition, JVCIR 2012

A Face Recognition System

- To develop a face recognition system:
  - face detection, facial feature detection, facial feature extraction; face recognition, face database and indexing.
Introduction

- Face recognition at normal or standard resolutions
  - Most of the existing face recognition algorithms were designed for images of standard resolutions, which can perform satisfactorily
  - Not much performance gain is achieved by existing face recognition methods when the image resolution increases, but this will become more computationally intensive

- High-resolution (HR) face recognition
  - New facial features can be explored
  - Rather than focusing on facial-feature appearances, such as the eyes, nose, mouth, etc., pore-scale features can be used
  - Our facial skin is rich in pore textures

- Three types of face recognition methods are considered
  - Low resolution
  - Standard resolution
  - High resolution
Low-Resolution Face Recognition

Approach 1:
- Face super-resolution, followed by feature extraction for face recognition

SR → Feature Extraction → Feature → Classification

Approach 2
- Facial-feature super-resolution, followed by face recognition

LR Feature → Feature SR → HR Feature → Classification
Low-Resolution Face Recognition

- Approach 3
  - Simultaneous face recognition and super-resolution
- By considering face super-resolution and recognition simultaneously, the accuracy of both super-resolution and recognition can be improved.

Simultaneous super-resolution and feature extraction for recognition of low-resolution faces, CVPR2008


Low-Resolution Face Recognition

- Facial features of a higher resolution (HrR) image are estimated directly from a low-resolution face image
- Gabor features are used
- Combine the Gabor features at different resolutions using Generalized Canonical Correlation Analysis (GCCA) for face recognition
Gabor Features

- A GW is a complex exponential modulated by a Gaussian function.
- These kernels are similar to the response of the two-dimensional receptive field profiles of the mammalian simple cortical cell.

$$\Psi(x, y, \omega, \theta) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x\cos\theta+y\sin\theta)^2 + (-x\sin\theta+y\cos\theta)^2}{2\sigma^2}} \cdot \left[ e^{i(\omega x \cos\theta + \omega y \sin\theta)} - e^{-\frac{\omega^2 \sigma^2}{2}} \right]$$

Gaussian function

sinusoidal function

$$e^{it} = \cos(t) + isin(t)$$

Innovative Technology Series

P.11
The GWs exhibit the desirable characteristics of capturing salient visual properties such as spatial localization, orientation selectivity, and spatial frequency selectivity.

- Gabor wavelets have been applied to:
  - Face detection
  - Face recognition
  - Object segmentation and tracking
  - Content-based texture image retrieval


The Fourier transform of a Gaussian function is also a Gaussian function.

\[ g(x) = e^{-ax^2} \leftrightarrow G(\omega) = \sqrt{\frac{\pi}{a}} e^{-\frac{\omega^2}{4a}} \]

- When multiplied by a sinusoidal function, the frequency spectrum of \( g(x) \) will be shifted up and down.

\[
\begin{align*}
  g(x)\sin\omega_0 x &\leftrightarrow \frac{i}{2} \left[ G(\omega + \omega_0) - G(\omega - \omega_0) \right] \\
  g(x)\cos\omega_0 x &\leftrightarrow \frac{1}{2} \left[ G(\omega + \omega_0) + G(\omega - \omega_0) \right]
\end{align*}
\]

Modulation Theorems
The Gabor functions with different scales and different orientations form a set of bandpass filters to extract features from an image.

Gabor Filters

\[ \omega_o^1 = \frac{\pi}{2} \]
\[ \omega_o^2 = \frac{\pi}{2\sqrt{2}} \]
\[ \omega_o^3 = \frac{\pi}{4} \]

Output of the Gabor Filters

\[ \theta^m = 0 \quad 0/8 \quad \pi/4 \quad 3\pi/8 \quad \pi/2 \quad 5\pi/8 \quad 3\pi/4 \quad 7\pi/8 \]
Feature Super-resolution

- The Gabor features of a higher resolution face image are estimated from its original LR face image
- **Local linear regression (LLR)** is employed to estimate the higher resolution features from an LR feature at a pixel position of a face image

Local linear regression

- A training set of \( N \) pairs of Gabor features of LR and HrR face images

\[
\mathbf{X}_{ij}^L = \left( \mathbf{x}_{ij,1}^L, \mathbf{x}_{ij,2}^L, \ldots, \mathbf{x}_{ij,N}^L \right) \leftrightarrow \mathbf{X}_{p,q}^H = \left( \mathbf{x}_{p,q,1}^H, \mathbf{x}_{p,q,2}^H, \ldots, \mathbf{x}_{p,q,N}^H \right)
\]

- \( \mathbf{x}_{ij,k}^L \): Gabor jet at \((i, j)\) of the \( k \)th LR training sample
- \( \mathbf{x}_{p,q,k}^H \): Gabor jet at \((p, q)\) of the \( k \)th HR training sample

\[
\mathbf{X}_{p,q}^H = \mathbf{A}_{p,q} \mathbf{X}_{ij}^L
\]

\[
\mathbf{A}_{p,q} = \mathbf{X}_{p,q}^H \left( \mathbf{x}_{ij}^L \right)^\top
\]

\[
= \mathbf{X}_{p,q}^H \left( \mathbf{x}_{ij}^L \right)^\top \left( \mathbf{x}_{ij}^L \left( \mathbf{x}_{ij}^L \right)^\top \right)^{-1}
\]
 Canonical Correlation Analysis (CCA)

- A method of correlating linear relationships between two sets of data
- Finding basis vectors for two sets of variables such that the correlations between the projections of the variables onto these basis vectors are mutually maximized
- Let $S_x$ denotes $(x_1, \ldots, x_N)$
  
  $S_y$ denotes $(y_1, \ldots, y_N)$

- These two sets of samples are projected onto $w_x$ and $w_y$ to form transformed samples
  
  $S_{x,w_x} = (w_x \cdot x_1, \ldots, w_x \cdot x_N)$
  
  $S_{y,w_y} = (w_y \cdot y_1, \ldots, w_y \cdot y_N)$


Innovative Technology Series
where \( y_i = u_i^T x \)

\[
\begin{align*}
\mathbf{x} &= y_1 u_1 + y_2 u_2 + y_3 u_3 \\
\mathbf{y} &= \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\
&= \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{u}^T \mathbf{x}
\end{align*}
\]
Choose \( \mathbf{w}_x \) and \( \mathbf{w}_y \) to maximize the correlation between the two vectors \( \mathbf{v}_i \) and \( \mathbf{z}_i \) \((i = 1, \ldots, N)\):

\[
\rho = \max_{\mathbf{w}_x, \mathbf{w}_y} \text{corr}(\mathbf{S}_{x,\mathbf{w}_x}, \mathbf{S}_{y,\mathbf{w}_y}) = \max_{\mathbf{w}_x, \mathbf{w}_y} \frac{\mathbf{S}_{x,\mathbf{w}_x} \cdot \mathbf{S}_{y,\mathbf{w}_y}}{\left\| \mathbf{S}_{x,\mathbf{w}_x} \right\| \left\| \mathbf{S}_{y,\mathbf{w}_y} \right\|}
\]

\[
= \max_{\mathbf{w}_x, \mathbf{w}_y} \frac{E[\mathbf{w}_x, \mathbf{x}]\mathbf{w}_y, \mathbf{y}]}{\sqrt{E[\mathbf{w}_x, \mathbf{x}]^2} E[\mathbf{w}_y, \mathbf{y}]^2]
\]

\[
= \max_{\mathbf{w}_x, \mathbf{w}_y} \frac{E[\mathbf{w}_x, \mathbf{x}]\mathbf{w}_y}{\sqrt{E[\mathbf{w}_x, \mathbf{x}]^2} E[\mathbf{w}_y, \mathbf{y}]^2}
\]

\[
= \max_{\mathbf{w}_x, \mathbf{w}_y} \frac{\mathbf{w}_x, E[\mathbf{x}, \mathbf{y}] \mathbf{w}_y}{\sqrt{E[\mathbf{x}, \mathbf{y}]^2} E[\mathbf{x}, \mathbf{w}_x] E[\mathbf{y}, \mathbf{w}_y]}
\]

\[
= \max_{\mathbf{w}_x, \mathbf{w}_y} \frac{\mathbf{w}_x, E[\mathbf{x}, \mathbf{y}] \mathbf{w}_y}{\sqrt{E[\mathbf{x}, \mathbf{y}]^2} E[\mathbf{x}, \mathbf{w}_x] E[\mathbf{y}, \mathbf{w}_y]}
\]

\[
\mathbf{w}_x \mathbf{x} = \mathbf{x}^T \mathbf{w}
\]

\[
\rho = \max_{\mathbf{w}_x, \mathbf{w}_y} \frac{\mathbf{w}_x^T \mathbf{C}_{xy} \mathbf{w}_y}{\sqrt{\mathbf{w}_x^T \mathbf{C}_{xx} \mathbf{w}_x \mathbf{w}_y^T \mathbf{C}_{yy} \mathbf{w}_y}}
\]

The solution can be obtained by solving the following equations:

\[
\mathbf{w}_y = \frac{\mathbf{C}_{yy}^{-1} \mathbf{C}_{yx} \mathbf{w}_x}{\lambda}
\]

\[
\mathbf{C}_{xy} \mathbf{C}_{yy}^{-1} \mathbf{C}_{yx} \mathbf{w}_x = \lambda^2 \mathbf{C}_{xx} \mathbf{w}_x
\]
Generalized CCA

- An extension of CCA
- Define the within-set scatter matrices of $S_x$ and $S_y$:

$$C_{W_x} = \sum_{i=1}^{c} P(\omega_i) \left[ \frac{1}{n_i} \sum_{j=1}^{n_i} (x_{ij} - m_x^i)(x_{ij} - m_x^i)^T \right]$$

$$C_{W_y} = \sum_{i=1}^{c} P(\omega_i) \left[ \frac{1}{n_i} \sum_{j=1}^{n_i} (y_{ij} - m_y^i)(y_{ij} - m_y^i)^T \right]$$

- Define the between-set scatter matrix:

$$L_{xy} = \frac{1}{n} \sum_{i=1}^{n} (x_i - m_x)(y_i - m_y)^T$$

“A theorem on the generalized canonical projective vectors,” Pattern Recognition, 2005

The criterion function:

$$J(\alpha, \beta) = \frac{\alpha^T L_{xy} \beta}{\sqrt{\alpha^T C_{W_x} \alpha \beta^T C_{W_y} \beta}}$$

- The two vectors $\{\alpha_i\}$ and $\{\beta_i\}$ that maximize the criterion function are called the generalized canonical projective vectors (GCPV)

- The generalized canonical discriminant features (GCDF):

$$X^* = [\alpha_1 \quad \alpha_2 \quad \ldots \quad \alpha_d]^T x = W_x^T x$$

$$Y^* = [\beta_1 \quad \beta_2 \quad \ldots \quad \beta_d]^T y = W_y^T y$$
The two GCDFs can be combined to form a single feature vector for classification:

\[
Z_1 = \begin{bmatrix} X^* \\ Y^* \end{bmatrix} = \begin{bmatrix} W_x^T \\ W_y^T \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = W_1 \begin{bmatrix} x \\ y \end{bmatrix}
\]

\[
Z_2 = X^* + Y^* = W_x^T x + W_y^T y = \begin{bmatrix} W_x \\ W_y \end{bmatrix}^T \begin{bmatrix} x \\ y \end{bmatrix} = W_2 \begin{bmatrix} x \\ y \end{bmatrix}
\]

– canonical correlation discriminant features (CCDFs)

---

**Low-resolution Face Recognition**

- Facial information at different resolutions is fused for face recognition
- The query face image is a LR image, whose Gabor feature vector, after applying PCA, is denoted as \(x^L\)
- Using Gabor-feature super-resolution to estimate the Gabor features at two higher resolutions, denoted as (medium-resolution) \(x^M\) and (high-resolution) \(x^H\)
- **Type-A CGCCA-LR:** $x^L$ and $x^M$ are fused at the first stage, and the combined feature is fused with $x^H$ at the second stage.

- **Type-B CGCCA-LR:** $x^L$ is fused with $x^M$ and $x^H$ at the first stage to form two combined feature vectors at the first stage, which are then fused to form a single feature vector for face recognition.

- The corresponding four CCDFs are:

$$Z^A_{1LR}, Z^A_{2LR}, Z^B_{1LR}, Z^B_{2LR}$$

Standard Resolution Face Recognition

- Facial information at different resolutions is fused for face recognition
- Each face image is down-sampled and up-sampled, so three face images at different resolutions are available:
  - higher resolution (HR)
  - original resolution (OR)
  - lower resolution (LR)

Multi-resolution feature fusion (MFF) face recognition
The three features at different resolutions can be fused in two stages; two features are fused in each stage

**Type-A CGCCA & Type-B CGCCA**

- **Type-A CGCCA**: The OR and HR Gabor features are fused in the first stage, which is then fused with the LR feature in the second stage
- **Type-B CGCCA**: The fusion between the OR and HR Gabor features, and between the OR and LR Gabor features are computed at the first stage, and the resulting features are fused at the second stage

**Experiments**

- We conduct experiments to evaluate the performance of
  - the Gabor-feature super-resolution method
  - the proposed MMF algorithm
  - the method for LR face recognition
- The proposed algorithms are compared with PCA and Laplacianfaces
- Two databases are used
  - ORL face database: 10 images for each of the 40 distinct subjects
  - FERET face database: 4 images for each of the randomly selected 200 subjects
Performance of Gabor-feature super-resolution

- The ORL database is used
- The original image resolution is set at 16×16, and high resolution is set at 48×48
- Gabor features of 5 scales and 8 orientations are extracted
- The HR Gabor features can also be generated using bicubic interpolation followed by Gabor filters, recognition rate = 0.686

<table>
<thead>
<tr>
<th>No. of training images for each person</th>
<th>Recognition rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.030</td>
</tr>
<tr>
<td>2</td>
<td>0.645</td>
</tr>
<tr>
<td>3</td>
<td>0.695</td>
</tr>
<tr>
<td>4</td>
<td>0.702</td>
</tr>
<tr>
<td>5</td>
<td>0.713</td>
</tr>
<tr>
<td>6</td>
<td>0.717</td>
</tr>
<tr>
<td>7</td>
<td>0.714</td>
</tr>
<tr>
<td>8</td>
<td>0.725</td>
</tr>
<tr>
<td>9</td>
<td>0.725</td>
</tr>
</tbody>
</table>

Performances of MFF face recognition algorithms

- Both ORL and FERET databases are used
- The resolutions used are: (LR, OR, HR) = (8×8, 24×24, 72×72)

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type-A CGCCA-1</td>
<td>Type-A feature-fusion method</td>
</tr>
<tr>
<td>Type-A CGCCA-2</td>
<td>Type-A feature-fusion method</td>
</tr>
<tr>
<td>Type-B CGCCA-1</td>
<td>Type-B feature-fusion method</td>
</tr>
<tr>
<td>Type-B CGCCA-2</td>
<td>Type-B feature-fusion method</td>
</tr>
<tr>
<td>Type-A CGCCA_FSR-1</td>
<td>Type-A feature-fusion method with FSR</td>
</tr>
<tr>
<td>Type-ACGCCA_FSR-1</td>
<td>Type-A feature-fusion method with FSR</td>
</tr>
<tr>
<td>Type-B CGCCA_FSR-2</td>
<td>Type-A feature-fusion method with FSR</td>
</tr>
<tr>
<td>Type-B CGCCA_FSR-2</td>
<td>Type-A feature-fusion method with FSR</td>
</tr>
<tr>
<td>Resolutions</td>
<td>Algorithms</td>
</tr>
<tr>
<td>-------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>OR = 24×24</td>
<td>PCA</td>
</tr>
<tr>
<td>OR = 24×24</td>
<td>Laplacianfaces</td>
</tr>
<tr>
<td>OR = 24×24</td>
<td>Type-A CGCCA-1</td>
</tr>
<tr>
<td>OR = 24×24</td>
<td>Type-A CGCCA-2</td>
</tr>
<tr>
<td>OR = 24×24</td>
<td>Type-B CGCCA-1</td>
</tr>
<tr>
<td>OR = 24×24</td>
<td>Type-B CGCCA-2</td>
</tr>
<tr>
<td>LR = 8×8</td>
<td></td>
</tr>
<tr>
<td>OR = 24×24</td>
<td>Type-A CGCCA_FSR-1</td>
</tr>
<tr>
<td>OR = 24×24</td>
<td>Type-A CGCCA_FSR-2</td>
</tr>
<tr>
<td>OR = 24×24</td>
<td>Type-B CGCCA_FSR-1</td>
</tr>
<tr>
<td>OR = 24×24</td>
<td>Type-B CGCCA_FH-2</td>
</tr>
<tr>
<td>HR = 72×72</td>
<td></td>
</tr>
</tbody>
</table>

**Performances of MFF-based LR face recognition**
- The query is a LR face image – denoted as OR face
- Gabor features at two higher resolutions: medium-resolution (MR) and high-resolution (HR) are estimated
- Two sets of resolutions (OR, MR, HR) are considered: (12×12, 24×24, 48×48) and (16×16, 32×32, 48×48)
- The ORL and FERET databases are used
- The four algorithms based on the CCDFs are:
  - **Type-A CGCCA-LR-1:** $Z^A_{1LR}$
  - **Type-A CGCCA-LR-2:** $Z^A_{2LR}$
  - **Type-B CGCCA-LR-1:** $Z^B_{1LR}$
  - **Type-B CGCCA-LR-2:** $Z^B_{2LR}$
<table>
<thead>
<tr>
<th>Resolutions</th>
<th>Algorithms</th>
<th>Recognition Rates on the ORL database</th>
<th>Recognition Rates on the FERET database</th>
</tr>
</thead>
<tbody>
<tr>
<td>OR = 12×12</td>
<td>PCA</td>
<td>0.625</td>
<td>0.530</td>
</tr>
<tr>
<td>OR = 12×12</td>
<td>Laplacianfaces</td>
<td>0.675</td>
<td>0.615</td>
</tr>
<tr>
<td>OR = 12×12</td>
<td>Type-A CGCCA-LR-1</td>
<td>0.825</td>
<td>0.795</td>
</tr>
<tr>
<td>MR = 24×24</td>
<td>Type-A CGCCA-LR-2</td>
<td>0.825</td>
<td>0.755</td>
</tr>
<tr>
<td>HR = 36×36</td>
<td>Type-B CGCCA-LR-1</td>
<td>0.825</td>
<td>0.795</td>
</tr>
<tr>
<td></td>
<td>Type-B CGCCA-LR-2</td>
<td>0.850</td>
<td>0.810</td>
</tr>
<tr>
<td>OR = 16×16</td>
<td>PCA</td>
<td>0.651</td>
<td>0.645</td>
</tr>
<tr>
<td>OR = 16×16</td>
<td>Laplacianfaces</td>
<td>0.775</td>
<td>0.680</td>
</tr>
<tr>
<td>OR = 16×16</td>
<td>Type-A CGCCA-LR-1</td>
<td>0.850</td>
<td>0.865</td>
</tr>
<tr>
<td>MR = 32×42</td>
<td>Type-A CGCCA-LR-2</td>
<td>0.850</td>
<td>0.850</td>
</tr>
<tr>
<td>HR = 48×48</td>
<td>Type-B CGCCA-LR-1</td>
<td>0.850</td>
<td>0.865</td>
</tr>
<tr>
<td></td>
<td>Type-B CGCCA-LR-2</td>
<td>0.865</td>
<td>0.860</td>
</tr>
</tbody>
</table>

**High-Resolution Face Recognition**

- Pore-scale features are explored
Pore-scale Facial-Feature Detection

DoG (Difference-of-Gaussian) Detector
- Only darker keypoints are detected as pores
- Similar quantity for different people
  → adaptive threshold
- Pore-scale facial-feature Modeling

\[ \text{pore}(x, y, \sigma) = 1 - 2\pi\sigma^2 G(x, y, \sigma) \]

Dong Li and Kin-Man Lam, "Design and Learn Distinctive Features from Pore-scale Facial Keypoints," Pattern Recognition, March 2015.

---

Gaussian kernel is used to create scale space
- i.e. to convolve an image with a Gaussian kernel of different scale or variance

\[ L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \]
\[ G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \]

- Stable keypoints can be located by detecting the extrema in the difference-of-Gaussian functions
Pore-scale Facial-Feature Detection

- **DoG**: difference of Gaussians

\[ D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma))^* \]

\[ l(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \]

Only the maxima of the DoG are detected to locate the darker keypoints in face regions.
Pore-scale Facial-Feature Detection

- The DoG is constructed in octaves, which have the $\sigma$ doubled in the scale space.

<table>
<thead>
<tr>
<th>Scale (first octave)</th>
<th>Scale (second octave)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma$</td>
<td>$k\sigma$</td>
</tr>
<tr>
<td>$k\sigma$</td>
<td>$k^2\sigma$</td>
</tr>
</tbody>
</table>

Difference of Gaussians (DoG)

- A sample point is selected only if it is larger than all its neighbors (3x3 regions) in the current and adjacent layers.

A DoG
By experiments, we found that an appropriate number of keypoints in a whole face region is about 5,000. This number is affected by the dense DoG responses at hairy areas.

To evaluate the skin conditions and determine the parameters, a hairless cheek region is cropped.

A small facial-skin region is sufficient for face recognition.

For the cropped region, the number of keypoints $N_k \in [450, 500]$.

A threshold should be set adaptively for selecting keypoints, which is related to the peak value $P$ of the DoG response.

Consider the DoG response of a pore:

$$D_{pore}(x, y, \sigma_1, \sigma_2) = [G(x, y, k\sigma_1) - G(x, y, \sigma_1)] \ast pore(x, y, \sigma_2)$$

$$= \int_{-\infty}^{\infty} [G(u, v, k\sigma_1) - G(u, v, \sigma_1)]pore(x - u, y - v, \sigma_2) dudv$$
Pore-scale Facial-Feature Detection

- The response at the centre \((x = 0, y = 0)\):
  \[
  D_{\text{pore}}(x = 0, y = 0, \sigma_1, \sigma_2) = \int_{-\infty}^{\infty} [G(u, v, k \sigma_1) - G(u, v, \sigma_1)]pore(-u, -v, \sigma_2) dudv
  \]
  \[
  = -\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} - \frac{\sigma_2}{k^2 \sigma_1^2 + \sigma_2^2}
  \]
  The maximum is determined by taking the derivative with respect to \(\sigma_1\) and setting it at zero:
  \[
  \hat{\sigma}_1 = k^{-1/2} \sigma_2
  \]
  - The peak value \(P = \frac{k - 1}{k + 1}\)
  - \(k\) is usually set at \(\sqrt{2}\)

- The adaptive threshold \(\tau\) is set within \([0, 0.2P]\) such that the resulting \(N_k \in [450, 500]\)

- Most of the pore-scale facial features are tiny and of low contrast, the parameters should be determined to achieve good performance

- The Bosphorus database was used: the 4 subjects that have been shown + 16 more face images

- Skin regions in different poses are grouped to form 4 datasets (10° and 20°, 10° and 30°, 20° and 30°, and 20° and 45°)
Pore-scale Facial-Feature Detection

- Parameter Selection: sampling frequency in scale

Note: RANSAC is used to remove the outliers

*Inlier Rate = No. of inliers / No. of matches

Eight scales are sampled in each octave

Innovative Technology Series 51

Pore-scale Facial-Feature Detection

- Parameter selection: Sampling frequency in the spatial domain

Innovative Technology Series 52
Different local descriptors can be used

We apply the SIFT descriptor to form a local descriptor for a pore keypoint – named PSIFT

<table>
<thead>
<tr>
<th>Parameters</th>
<th>PSIFT</th>
<th>SIFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of subregions</td>
<td>8×8</td>
<td>4×4</td>
</tr>
<tr>
<td>Support size of total subregions</td>
<td>48×σ₂</td>
<td>12×σ₂</td>
</tr>
<tr>
<td>Support size of each subregion</td>
<td>6×σ₂</td>
<td>3×σ₂</td>
</tr>
<tr>
<td>No. of orientation bins</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Dimension of the feature</td>
<td>512</td>
<td>128</td>
</tr>
</tbody>
</table>
Relative-position Descriptor

- Use of gradient orientation histograms
  - Robust representation
- Example: $2 \times 2$ subregions
  - Four 8-bin histograms
  - Each bin represents a range of certain orientations
  - Bin value = sum of gradient magnitudes

Pore-to-pore Correspondences Dataset

- Face images with 10-, 20-, 30- and 45-degree poses of 105 subjects in Bosphorus Database are used
- 420 cropped cheek-region images
- Matching based on PSIFT and RANSAC
- A track is a set of matched keypoints across the face images of a subject at different poses
- 4,240 tracks is established containing 4 keypoints corresponding to the 10-, 20-, 30- and 45-degree pose
Pore-to-pore Correspondences Dataset

- A supervised learning procedure based on Linear Discriminant Analysis (LDA) is proposed
- 4,240 classes (tracks), 4 pore images in a class
- By projecting the PSIFT descriptor to the projection vectors learned, the resulting descriptors are more discriminative and have a more compact size
  - Namely, LDAPSIFT

Experiments on Cropped Skin images

- Dataset
  - Bosphorus face database
  - 105 skin-region pairs from 210 face images, captured at the poses of 10° and 45°
  - Original resolution is about 1,400×1,200 pixels
  - After cropped, the skin region is about 350×350 pixels
Experiments on Cropped Skin images

<table>
<thead>
<tr>
<th>Method</th>
<th>Average No. of Inliers</th>
<th>No of image pairs with more than 20 inliers</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDAPSIFT</td>
<td>89.33</td>
<td>102</td>
</tr>
<tr>
<td>PSIFT</td>
<td>73.86</td>
<td>96</td>
</tr>
<tr>
<td>SIFT detector + PSIFT</td>
<td>25.94</td>
<td>44</td>
</tr>
<tr>
<td>PSIFT detector + SIFT</td>
<td>8.65</td>
<td>11</td>
</tr>
<tr>
<td>SIFT</td>
<td>3.66</td>
<td>5</td>
</tr>
</tbody>
</table>

Face Matching Results
Face Matching of Identical Twin Images

Face Matching for Images Captured by iPhone 5s
Conclusions

- Multi-resolution facial-feature fusion can be employed for face recognition at different resolutions
- High-resolution face recognition based on pore-scale features is a new research area
- This approach can solve the problems (pose, facial expression, lighting, occlusion, etc.) of existing face recognition algorithms at the same time
- Particularly effective to recognize identical twins
- Useful for 3D face reconstruction
- Future work:
  - Skin decomposition, non-rigid point matching, build a high-resolution face database, etc.
- End -