Automatic Asymmetric 3D-2D Face Recognition

Di Huang[1], Mohsen Ardabilian[1], Yunhong Wang[2], and Liming Chen[1]
[1]Université de Lyon, CNRS, Ecole Centrale Lyon, LIRIS UMR, 69134, Lyon, France
[2]School of Computer Science and Engineering, Beihang University, 100091, Beijing, China

...@liris.cnrs.fr - http://liris.cnrs.fr/...

Laboratoire d'InfoRmatique en Image et Systèmes d'information

LIRIS UMR 5205 CNRS/INSA de Lyon/Université Claude Bernard Lyon 1/Université Lumière Lyon 2/Ecole Centrale de Lyon
Université Claude Bernard Lyon 1, bâtiment Nautibus
43, boulevard du 11 novembre 1918 — F-69622 Villeurbanne cedex
http://liris.cnrs.fr

Outline

- Background
- The Proposed Method Overview
- Data Preprocessing
- Asymmetric 3D-2D Face Recognition
- Experimental Results
- Conclusions

Background

- General difficulties of face recognition
  - 2D facial image based recognition
    - Illumination
    - Expression
    - Pose
  - 3D face model based recognition
    - Expression
    - Acquisition and Computation Cost (Registration)

Background

- Comparisons of 2D and 3D face recognition
  - 3D data has more information for face representation
  - 3D face recognition is robust to illumination and pose
  - 2D data is easier for acquisition
  - 2D face recognition needs less computation cost

Background

- Goal: combining the advantages of 2D and 3D face data for robust face recognition; and limiting the use of 3D data to where it really helps to improve performance.
  - Accuracy: better than 2D-2D, comparable to 3D-3D
  - Time-Consuming: much less than 3D-3D

Background

- The state-of-the-art
  - Rama et al. proposed to project 3D texture information in cylindrical coordinate as gallery and 2D image was used as probe, then applied Partial Principle Component Analysis for feature extraction. However, 3D shape information was actually not used. [ICASSP 2006]
  - Riccio et al. proposed to utilize control keypoints to compute several geometrical invariants for face recognition. 3D shape is as gallery and 2D images as probe. But it is not an easy task to accurately locate the control points. [ICIAP 2005]
  - Yang et al. applied patch based kernel CCA to learn mapping between range (gallery) and 2D images (probe) directly. But the performance was not reliable enough. [FG 2008]

Background

What's new in our method?

- More information:

  Gallery Set: Textured 3D face (3D point-clouds and corresponding 2D images)

  Probe Set: 2D facial images

- Illumination normalization and pose correction
- Description of local texture changes of 2D face images as well as local shape variations of facial range images

Outline

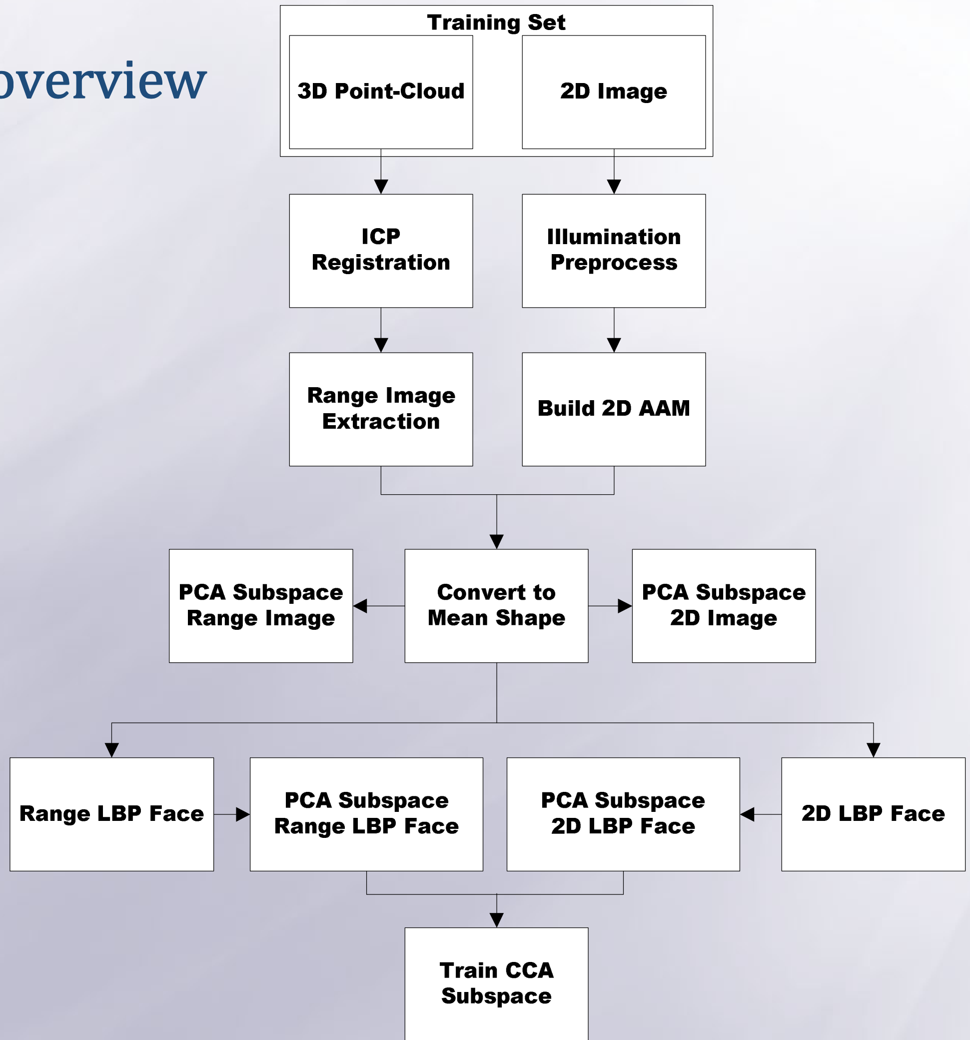
- Background
- The Proposed Method Overview
- Data Preprocessing
- Asymmetric 3D-2D Face Recognition
- Experimental Results
- Conclusions

Overview

- The proposed method overview
  - 2D-2D face matching
  - 3D-2D face matching
  - Score fusion

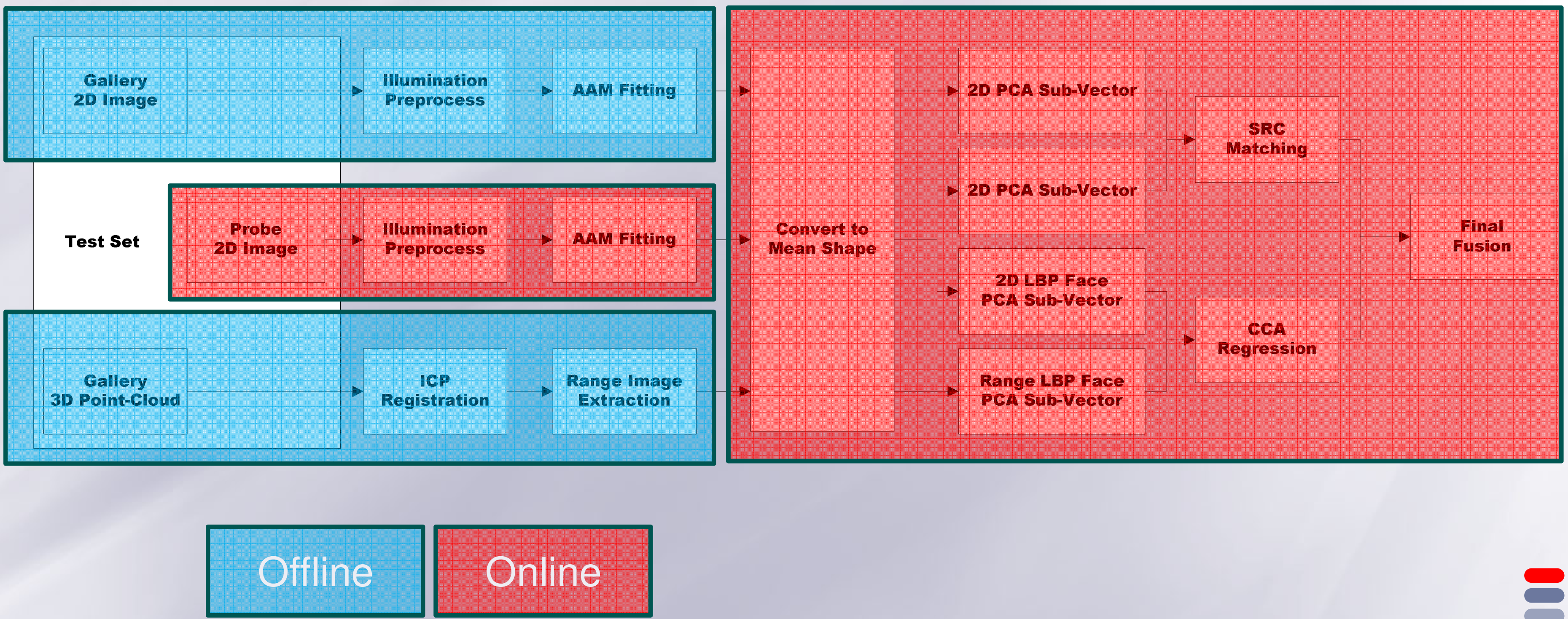
Overview

- The proposed method overview
  - Training stage
  - Test stage



Overview

- The proposed method overview
  - Training stage
  - Test stage



Outline

- Background
- The Proposed Method Overview
- Data Preprocessing
- Asymmetric 3D-2D Face Recognition
- Experimental Results
- Conclusions

Data preprocessing

2D data preprocessing

- Logarithmic Total Variation (LTV) based Illumination Normalization

It works on any single image without any prior information on 3D face geometry or light source. It not only inherits ability from the TV-L1 model to decompose image f into a large-scale output u and a small-scale output v, but also brings properties of edge-preserving and multiscale additive signal decomposition.

$$I(x,y) = \rho(x,y) \cdot S(x,y)$$

$$f(x,y) = \log(I(x,y)) = \log(\rho(x,y)) + \log(S(x,y))$$

$$u^* = \arg\min\{\int |\nabla u| + \lambda \|f - u\|_L\}$$

$$v^* = f - u^*$$

$$S \approx \exp(u^*), \;\; \rho \approx \exp(v^*)$$

[Ref.] T. Chen, X. S. Zhou, D. Comaniciu and T.S. Huang, “Total Variation Models for Variable Lighting Face Recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 28, no. 9, pp. 1519-1524, 2006.

Data preprocessing

2D data preprocessing

- Active Appearance Model (AAM) based Pose Correction

$$x = \bar{x} + Q_x c$$

$$g = \bar{g} + Q_g c$$

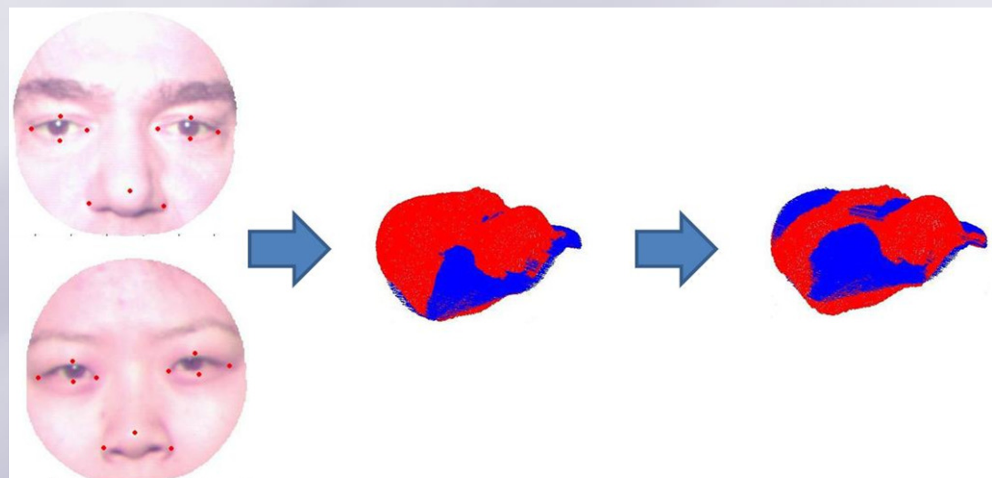
[Ref.] T. F. Cootes, G. J. Edwards, and C. J. Taylor, “Active Appearance Models,” *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 23, no. 6, pp. 681-685, 2001.

Data preprocessing

- 3D data preprocessing
  - Region based Iterative Closet Point (R-ICP) Registration

    Rigid face region

    Coarse to fine Strategy



Outline

- Background
- The Proposed Method Overview
- Data Preprocessing
- Asymmetric 3D-2D Face Recognition
- Experimental Results
- Conclusions

Asymmetric 3D-2D Face Recognition

- Asymmetric 3D-2D face recognition
  - 2D-2D matching
  - 3D-2D matching
  - Score fusion

Asymmetric 3D-2D Face Recognition

2D-2D matching based on Sparse Representation Classifier (SRC)

k classes and n_i feature vectors, $v_{i,j} \in R_m$ are utilized for training from the i^{th} class, $i = \{1, 2, \ldots, k\}$ and j is the index of the training sample, $j = \{1, 2, \ldots, n_i\}$. All the training data from the i^{th} class are placed in a matrix $A_i = [v_{i,1}, v_{i,2}, \ldots, v_{i,ni}] \in R_{m \times ni}$. A dictionary matrix A for k classes is developed by concatenating A_i, $i=1, \ldots, k$. A test pattern y can be represented as a linear combination of all n training samples ($n = n_i \times k$): $y = Ax$

where x is an unknown coefficient vector. It is straightforward to note that only the entries of x that are non-zero correspond to the class of y. It can be solved as long as its solution is known to be sufficiently sparse. An equivalent L_1-norm minimization:

$$(L1): x_1 = \arg\min \|x\|_1 ;\ \ Ax = y$$

can be solved as a good approximation. With the solution x_1, we can compute the residual between a given probe face and each individual as:

$$R_i = \left\| y - \sum_{j=1}^{k} x_{1_{i,j}} v_{i,j} \right\|_2$$

The identity of any given probe face is determined as the one with the smallest residual R.

[Ref.] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 31, no. 2, pp. 210-227, 2009.

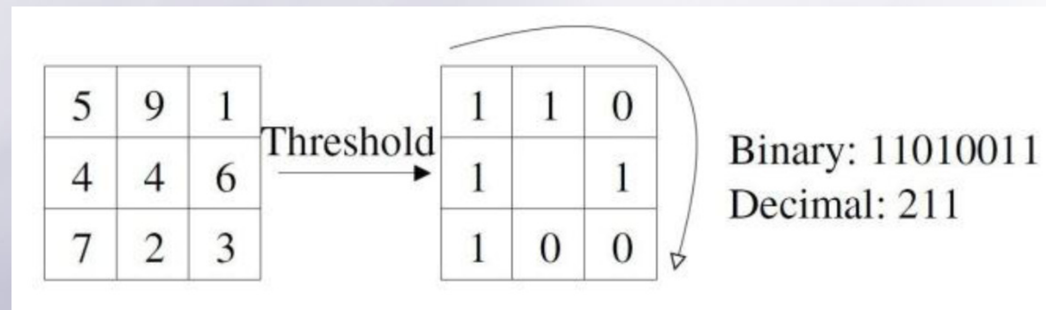
Asymmetric 3D-2D Face Recognition

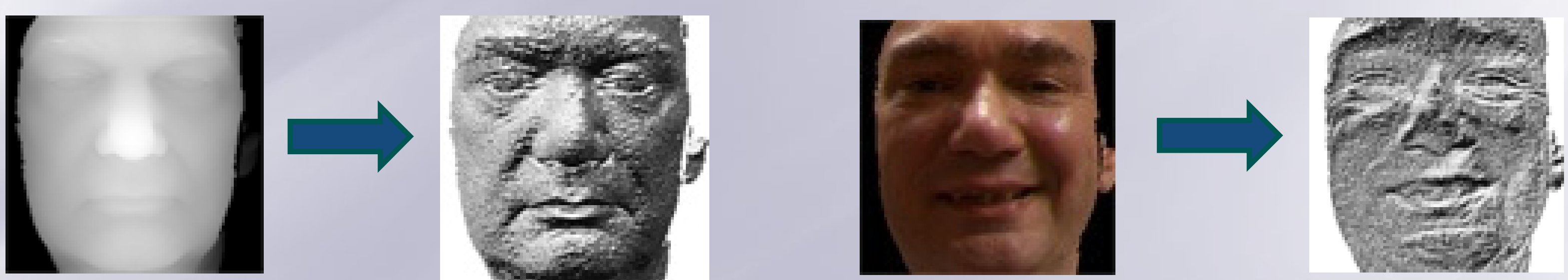
- Description of local texture and shape variations

  The multi-scale Local Binary Patterns (LBP) based method is applied to describe local texture variations of 2D facial images and local shape changes of facial range images.

$$LBP(x_c, y_c) = \sum_{n=0}^{P-1} s(i_n - i_c)2^n$$

$$s(x) = \begin{cases} 1 & if \quad x \geq 0 \\ 0 & if \quad x < 0 \end{cases}$$





Asymmetric 3D-2D Face Recognition

3D-2D matching based on Canonical Correlation Analysis (CCA)

CCA is a powerful analysis algorithm especially useful for relating two sets of variables, by maximizing the correlation in the CCA subspace.

Given N pairs of samples (x_i, y_i) of (X, Y), i=1, 2,..., N, where , $X \in \mathbb{R}^p \; Y \in \mathbb{R}^q$, with mean value of zero. The goal of CCA is to learn a pair of directions w_x and w_y to maximize correlation between $x=w_x^T X$ and $y=w_y^T Y$. In the context of CCA, two projections: x and y are also referred to as the canonical variant. Formally, the directions can be found as the maxima of the function:

$$\rho = \frac{E[w_x^T X Y^T w_y]}{\sqrt{E[w_x^T X X^T w_x] E[w_y^T Y Y^T w_y]}}$$

To test new pairs of variables, we first project them into CCA subspace $x'=w_x^T X'$; $y'=w_y^T Y'$; and then their similarity is calculated, and bigger value indicates higher similarity.

$$S(x', y') = \frac{x' \cdot y'}{\|x'\| \|y'\|}$$

Asymmetric 3D-2D Face Recognition

Score fusion

Finally, using Min-Max normalization, both the matching scores from 2D-2D and 3D-2D are normalized to [0, 1], and a weighted sum rule is used for fusion.

$$F = w_S * S_S + w_R * (1 - S_R)$$

The corresponding weights w_S and w_R are calculated dynamically during the online step using:

$$w_{S_i} = \frac{mean(S_i) - \min_1(S_i)}{mean(S_i) - \min_2(S_i)}$$

where i corresponds to the two similarity measures: S, and R, and operators $\min_1(S_i)$ and $\min_2(S_i)$ produce the first and second minimum value of the vector S_i.

Outline

- Background
- The Proposed Method Overview
- Data Preprocessing
- Asymmetric 3D-2D Face Recognition
- Experimental Results
- Conclusions

Database

- The complete FRGC v2.0 dataset
  - 4007 textured 3D face models of 466 subjects
  - Median filter is utilized for removing the spikes
  - Cubic interpolation is adopted for filling the holes
- Settings
  - The first model with neutral expression of each subject makes up a gallery set of 466, while the other models (4007-466=3541) are used for test. All images are cropped to 175×190 pixels.

Results

Face recognition

Rank-one recognition rate of 2D-2D matching using different LBP operators

2D-2D	R-1	R-2	R-3	R-4	R-5	R-6	R-7	R-8	Fusion
P-08	0.7617	0.8204	0.8323	0.8371	0.8456	0.8416	0.8156	0.7631	0.8735
P-12	0.7535	0.7899	0.8250	0.8252	0.8340	0.8224	0.7959	0.7504	0.8775
P-16	0.7521	0.7928	0.7899	0.8106	0.8128	0.8046	0.7801	0.7250	0.8662

Results

Face recognition

Rank-one recognition rate of 2D-2D matching using different LBP operators

3D-2D	R-1	R-2	R-3	R-4	R-5	R-6	R-7	R-8	Fusion
P-08	0.6843	0.7648	0.7868	0.8080	0.8210	0.8331	0.8303	0.8204	0.8528
P-12	0.7032	0.7436	0.7953	0.8145	0.8267	0.8233	0.8272	0.8278	0.8530
P-16	0.7171	0.7623	0.7866	0.8086	0.8139	0.8202	0.8199	0.8182	0.8564

Results

Face recognition (Rank-One)

2D-2D Matching	Results
(S1) Original faces	0.786
(S2) *LBP* Histograms	0.846
(S3) *MS-LBP* Histograms	0.876

3D-2D Matching	Results
(A1) Original faces	0.815
(A2) *LBP* Face Map	0.833
(A3) *MS-LBP* Face Map	0.856

Final Matching	Results
(F1): (S1) + (A1)	0.832
(F2): (S3) + (A3)	0.903

Outline

- Background
- The Proposed Method Overview
- Data Preprocessing
- Asymmetric 3D-2D Face Recognition
- Experimental Results
- Conclusions

Conclusions

- The proposed asymmetric 3D-2D face recognition works better than 2D-2D, since it uses 3D information, while it avoids high online cost of data acquisition and registration.
- CCA learning based on LBP faces outperforms that on the original images.

Future Work

- Enhance AAM with more samples in different poses
- Other descriptors to enhance local variations of faces to further improve the result
- Test the robustness of asymmetric face recognition approach especially to facial expression variations.

The End

Thank you!

Q&A

Q&A