Multi-modal Face Recognition

Hu Han
hanhu@ict.ac.cn
http://vipl.ict.ac.cn/members/hhan
2016/04/06
Outline

- Background
- Related work
- Multi-modal & cross-modal FR
- Trend on multi-modal (face) recognition
- Conclusion and discussion
Background

Unconstrained sensing & uncooperative subject scenario poses great challenges to unimodal FR system

“It’s our intention to go through every frame of every video.”
– Boston Police Commissioner, Ed Davis

“We are particularly interested in reviewing video footage captured by bystanders with cell phones or personal cameras near either of the blasts… In an investigation of this nature, no detail is too small.”
– Attorney General, Eric H. Holder Jr.
Background

- Unimodal FR
  - A manually selected probe face image of the suspect (Tamerlan Tsarnaev) with the best quality is matched with its true mate by a COTS with rank-5000 among a 1M gallery set
Background

- Challenges
  - Low quality surveillance videos and images
  - Intentional thwarting of identification (e.g. sunglasses and hats)
  - Daunting amount of data
  - Videos or images are n/a
  - ...

- Multi-modal FR is a possible solution
  - Advances in computing and imaging tech.
    - RGB, depth, NIR, 3D, sketch, etc.
  - Multi-modality, multi-view, multi-biometrics
Background

Traditional Forensic Investigation Workflow

Human operators manually review $K \times n$ images ($n =$ # of images in the face media collection)
Outline

- Background
- Related work
- Multi-modal & cross-modal FR
- Trend on multi-modal (face) recognition
- Conclusion and discussion
Related work

- Multi-modal FR
  - 2D + 3D
    - Beumier and Acheroy, PRL’01
    - Chang et al., ACM-W’03
    - ...
  - 2D + depth
    - Lu and Jain, TPAMI’06
  - 2D + 3D + NIR
    - Bowyer et al., 2003-2011

- Most are: Per-modal matching + score-level fusion
Related work

- **Cross-modal FR**
  - Modality transformation
    - Wang & Tang, TPAMI’09 (sketch vs. photo)
    - Gao et al., TCSVT’12 (sketch vs. photo)
    - 3D face modeling, Blanz & Vetter’03 (2D vs. 3D)
    - ...

- Invariant features
  - Lei & Li, CVPR’09
    - VIS-NIR
  - Klare & Jain, TPAMI’13; Han & Jain TIFS’13; Klum et al., TIFS’14
    - VIS-NIR, forensic sketch, VIS-TIR
Outline

- Background
- Related work
- Multi-modal & cross-modal FR
  - Multi-modal FR
  - Trend on multi-modal (face) recognition
- Conclusion and discussion
Multi-modal face recognition

Gender: Male  Race: White  Age: 60-70

Still image

Sketch

Video

3D

Human operators manually review 1*K images
Multi-modal face recognition

A hierarchical quality-based fusion

Quality measures: Q1, Q2, Q3, Q4

1M Mugshot, True mate is matched at rank-112 (vs. rank-5000 in unimodal)

\[ s = \frac{1}{m} \sum_{i=1}^{m} q_i s_i = Q^T S \]
Multi-modal face recognition

Matching a Face Track from a Video

Multi-frame Score-level Fusion:
- mean
- median
- max
- min

\[ U = u_1, u_2, ..., u_a \]
\[ V = v_1, v_2, ..., v_b \]

All Frame Pairs

COTS Face Matcher

\[ s(U, V) \]

\[ s(u_i, v_j) \]

Similarity Matrix

Not Same

\[ s(U, V) \geq t \]

\[ s(U, V) < t \]
Multi-modal face recognition

- Pose Correction via 3D Face Modeling

\[ q(I_{PC}) = \frac{1}{t} \sum_{i=1}^{t} SSIM(I_{PC}, R_i) \]

\[ = \frac{1}{t} \sum_{i=1}^{t} I(I_{PC}, R_i)^{\alpha} \cdot c(I_{PC}, R_i)^{\beta} \cdot s(I_{PC}, R_i)^{\gamma} \]


Multi-modal face recognition

- Close set identification
  - 4,249 gallery images, 596 probe subjects

<table>
<thead>
<tr>
<th># images or videos per subj.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7+</th>
</tr>
</thead>
<tbody>
<tr>
<td># subjects (images in LFW)</td>
<td>238</td>
<td>110</td>
<td>78</td>
<td>57</td>
<td>25</td>
<td>12</td>
<td>76</td>
</tr>
<tr>
<td># subjects (videos in YTF)</td>
<td>204</td>
<td>190</td>
<td>122</td>
<td>60</td>
<td>18</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

Images

Videos

Multi-modal

2016/4/6 hanhu@ict.ac.cn
Multi-modal face recognition

- Close set identification
  - 4,249 gallery images + 1M background mugshots, 596 probe subjects
Multi-modal face recognition

- Open set identification
  - The person of interest may not be present in legacy face databases
  - The gallery consists of 596 subjects with at least two images in the LFW database and at least one video in the YTF database

![Graphs showing detection and identification rate vs false alarm rate](image)

(a) Probe: Single Image
(b) Probe: Single Video Track
Multi-modal face recognition

- Quality based fusion

(a) Quality values: 0.78, 0.87
   SUM rule rank: 286
   QBF rule rank: 163

(b) Quality values: 0.42, 0.98
   SUM rule rank: 95
   QBF rule rank: 150

(c) Quality values: 0.78, 0.30
   SUM rule rank: 202
   QBF rule rank: 149

(d) Quality values: 0.46, 0.99
   SUM rule rank: 7
   QBF rule rank: 1

(e) Quality values: 0.78, 0.67
   SUM rule rank: 124
   QBF rule rank: 243
Multi-modal face recognition

- A Case Study on the Boston Bomber
  (Gallery of one million mugshots)

Race: White
Gender: Male
Age: 20 to 30

(a) Without Demographic Filtering

<table>
<thead>
<tr>
<th></th>
<th>( l_a )</th>
<th>( l_b )</th>
<th>( l_c )</th>
<th>max</th>
<th>sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>( l_x )</td>
<td>117,322</td>
<td>475,769</td>
<td>8,285</td>
<td>18,710</td>
<td>27,673</td>
</tr>
<tr>
<td>( l_y )</td>
<td>12,444</td>
<td>440,870</td>
<td>63,313</td>
<td>38,298</td>
<td>28,169</td>
</tr>
<tr>
<td>( l_z )</td>
<td>87,803</td>
<td>237,704</td>
<td>53,771</td>
<td>143,389</td>
<td>55,712</td>
</tr>
<tr>
<td>max</td>
<td>9,409</td>
<td>117,623</td>
<td>6,259</td>
<td>14,977</td>
<td>6,281</td>
</tr>
<tr>
<td>mean</td>
<td>13,658</td>
<td>125,117</td>
<td>8,019</td>
<td>20,614</td>
<td>8,986</td>
</tr>
</tbody>
</table>

(b) With Demographic Filtering (white male, 20-30)

<table>
<thead>
<tr>
<th></th>
<th>( l_a )</th>
<th>( l_b )</th>
<th>( l_c )</th>
<th>max</th>
<th>sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>( l_x )</td>
<td>5,432</td>
<td>27,617</td>
<td>112</td>
<td>114</td>
<td>353</td>
</tr>
<tr>
<td>( l_y )</td>
<td>518</td>
<td>25,780</td>
<td>1,409</td>
<td>1,656</td>
<td>686</td>
</tr>
<tr>
<td>( l_z )</td>
<td>3,958</td>
<td>14,670</td>
<td>1,142</td>
<td>2,627</td>
<td>1,416</td>
</tr>
<tr>
<td>max</td>
<td>374</td>
<td>6,153</td>
<td>94</td>
<td>109</td>
<td>106</td>
</tr>
<tr>
<td>mean</td>
<td>424</td>
<td>5,790</td>
<td>71</td>
<td>109</td>
<td>82</td>
</tr>
</tbody>
</table>
Multi-modal face recognition

- Forensic Sketches from Low Quality Video

Retrieval ranks without and with demographic filtering are given as #(#).
Deep multi-modal FR
Deep multi-modal FR

- Deep RGBD face recognition
  - 900,000 RGBD images of 700 subjects

<table>
<thead>
<tr>
<th>Modality</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>RGB</td>
<td>0.93</td>
</tr>
<tr>
<td>Depth</td>
<td>0.86</td>
</tr>
<tr>
<td>Deep RGB-D</td>
<td>0.98</td>
</tr>
</tbody>
</table>
Outline

- Background
- Related work
- Multi-modal & cross-modal FR
  - Cross-modal FR
- Trend on multi-modal (face) recognition
- Conclusion and discussion
Cross-modal face recognition

- Compatible with huge existing 2D face images
  - RGBD vs. RGB
- Modality is not available
  - NIR vs. VIS
  - Sketch vs. photo
Cross-modal face recognition

- Sketch to mugshot matching
  - Viewed sketch: drawing/synthesizing a sketch while looking at a subject/photo
  - Forensic sketch: drawing/synthesizing a sketch based on verbal description from the victim or eyewitness
  - COTS matcher for photo-to-photo matching can achieve over than 85% rank-1 identification rate for viewed sketch in 2013, while its performance for forensic sketch identification is less than 10%
Cross-modal face recognition

- Sketch to mugshot matching
Sketch to mugshot matching

- Sketch leads to arrest of suspects

Timothy McVeigh (the Oklahoma City bomber)  David Berkowitz (Son of Sam)  Ted Kaczynski (the Unabomber)
Sketch to mugshot matching

- Local and holistic matching
  - Local matching: component based rep.

Sketch to mugshot matching

- Component based rep. is an inverse process of sketch composition
Sketch to mugshot matching

- Local and holistic matching
  - Holistic matching: dense keypoint features

Sketch → Eye Detection → Alignment & CSDN → Patch-wise Segmentation → Dense Grid of Keypoints

MLBP → SIFT
Sketch to mugshot matching

- Complementarity

![Graph showing retrieval rates for different algorithms.]

- Algorithm:
  - Component-Based
  - COTS-1
  - COTS-2
  - COTS-3
  - Holistic
  - Holistic + Component-Based
# Sketch to mugshot matching

- **Hand-drawn and software-generated forensic sketch**

<table>
<thead>
<tr>
<th>Matcher</th>
<th>Testing Database</th>
<th>TAR @ FAR =</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>FaceSketchID</td>
<td>PRIP-HDC</td>
<td>0.255 ± 0.020</td>
<td>0.565 ± 0.028</td>
</tr>
<tr>
<td></td>
<td>PRIP-SGC</td>
<td>0.251</td>
<td>0.613</td>
</tr>
<tr>
<td>COTS-1</td>
<td>PRIP-HDC</td>
<td>0.122 ± 0.049</td>
<td>0.381 ± 0.047</td>
</tr>
<tr>
<td></td>
<td>PRIP-SGC</td>
<td>0.129</td>
<td>0.473</td>
</tr>
<tr>
<td>COTS-2</td>
<td>PRIP-HDC</td>
<td>0.115 ± 0.028</td>
<td>0.491 ± 0.255</td>
</tr>
<tr>
<td></td>
<td>PRIP-SGC</td>
<td>0.170</td>
<td>1.000</td>
</tr>
<tr>
<td>COTS-3</td>
<td>PRIP-HDC</td>
<td>0.064 ± 0.018</td>
<td>0.249 ± 0.050</td>
</tr>
<tr>
<td></td>
<td>PRIP-SGC</td>
<td>0.069</td>
<td>0.232</td>
</tr>
</tbody>
</table>
Sketch to mugshot matching

- Software-generated viewed sketch
Generalized cross-modal FR

- Cross-distance and cross-spectral matching in nighttime FR

150m NIR at night

Enrolled VIS

2016/4/6 hanhu@ict.ac.cn
Cross-distance and cross-spectral FR

- A learning based image restoration method to recover a high-quality face image from a low-quality image.
  - Dictionary building

- High-quality image (1m NIR)

- Low-quality image (150m NIR)

- Median filtering & normalization

- Random patches

- Dictionary of patches

- K-means clustering

- Clusters
Cross-distance and cross-spectral FR

- Per-patch recovery using LLE
Cross-distance and cross-spectral FR

- Restored images

Fig. 19. Examples of (a) correct matches and (b) incorrect matches by the proposed approach at 0.1% FAR in intra-spectral and cross-distance matching.
Cross-distance and cross-spectral FR

- Cross-distance and intra-spectral test

![Graphs showing true accept rate vs. false accept rate for different resolutions and spectral bands.]
Cross-distance and cross-spectral FR

Cross-distance and cross-spectral test
Sketch to mugshot matching

- Licensed to MorphoTrak, one of the world’s leading biometrics companies

- Nighttime FR system received funding from FBI
Outline

- Background
- Related work
- Multi-modal & cross-modal FR
  - Multi-modal FR
- Trend on multi-modal (face) recognition
- Conclusion and discussion
Trend on multi-modal (face) recognition

- The other biometrics or multi-biometrics
  - Tattoo, gesture,...
  - Google Abacus project

- General object
  - RGB-D, 300 common household objects
Trend on multi-modal (face) recognition

- The other biometrics or multi-biometrics
  - Tattoo, gesture,...
  - Google Abacus project

Masked ringleader of crowd trouble during Italy-Serbia clash identified by his tattooed arms [1].

Trend on multi-modal (face) recognition

- The other biometrics or multi-biometrics
  - Tattoo, gesture,...
  - Google Abacus Project (Google I/O 2015)
Trend on multi-modal (face) recognition

- The other biometrics or multi-biometrics
  - Tattoo, gesture,...
  - Google Abacus Project (Google I/O 2015)

3TB data

You are your password!
Outline

- Background
- Related work
- Multi-modal & cross-modal FR
  - Multi-modal FR
- Trend on multi-modal (face) recognition
- Conclusion and discussion
What I want to convey...

- Multi-modal FR significantly boosts the face recognition performance, particularly in unconstrained scenarios; but the optimum process pipelines of individual modalities and fusing scheme are still not known.
- Cross-modality FR, particularly forensic sketch recognition, has wide applications, but remains an open problem.
- Download
  - [Data] Still & video & sketch & 3D face images
  - [Data] Cross-distance, cross-spectral face images
  - [Data] Computer generated viewed-sketches
  - [Protocol] Open-set identification protocol
  - http://biometrics.cse.msu.edu/pubs/databases.html
Related papers


Thank You!

hanhu@ict.ac.cn
http://vipl.ict.ac.cn/members/hhan