Evaluation of Online Face Processing 2013: Methodology and Results
Background

- IBG contracted by USG agency to study online face processing – technologies, capabilities, performance
- Recently-completed study compares performance in 2011 and 2013
- Online face processing is the largest commercial use of biometrics, performance and capabilities have not been systematically analyzed
- Interested in identity-related risks, threats, and opportunities
2013 Performance Evaluation: Goals, Parameters

- Has online face processing performance improved since 2011?
- Does performance differ significantly across services?
- Services under evaluation: Facebook, Google+, and PittPatt
  - For Facebook, a relatively direct comparison can be made between 2011 and 2013 – fundamental face grouping workflow unchanged
  - Google+ replaces Picasa in 2013 evaluation
  - PittPatt replaces Face.com in 2013 evaluation
- Why PittPatt?
  - Blanket license for USG use
  - Designed for low-quality faces, such as those often uploaded to SNS
  - PittPatt clustering is analogous to SNS grouping functions
- Focus on grouping performance (false detections are rare)
Online Face Processing History and Timeline

- **2008**
  - September: Google adds FR to Picasa Web Albums
  - January: Apple releases iPhoto '09 with FR

- **2009**
  - September: Google releases Picasa 3.6 with FR

- **2010**
  - September: Facebook rolls out group face tagging

- **2011**
  - June: Facebook acquires Face.com
  - July: Google acquires PliPalt
  - December: Apple acquires Polar Rose

- **2012**
  - June: Facebook rolls out Tag Suggestions worldwide
  - July: Facebook suspends Tag Suggestions in US
  - October: Google acquires Viewdle
  - December: Apple issues patent for controlling mobile and desktop devices with FR

- **2013**
  - January: Facebook re-enables Tag Suggestions in US
  - February: Facebook de-lists face templates in EU

- **2014**
  - March: Facebook announces neural network-aided DeepFace algorithm
Challenge: Lack of Representative Images

- Civil / criminal face image datasets unrepresentative of social photos
  - Controlled capture conditions, or not “social”

- Open-source online face image datasets also not representative
  - Public figures, celebrities
  - Scanned images from newspapers
  - Do not reflect advances in smartphone camera capabilities

- What is “representative”?
  - Life events
  - Candid photos
  - Uncontrolled collection
  - Range of cameras and image formats

- Identity of individuals in photos unknown
SOCIAL-ID: Online Images for Evaluation, Testing

- **Sanitized Online Collection and Identity Analysis Library - Image Dataset**
- Face image dataset collected to evaluate online face processing site / service performance
- Images downloaded from [www.flickr.com](http://www.flickr.com) using the Flickr API
- For 2011 testing, >700K photos retrieved from >2K accounts
- 41 countries represented (most-represented: UK, IT, AU, ES, FR)
- Limited to non-US persons using account data, photo coordinates
Use, Selection of Flickr Accounts in SOCIAL-ID

- Photos retrieved for a given account holder referred to as a **dataset**
- Each account holder’s dataset is a discrete universe of people **separate from all other account holders’ datasets**
- **All matching and grouping in this study is intra-dataset**
- Datasets analyzed prior to testing to find “face-rich” datasets
  - IBG enrolled and searched datasets through Neurotechnology VeriLook
  - Each dataset searched **intra-set** to count detected faces and potential matches
- **Dataset selection criteria**
  - <500 photos (else excessive manual effort to review, process)
  - >25 detected faces (ensures that datasets are “face-rich” and warrant review)
  - 15 to 300 potential matches (we want to study datasets with potential matches)
SOCIAL-ID Characteristics

Distribution of Roll, Pitch, and Yaw

Number of Faces

Degrees

Copyright © 2014 IBG, A Novetta Solutions Company – 8
Automating SNS Uploads and Results Retrieval

- 2011 performance evaluation was extremely labor intensive
  - Uploads failed (i.e. face processing did not work), adjudication was real-time
- For 2013, IBG developed “pipeline” application to automate uploading photos to services, retrieving outputs, parsing results
  - Allows analysts to focus solely on adjudication as opposed to processing
- Pipeline automatically…
  - Uploads a set of images specified by the user (can process all 180 sets)
  - Saves html files with contain detection and grouping information
  - Extracts all relevant information (e.g. grouped faces) by parsing html files
  - Routes results to SOCIAL-GT adjudication interface for offline adjudication
- Enables iterative OFP performance assessment at frequent intervals
High-Level Processing for Facebook and Google+

Logs into SNS → Uploads SOCIAL-ID datasets → Parses face detection and grouping pages → Obtains face coordinates and groups from HTM files → Writes outputs to SOCIAL-GT

facebook

Google+

180 Datasets
37,778 Photos
47,048 Detected Faces
3,875 Grouped Faces
Automation of Google+ Face Processing (1 of 2)

- **Uploading photo albums**
  - Using macros that interact with Google+ UI, pipeline logs in & uploads album
  - “Done” button selected when available; Google+ begins grouping detected faces
  - Grouping page saved as a HTML file; User ID and album ID comprise an RSS feed accessible at a special URL
    
    https://picasaweb.google.com/data/feed/api/user/ [user ID] /albumid/ [album ID]

- **Parsing HTML results**
  - Pipeline uses Beautiful Soup Python library
  - Pipeline determines source photo for each detected face, extracts source path, coordinate, and group number
  - When a HTML page is saved on Firefox, Firefox creates a directory where images are automatically downloaded
  - Google+ coordinate format is x and y coordinates of the top left and bottom right corner of the detection
Automation of Google+ Face Processing (2 of 2)

- HTML tags denote the presentation of a group
- By searching for these tags and finding nested faces, pipeline assigns all faces in that group the same group number
- Group 0 represents set of faces not grouped with any other faces
- Other group numbers (1, 2, 3, etc.) assigned sequentially from top to bottom
Automation of Facebook Face Processing

- Multiple “Test User” accounts created to enable parallel testing
- Uploading photo albums
  - Using macros that interact with Facebook UI, pipeline logs in & uploads album
  - When title bar displays “Upload Complete”, Facebook saves upload page as an HTML file; pipeline clicks on Post Photos” to proceed to grouping page
  - Pipeline saves page such that grouping HTML has a URL for each face
- Parsing HTML results
  - To find the group number for each detected face, the pipeline searches through HTML tags to find which faces belong to each group
  - Group 0 represents ungrouped faces, other group numbers assigned sequentially by appearance
  - Pipeline parses URLs for photo filename and face coordinates; Facebook gives top left x and y coordinate, width, and height of the detection
PittPatt SDK Clustering Optimization

Level 0
No Grouping

Level 1
Minimal Grouping

Level 3
Moderate Grouping

Level 7
Aggressive Grouping
Analyst Adjudication Interface (SOCIAL-GT)

- Results for each service presented to analysts for offline adjudication
- For each dataset, results presented in SNS-defined groups
- Ungrouped images are also shown, enabling tabulation of false negative identifications (ungrouped primary subjects)
Categorizing Detected Faces through SOCIAL-GT
Detection of Partially Obscured Faces...
Correct Grouping with Various Angles/Expressions
Calculating Grouping Rates

- Find all groups in which Primary Subject face appears 1+ times
  - These are defined as “Primary Subject Groups”
- Count Primary Subject and Non-Primary Subject Faces

\[
\text{Correct Grouping Rate} = \frac{\text{Primary Subject Faces in Primary Subject Groups}}{\text{Total Faces in Primary Subject Groups}}
\]

- Example
  - Album with 120 photos and 375 faces is uploaded
  - Facebook creates 6 groups with 1+ Primary Subject faces; we ignore groups without Primary Subject faces
  - 24 faces are present in these 6 groups: 18 Primary Subject Faces and 6 Non-Primary Subject Faces (“impostors”)
  - Correct Grouping Rate for this dataset = 75% (18/24)
- Ungrouped Primary Subject faces are not part of this calculation – these are dealt with separately
Results: Correct Grouping Rates

Comparative Correct Grouping Rates

- PittPatt
- Facebook
- Google+

Correct Grouping Rate

Groups (Cumulative)

<table>
<thead>
<tr>
<th>Groups</th>
<th>Correct Grouping Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>95.64%</td>
</tr>
<tr>
<td>2</td>
<td>95.22%</td>
</tr>
<tr>
<td>3</td>
<td>95.13%</td>
</tr>
<tr>
<td>4</td>
<td>94.94%</td>
</tr>
<tr>
<td>5+</td>
<td>94.70%</td>
</tr>
</tbody>
</table>

94.66%  93.83%  93.51%  93.19%  93.38%

94.23%  93.40%  93.34%  92.95%  92.98%
Results: Facebook Grouping Rate

Facebook Correct Grouping Rates (Cumulative)

Correct Grouping Rate

94.67%  93.41%  93.35%  93.19%  93.39%
91.76%  89.95%  89.92%  89.58%  89.29%

Group

Copyright © 2014 IBG, A Novetta Solutions Company – 21
Results: Facebook Grouping (2011 vs. 2013)

Correct and Incorrect Facebook Matches by Group

- 2013 Correct Matches
- 2013 Incorrect Matches
- 2011 Correct Matches
- 2011 Incorrect Matches

Number of Grouped Faces

Group

1
2
3
4
5+

1047
891
48
16
128

28
186
86
50
109

59
80
7
6
18

Copyright © 2014 IBG, A Novetta Solutions Company – 22
Results: Face Grouping Performance

Face Grouping Performance

- **Google+**
  - Correctly Grouped PS Faces: 2107 (62.6%)
  - Ungrouped PS Faces: 1748
  - Incorrectly Grouped Faces: 1615 (48.0%)

- **PittPatt**
  - Correctly Grouped PS Faces: 1497 (44.5%)
  - Ungrouped PS Faces: 1866
  - Incorrectly Grouped Faces: 90

- **Facebook**
  - Correctly Grouped PS Faces: 106
  - Ungrouped PS Faces: 1256
  - Incorrectly Grouped Faces: 165
Results: Facebook Detection (2011 vs. 2013)

- In 180 datasets in 2013, Facebook detected 24,373 live faces
- In 180 datasets in 2011, Facebook detected 23,208 live faces
- An increase of 1165 (5.02%) in 2013
SOCIAL-ID 2013: Expanded Retrieval

- New SOCIAL-ID software suite consists of command-line Java applications that retrieve images from Facebook, Flickr, and Twitter.
- Acquired photos are processed with Neurotechnology VeriLook to provide a rough estimation of face content.
- Applications run in either ‘new’ or ‘update’ mode.

<table>
<thead>
<tr>
<th></th>
<th>Images with Face(s)</th>
<th>Faces Detected</th>
<th>Matches within User Accounts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Twitter</td>
<td>1,392,400</td>
<td>2,047,775</td>
<td>786,842</td>
</tr>
<tr>
<td>Instagram</td>
<td>1,288,493</td>
<td>1,951,107</td>
<td>4,054,018</td>
</tr>
<tr>
<td>Facebook</td>
<td>322,861</td>
<td>365,162</td>
<td>1,129</td>
</tr>
<tr>
<td>Flickr</td>
<td>201,269</td>
<td>286,310</td>
<td>3,047,754</td>
</tr>
<tr>
<td>Twitpic</td>
<td>127,913</td>
<td>172,552</td>
<td>642,682</td>
</tr>
<tr>
<td>yFrog</td>
<td>48,851</td>
<td>67,273</td>
<td>8,213</td>
</tr>
</tbody>
</table>
Conclusions and Future Work

- Online face processing performance evaluation can be partially automated
- ~93-95% of SNS groupings are correct
  - That is, 5-7% of faces are grouped incorrectly
- Google+ grouped ~600 more faces than Facebook out of ~3300 possible faces
- PittPatt can be used to roughly approximate SNS performance by using different clustering strengths

Other areas of interest
- Expanding to other demographic groups (to date, mostly Caucasian or Asian)
- Processing images with large inter-eye distances
- Evaluating tagging performance
- Establishing recurring processing effort to identify SNS face processing upgrades
- Using PittPatt processing to emulate Facebook, Google+ performance

For more information: mthieme@novetta.com