# IARPA Janus Benchmark-B Face Dataset

May 15, 2017

## Contents

# 1 Legal Notice

IARPA Janus Benchmark-B (IJB-B) (the Database), is made available under different Creative Commons license variants found in `ijbb_licenses_and_sources.xlsx` A Creative Commons (CC) license is one of several public copyright licenses that enable the free distribution of an otherwise copyrighted work. Should one or more samples be alternatively licensed then the owner or legal representative should contact NIST via FaceChallenges@nist.gov and identify particular images.

The Database consists of still images, frames and videos of celebrities and Internet personalities collected from the web. There are 1845 subjects in the Database and the Database is designed to have no overlap with the popular face recognition benchmarks, such as University of Oxfords VGG-Face dataset and the CASIA WebFace dataset.

Users and prospective users of the database will:

1. Comply with the licensing terms of the Database detailed in `ijbb_licenses_and_sources.xlsx` and `https://creativecommons.org/licenses/`
2. Comply with best practices for attribution found here: `https://wiki.creativecommons.org/wiki/best_practices_for_attribution`
3. Note binding licensing terms at https://creativecommons.org/licenses/
4. Understand that the Database is subject to change at any time.
5. Understand that U.S. Government makes no guarantee or warranty of non-infringement with respect to the Database.
6. Note that the U.S. Government is relying on the stated and represented due diligence of contractors who in part prepared this Database.

# 2 Notes

- An EXIF parsing bug exists in OpenCV 3.1.0 which causes `imread()` to hang indefinitely while decoding some JPEG images. See: `https://github.com/opencv/opencv/pull/6800`
- Face detection protocol now includes an ignore flag. This flag is set to 1 if a Turker noted any errors (extra box, missing box, incorrect box) with the bounding box(es) for a given piece of media.
- All coordinates are integers

# 3 Directory Structure

```
IJB-B/
|-- cluster_eval/
|   |-- association_modified_bcubed.py
|   `-- modified_bcubed.py
|-- frames/
|-- ijbb_licenses_and_sources.xlsx
|-- img/
|-- nonfaces/
|-- protocol/
|   |-- cluster/
|   |   |-- test7/
|   |   |   |-- ijbb_clustering_1024_hint_10000.csv
|   |   |   |-- ijbb_clustering_128_hint_1000.csv
|   |   |   |-- ijbb_clustering_1845_hint_10000.csv
|   |   |   |-- ijbb_clustering_256_hint_1000.csv
|   |   |   |-- ijbb_clustering_32_hint_100.csv
|   |   |   |-- ijbb_clustering_512_hint_1000.csv
|   |   |   `-- ijbb_clustering_64_hint_100.csv
|   |   `-- test8/
|   |       |-- ijbb_detection_clustering_ground_truth.csv
|   |       `-- ijbb_detection_clustering_hint_100000.csv
|   |-- ijbb_11_covariate_matches.csv
|   |-- ijbb_11_covariate_probe_reference_metadata.csv
|   |-- ijbb_11_S1_S2_matches.csv
|   |-- ijbb_1N_gallery_S1.csv
|   |-- ijbb_1N_gallery_S2.csv
|   |-- ijbb_1N_probe_img.csv
|   |-- ijbb_1N_probe_mixed.csv
|   |-- ijbb_1N_probe_video.csv
|   |-- ijbb_face_detection.csv
|   |-- ijbb_metadata.csv
|   `-- ijbb_subject_names.csv
|-- README/
|-- README.pdf
`-- video/
```

# 4 Dataset Summary

## 4.1 Media Statistics

| Media Type | Count |
|---|---|
| Images | 11754 |
| Frames | 55026 |
| Videos | 7011 |
| Nonfaces | 10044 |

Table 1: Media Counts

| Test | Mode | Purpose | "Curation" |
|---|---|---|---|
| 1 | 1:1 Verification | Templates are comprised of mixed media (frames and stills) | All templates |
| 2 | 1:1 Covariate Verification | Single still-image comparisons to perform failure analysis. *e.g. yaw difference* | All templates |
| 3 | 1:N Still Image Search | Open set 1:N protocol using only still images as probe | Probe and G1/G2 |
| 4 | 1:N Mixed Search | Open set 1:N protocol using mixed media (frames, stills) as probe | G1/G2 |
| 5 | 1:N Video Search | Open set 1:N protocol using full motion video as probe. Frame and frame index with bounding box of the first occurrence of the subject of interest provided as input | G1/G2 |
| 6 | Face Detection | Pure detection, without recognition. Includes pile of non-face imagery | – |
| 7 | Clustering | Cluster cropped faces. Bounding box and landmark annotations are provided to delineate the subject of interest. | – |
| 8 | Detection and Clustering | Same as 7 but requires bounding box association to identify the subjects of interest. Detect all faces in a pile of media and cluster them accordingly. This test also includes a set of non-face media as distractors. | – |

Table 2: Summary of Protocol Descriptions. **NOTE:** *Galleries G1 and G2 are shared between tests 3-5.*

## 4.2 Protocol Descriptions

#### 4.2.1 Test 1: 1:1 Verification

*See Table 2.*

#### 4.2.2 Test 2: 1:1 Covariate Verification

*See Table 2.*

#### 4.2.3 Test 3: 1:N Still Image Search

*See Table 2.*

#### 4.2.4 Test 4: 1:N Mixed Search

*See Table 2.*

#### 4.2.5 Test 5: 1:N Video Search

*See Table 2.*

#### 4.2.6 Test 6: Face Detection

*See Table 2.*

#### 4.2.7 Test 7: Clustering

The clustering protocol is designed to test an algorithm's ability to identify multiple instances of the same subject from a collection of various pieces of media. In experiment 7, there are seven clustering sub-protocols, each with a different number of subjects and hint. See Table 3.

| **Subjects** | 32 | 64 | 128 | 256 | 512 | 1024 | 1845 |
|---|---|---|---|---|---|---|---|
| **Hints** | 100 | 100 | 1000 | 1000 | 1000 | 10000 | 10000 |

Table 3: Test 7 Sub-Protocol Subjects and Hints

The hints serve as a coarse upper bound on the number of true clusters within each sub-protocol. The hint is calculated by the following equation: $10^{\lceil \log_{10} |\text{subjects}| \rceil}$. For each sub-protocol, all imagery for each selected subject in IJB-B is used (still images and video frames) and is a superset of the previous sub-protocol (e.g., Clustering-64 contains all media from Clustering-32 plus media from 32 additional subjects). The input in this protocol is a collection of face images and bounding boxes that delineate the face of interest. The faces are then treated as individual items to be clustered. If an FTE occurs, the cluster number should be set to -1. BCubed [1] precision and recall, and F-measure are used as the primary evaluation metric. FTEs, although not clustered, still count towards evaluation with 0 precision and 0 recall.

#### 4.2.8 Test 8: Detection and Clustering

The detection and clustering protocol is designed to test an algorithm's ability to detect faces in an image and cluster all instances of the same subject together. The Clustering-1845 sub-protocol is used for this evaluation.

A template ID and bounding box for each detected face in a piece of media shall be returned. Note that the non-face imagery used in the IJB-B face detection protocol will also be used as a distractor set. After clustering, associations (by calculating bounding box overlap with ground truth bounding boxes) are assigned for each piece of media as either an IJB-B subject ID or a background subject ID. **All faces of IJB-B Subjects of Interest that were not detected will be treated as FTEs and evaluated in the same manner as listed in Experiment 7.** BCubed precision and recall, and F-measure will be used as the primary evaluation metric but modified as described below.

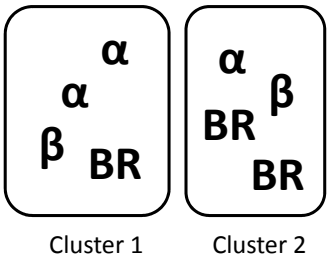


Figure 1: Example Clusters

- Precision is calculated only for associated events that have a IJB-B subject ID, but background subjects will be included in the calculation of those precision events. Consider Figure 1. Let Cluster 1 contain two faces of Hillary Clinton ($\alpha$), a face of Donald Trump ($\beta$), and a background face (BR). Further, let Cluster 2 contain one face of Hillary Clinton, one face of Donald Trump, and two background faces. The overall precision for the two clusters in this scenario is computed as:

$$\frac{\left(\frac{2}{4} + \frac{2}{4} + \frac{1}{4}\right) + \left(\frac{1}{4} + \frac{1}{4}\right)}{5} = 0.35$$

- Recall will be calculated only for associated events that have a IJB-B subject ID. For the preceeding example, the Recall for the sample clustering including Hillary Clinton ($\alpha$) and Donald Trump ($\beta$) is

computed as (*Note that all recalls for background events (BR) are ignored.*):

$$\frac{\frac{2}{3}+\frac{2}{3}+\frac{1}{2})+\ \ \frac{1}{3}+\frac{1}{2})}{5} \approx 0.533$$

## 4.3 Metadata Description

Attributes collected for images in the Janus program were annotated using crowd sourcing. As a result, attributes are defined in layman's terms, instead of strict scientific definitions. For more information, see `http://libjanus.org`

| **0-19** | **20-34** | **35-49** | **50-64** | **65+** | **Unknown** | **No Annotation** |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 0 | NaN |

Table 4: Age Enumeration

| **Outdoor** | **Indoor** | **No Annotation** |
|---|---|---|
| 0 | 1 | NaN |

Table 5: Indoor/Outdoor

| | |
|---|---|
| **Light Pink** | 1 |
| **Light Yellow** | 2 |
| **Medium Pink/Brown** | 3 |
| **Medium Yellow/Brown** | 4 |
| **Medium-Dark Brown** | 5 |
| **Dark Brown** | 6 |
| **No Annotation** | NaN |

Table 6: Skin Tone Enumeration

| **No Facial Hair** | **Moustache** | **Goatee** | **Beard** | **No Annotation** |
|---|---|---|---|---|
| 0 | 1 | 2 | 3 | NaN |

Table 7: Facial Hair Enumeration

| **Female** | **Male** | **No Annotation** |
|---|---|---|
| 0 | 1 | NaN |

Table 8: Gender

| Roll | | Yaw | |
|---|---|---|---|
| **Min** | **Max** | **Min** | **Max** |
| $-53°$ | $61°$ | $-87°$ | $78°$ |

Table 9: Face Orientation

# 5 Cluster Evaluation

Two scripts are provided for cluster evaluation in the `cluster_eval` directory which contain implementations of the BCubed evaluation metric relevant to IJB-B tests 7 and 8.

- `modified_bcubed.py`: This is used to evaluate clustering Test 7 (i.e. IJB-B subject of interests when given the face bounding box). The script can be called as such:

  `./modified_bcubed.py cluster_file ground_truth [-no_fte][-h]`

  - `cluster_file` is the output of Test 7 and should have a header of:
    `TEMPLATE_ID,FILENAME,CLUSTER_INDEX,CONFIDENCE`
  - `ground_truth` file contains the mapping between template IDs and filenames to subject IDs. It should have a header of:
    `TEMPLATE_ID,SUBJECT_ID,FILENAME`
  - `[-no_fte]` If set, all FTEs with a cluster assignment set to -1 will be ignored for all calculations.
  - `[-h]` Help

- `association_modified_bcubed.py`: This is used to evaluate clustering Test 8 (i.e. face detection and clustering of all imagery in IJB-B). The script can be called as such:

  `./association_modified_bcubed.py detection_cluster_file ground_truth [-no_fte][-h]`

  - `detection_cluster_file` is the output of clustering Test 8 and should have a header of:
    `TEMPLATE_ID,FILENAME,CLUSTER_INDEX,CONFIDENCE,FACE_X,FACE_Y,FACE_WIDTH,FACE_HEIGHT`
  - `ground_truth` associates the detected bounding boxes to that of IJB-B Subjects of Interest (SOI). It should have a header of:
    `SUBJECT_ID,FILENAME,FACE_X,FACE_Y,FACE_WIDTH,FACE_HEIGHT`
  - `[-no_fte]` If set, all FTEs with a cluster assignment set to -1 will be ignored for all calculations.
  - `[-h]` Help

The provided implementation of the BCubed metric has been modified according to the following guidelines:

- For test 7, any face that is a Failure To Enroll (FTE) will be counted as an event and evaluated with 0 precision and 0 recall
- For test 8, any IJB-B SOI event that fails to have an associated bounding box will be evaluated in the same manner as an FTE described above
- For test 8, any bounding box that has been associated with a IJB-B SOI, but fails to enroll will be evaluated as a FTE described above
- For test 8, any bounding box that is associated as a Background subject (BR) will be treated in accordance with the following criteria:
  - BR events clustered with IJB-B SOIs are used in the calculation of precision
  - Pure BR events will be ignored completely during evaluation

# References

[1] Enrique Amigo, Julio Gonzalo, Javier Artiles, and Felisa Verdejo. A comparison of extrinsic clustering evaluation metrics based on formal constraints. *Inf. Retr.*, 12(4):461–486, August 2009.