IARPA Janus Benchmark-B Face Dataset

Cameron Whitelam, Emma Taborsky*, Austin Blanton, Brianna Maze*, Jocelyn Adams*, Tim Miller*, Nathan Kalka*, Anil K. Jain, James A. Duncan*, Kristen Allen, Jordan Cheney*, Patrick Grother*****

CVPR July 21-26 2017 HONOLULU

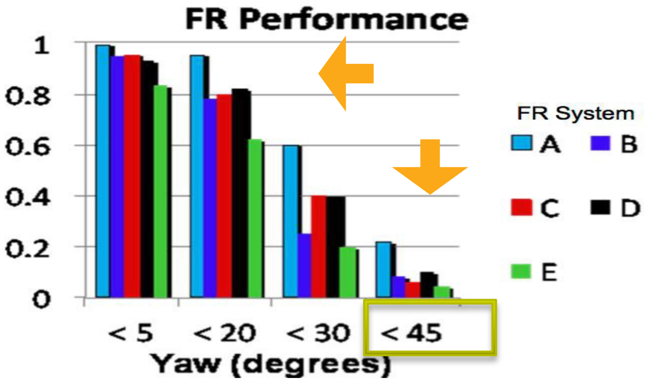
21 July 2017

Noblis*
Michigan State University**
NIST***

noblis MICHIGAN STATE UNIVERSITY NIST

State-of-the-Art & Limitations of Face Recognition

- COTS face recognition algorithms perform best on well-posed, frontal facial photos taken for identification purposes
  - Janus focuses on full range of roll, pitch, and yaw
- Face recognition performance is brittle with respect to factors such as Age, Pose, Illumination & Expression (A-PIE)



Negative impact on performance when a single factor, such as yaw is Changed. -- *NIST Multiple Biometric Evaluation (MBE) 2010*

Levels of Difficulty in Face Recognition

Roberto Stuckert Filho/PR

Agência Brasil

Roberto Stuckert Filho/PR

Image Type	**Frontal, Cooperative, Controlled**	**Near frontal, uncooperative**	**Full variation in pose, illumination, environment, uncooperative**
Face Detection performance	Human	Near human	Limited
Automated FR performance	Human	Near human	Limited

© 2017 Noblis, Inc.

Limitations of Prior Datasets

Previous datasets do not meet the requirements to push state of the art in unconstrained face recognition

- "Media in the Wild" datasets ushered in a new era of algorithmic approaches but were quickly saturated
  - E.g., LFW, PubFig, YTF
- These datasets are limited by at least one of the following factors:
  - Subjects were located using a commodity face detector
  - Lack of media type diversity
  - Lack of geographic diversity
  - No clear legal authority for redistribution of images with respect to data copyrights
  - Unlabeled subject identities (e.g. MegaFace)
  - No testing protocols, or only face detection protocols (e.g. WIDER FACE)
- IJB-A [1] addressed the above issues, but lacks the number of subjects to accurately test algorithms at low ends of the ROC curve



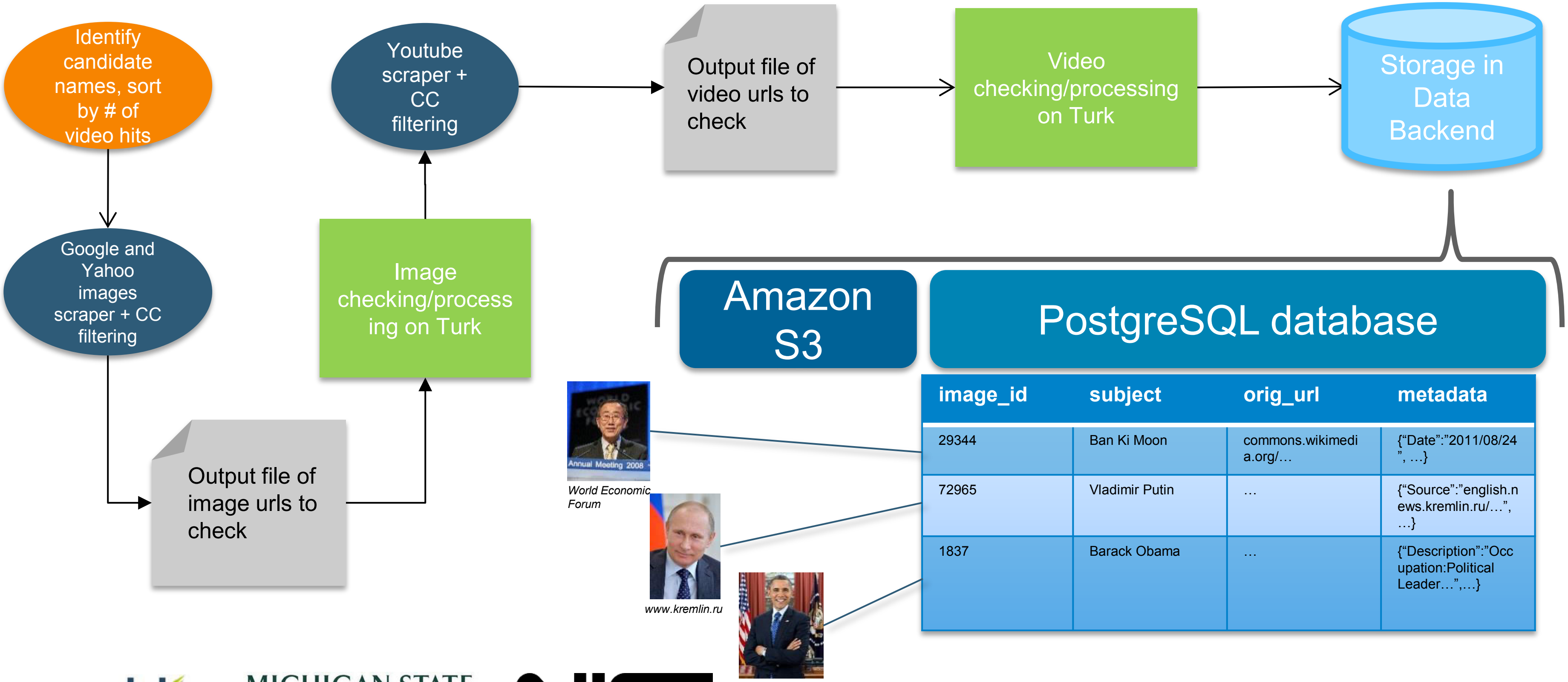

[1] B. F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, M. Burge and A. K. Jain, "Pushing the Frontiers of Unconstrained Face Detection and Recognition: IARPA Janus Benchmark A", CVPR, Boston, Massachusetts, June 8-10, 2015.

Images from:

Face recognition in unconstrained environments. Vol. 1. No. 2. Technical Report 07-49, University of Massachusetts, Amherst, 2007
Wolf, Lior, Tal Hassner, and Itay Maoz. "Face recognition in unconstrained videos with matched background similarity." CVPR, 2011.

© 2017 Noblis, Inc.

Data Development
Collection & Storage



image_id	subject	orig_url	metadata
29344	Ban Ki Moon	commons.wikimedia.org/…	{"Date":"2011/08/24", …}
72965	Vladimir Putin	…	{"Source":"english.news.kremlin.ru/…", …}
1837	Barack Obama	…	{"Description":"Occupation:Political Leader…",…}

noblis MICHIGAN STATE UNIVERSITY NIST

© 2017 Noblis, Inc.

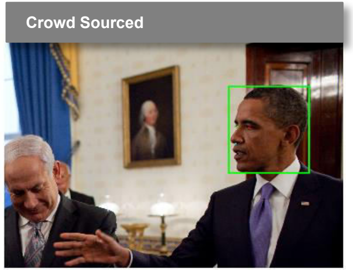
Data Development

Annotation Methodology

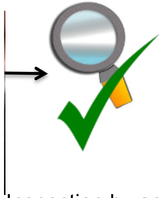
Start with subject imagery

Bounding box (BB) location for all faces

Select BB for POI



Label image and subject attributes



Inspection by analyst

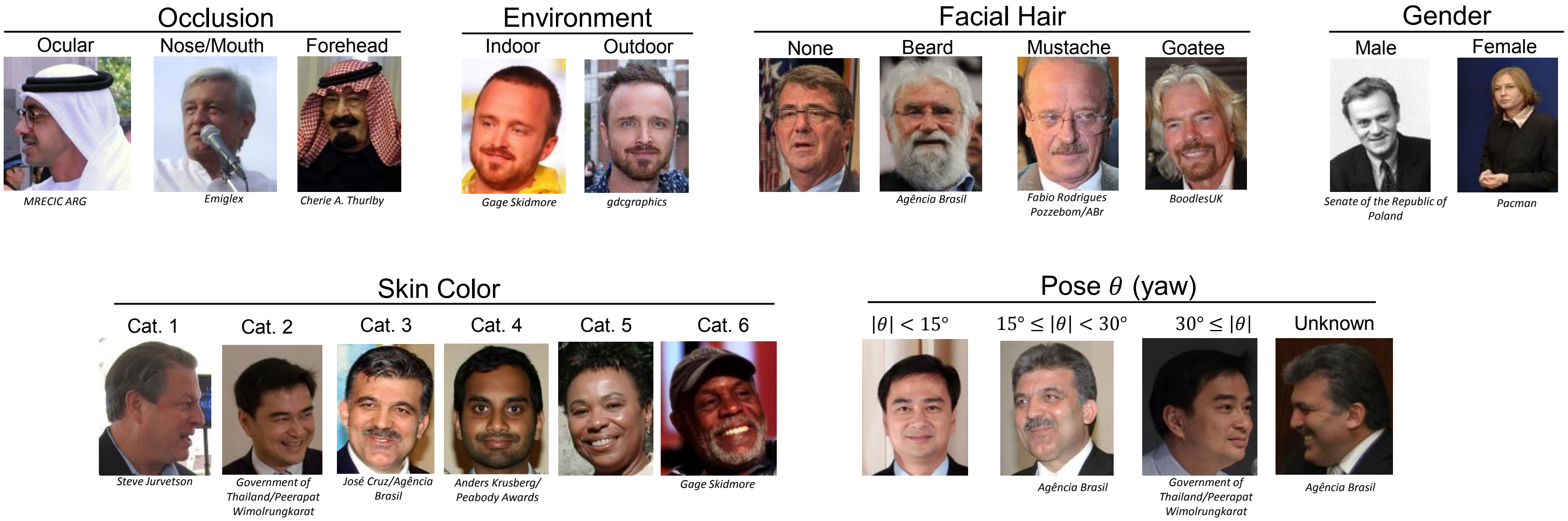
- Manual annotations were gathered using Amazon Mechanical Turk (AMT)
- AMT cannot be used “out of the box” for annotating geometric primitives (e.g., bounding boxes)

cc = Creative Commons

© 2017 Noblis, Inc.

Annotation Activities

Covariates examined/annotated across entire Janus imagery dataset (to date)

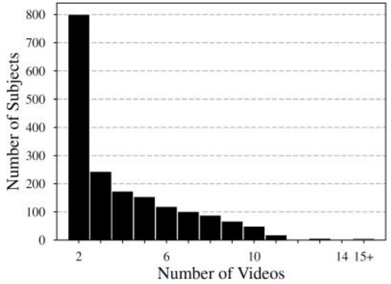


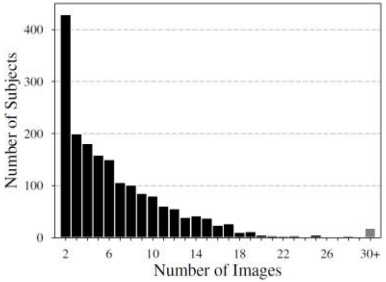
noblis MICHIGAN STATE UNIVERSITY NIST

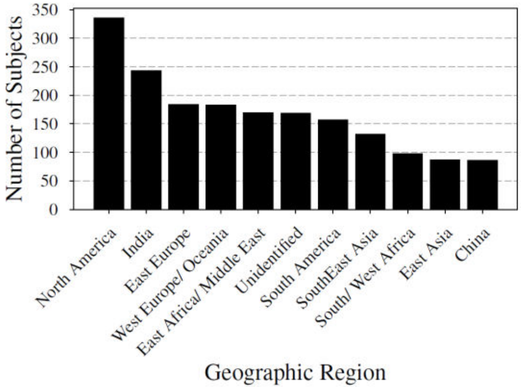
© 2017 Noblis, Inc.

IJB-B Overview

- Extension of IJB-A
    - Addition of 1,345 new subjects
- 11,754 still images – 7,011 videos – 1,845 subjects
    - 6.37 images & 3.8 videos on average per subject
    - 125,474 faces with an average ~2 faces per piece of media
    - 10,044 non-face images
- Over 3.8 million manual annotations to date
- Protocols supporting face detection, 1:1, 1:N, and clustering
- Improved geographic distribution
- Data licensed for redistribution
- Features non-frontal pose, heavy occlusion, and low resolution images







© 2017 Noblis, Inc.

IJB-B Example Imagery

Keith Allison

JK the Unwise

Mohamed CJ

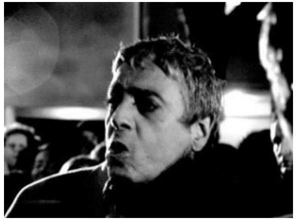
Pose

Archivos de Jota Linderos

Adrien Barbier

Illumination

akynou



Sven-Sebastian Sajak

Egyptian Chronicles

Expression

Andrew Baron

www.filmitadka.com

Xmrsmile

Occlusion

noblis MICHIGAN STATE UNIVERSITY NIST

© 2017 Noblis, Inc.

Test Protocols

Face Detection
- All images and video frames

1:N Identification
- Still images only
- Mixed images and frames
- Video

1:1 Verification
- Mixed images and frames (same templates used for 1:N mix)
- Covariate (single image/frame templates)

Clustering
- 7 sub-protocols:
  - 32 subjects
  - 64 subjects
  - 128 subjects
  - 256 subjects
  - 512 subjects
  - 1024 subjects
  - 1845 (all) subjects

Detection + Clustering
- 7 sub-protocols:
  - 32 subjects
  - 64 subjects
  - 128 subjects
  - 256 subjects
  - 512 subjects
  - 1024 subjects
  - 1845 (all) subjects

All 1:N probe templates are searched against disjoint gallery sets S1 and S2 to allow for open set analysis

© 2017 Noblis, Inc.

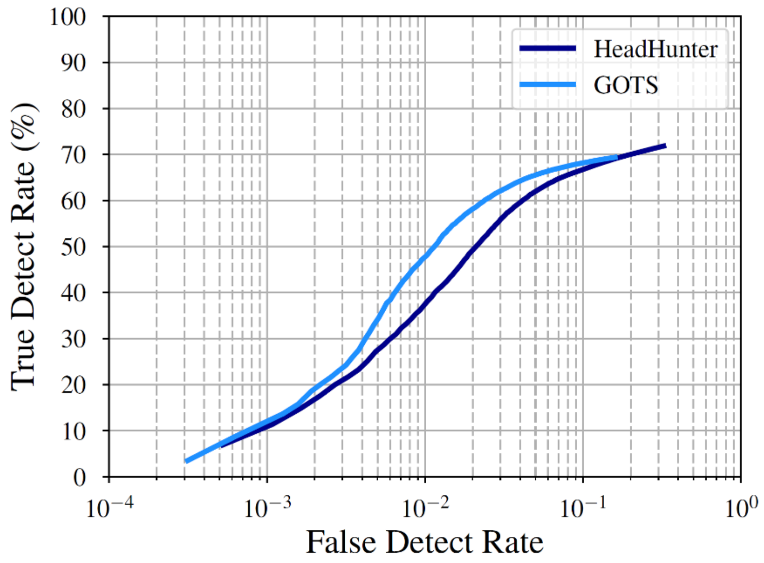
Face Detection

- **125,474 faces across 76,824 images and frames** (66,780 with faces and 10,044 non-face)
  - Protocol is augmented with 10,044 still images that do not contain any faces to test operationally relevant cases
  - Average ~2 faces per piece of media
- Evaluation metrics
  - ROC curve of True Detect Rate (TDR) and False Detect Rate (FDR)
  - *Extension metric* – average time to detect all faces in an image

© 2017 Noblis, Inc.

Baseline Results – Face Detection

- GOTS detector
  - Top performing detector in a recent face detection benchmark
  - Shown to achieve results similar to top published performers on FDDB
- Open-sourced HeadHunter
  - Specifically designed for unconstrained imagery
- GOTS performs ~10% better at FDR 10^{-2}



Face detection results on the IJB-B face detection protocol

© 2017 Noblis, Inc.

1:N Recognition

1:N Still	8,104 probe templates extracted from 5,732 still images
1:N Mixed	10,270 probe templates containing 60,758 still images and video frames
1:N Video	7,110 probe templates

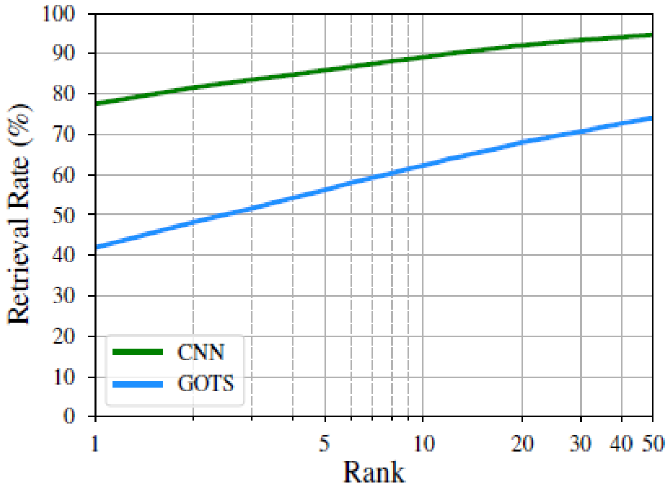
Evaluation Metrics

- CMC – measures closed set performance
- DET/IET – measures open set performance
- *Extension metric* – Mean compute time of template generation and probe searches

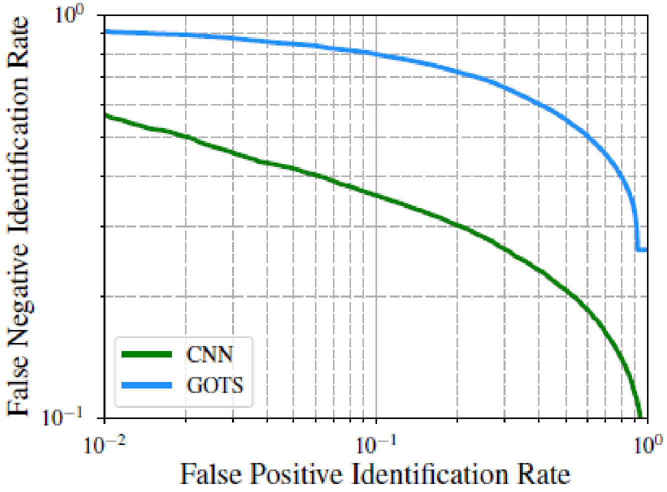
© 2017 Noblis, Inc.

Baseline Results – 1:N Mixed Media

- The CNN benchmark used an open-source, state-of-the-art model trained on the VGG face dataset



Average CMC performance across gallery sets S1 and S2



Average IET performance across gallery sets S1 and S2

© 2017 Noblis, Inc.

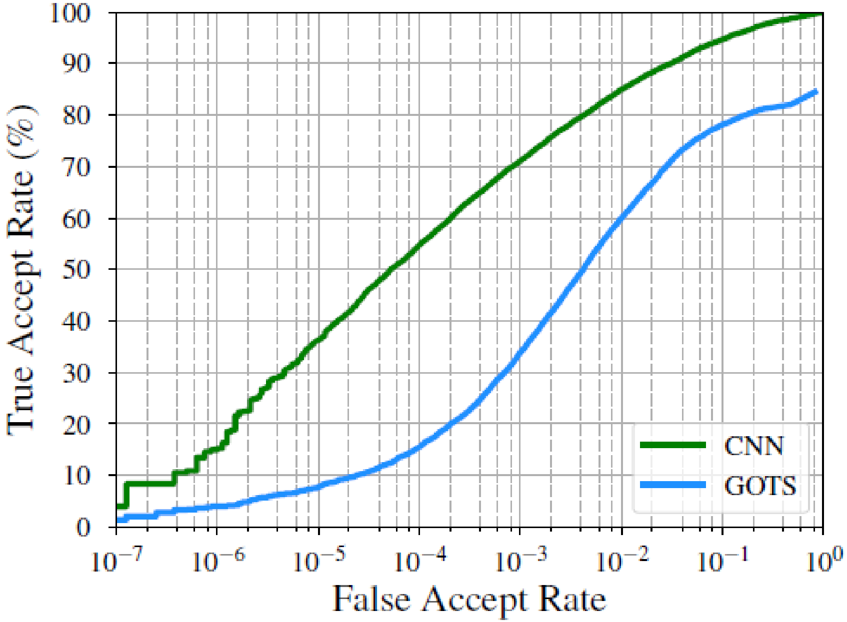
1:1 Verification

1:1 Baseline	***10,270 genuine*** comparisons and ***8,000,000 impostor*** comparisons
1:1 Covariate	***3,867,417 genuine*** and ***16,402,860 impostor*** comparisons between ***single image gallery and probe templates***

- Evaluation Metrics
  - Receiver Operating Characteristic (ROC) – metrics of TAR at a FAR of 10^{-2} and 10^{-4}
  - *Extension metric* – Mean duration of template generation and comparisons

© 2017 Noblis, Inc.

Baseline Results – 1:1 Verification



ROC constructed across 1:1 baseline matches

© 2017 Noblis, Inc.

Clustering

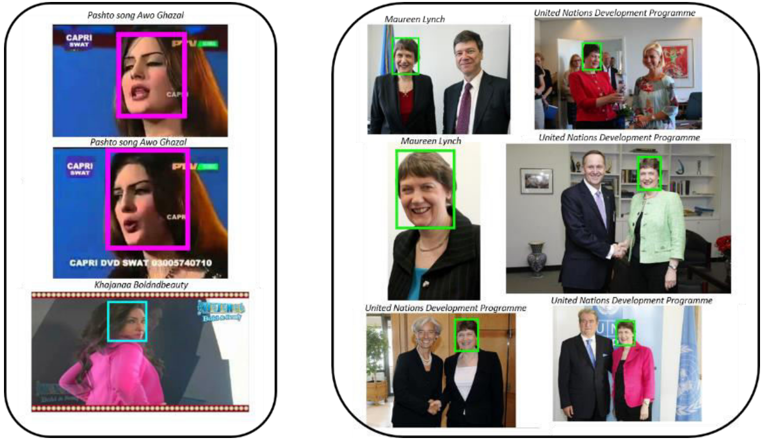
- Seven sub-protocols that test an algorithm's ability to cluster at different scales
  - All imagery for each selected subject is used and is a superset of the previous sub-protocol
  - Input to the clustering protocol is an image and a bounding box
  - A hint is provided for each sub-protocol

$$10^{\lceil \log_{10} |subjects| \rceil}$$

- Evaluation Metrics
  - BCubed Precision and Recall
    - Accounts for normal precision/recall edge cases by averaging across items
  - F-measure

© 2017 Noblis, Inc.

Baseline Results – Clustering



Sample clustering results; unique subject identities are represented with different color bounding boxes

	Hint	Precision	Recall	F-measure	Run Time	Percent of FTEs
Clustering-32	100	0.589	0.298	0.395	0.32m	20.9
Clustering-64	100	0.578	0.302	0.396	1.15m	20.0
Clustering-128	1,000	0.605	0.352	0.445	7.23m	15.5
Clustering-256	1,000	0.581	0.362	0.446	25.12m	14.8
Clustering-512	1,000	0.516	0.328	0.401	76.92m	16.7
Clustering-1024	10,000	0.485	0.345	0.403	310.5m	15.5
Clustering-1845	10,000	NA				

GOTS performance on the IJB-B clustering protocols; Note that results for Clustering-1845 are not available due to memory constraints

© 2017 Noblis, Inc.

Questions?

IJB-B is available for download at http://nigos.nist.gov:8080/facechallenges/IJBB/

This research is based upon work supported by the Office of the Director of National Intelligence (ODNI) and the Intelligence Advanced Research Projects Activity (IARPA). The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

© 2017 Noblis, Inc.