Pushing the Frontiers of Unconstrained Face Detection and Recognition: IARPA Janus Benchmark A

Brendan F. Klare, Emma Taborsky , Austin Blanton , Jordan Cheney , Kristen Allen , Patrick Grother , Alan Mah , Anil K. Jain

The development of accurate and scalable unconstrained face recognition algorithms is a long term goal of the biometrics and computer vision communities. The term "unconstrained" implies a system can perform successful identifications regardless of face image capture presentation (illumination, sensor, compression) or subject conditions (facial pose, expression, occlusion). While automatic, as well as human, face identification in certain scenarios may forever be elusive, such as when a face is heavily occluded or captured at very low resolutions, there still remains a large gap between automated systems and human performance on familiar faces. In order to close this gap, large annotated sets of imagery are needed that are representative of the end goals of unconstrained face recognition. This will help continue to push the frontiers of unconstrained face detection and recognition, which are the primary goals of the IARPA Janus program.

The current state of the art in unconstrained face recognition is high accuracy (roughly 99% true accept rate at a false accept rate of 1.0%) on faces that can be detected with a commodity face detectors, but unknown accuracy on other faces. Despite the fact that face detection and recognition research generally has advanced somewhat independently, the frontal face detector filtering approach used for key in the wild face recognition datasets means that progress in face recognition is currently hampered by progress in face detection. Hence, a major need exists for a face recognition dataset that captures as wide of a range of variations as possible to offer challenges to both face detection as well as face recognition.

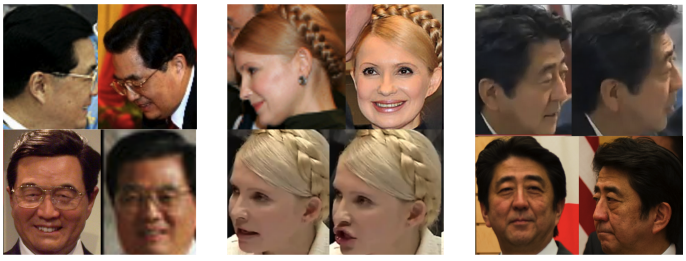
In this paper we introduce the IARPA Janus Benchmark A (IJB-A), which is publicly available for download. The IJB-A contains images and videos from 500 subjects captured from "in the wild" environment. All labelled subjects have been manually localized with bounding boxes for face detection, as well as fiducial landmarks for the center of the two eyes (if visible) and base of the nose. Manual bounding box annotations for all non-labelled subjects (i.e., other persons captured in the imagery) have been captured as well. All imagery is Creative Commons licensed, which is a license that allows open re-distribution provided proper attribution is made to the data creator. The subjects have been intentionally sampled to contain wider geographic distribution than previous datasets. Recognition and detection protocols are provided which are motivated by operational deployments of face recognition systems. An example of images and video from IJB-A can be found in Figure 3.

The IJB-A dataset has the following claimed contributions: (i) The most unconstrained database released to date; (ii) The first joint face detection and face recognition benchmark dataset collected in the wild; (iii) Meta-data providing subject gender and skin color, and occlusion (eyes, mouth/nose, and forehead), facial hear, and coarse pose information for each imagery instance; (iv) Widest geographic distribution of any public face dataset; (v) The first in the wild dataset to contain a mixture of images and videos; (vi) Clear authority for re-distribution; (vii) Protocols for identification (search) and verification (compare); (viii) Baseline accuracies from off the shelf detectors and recognition algorithms; and (ix) Protocols for both template and model-based face recognition.

Every subject in the dataset contains at least five images and one video. IJB-A consists of a total of 5,712 images and 2,085 videos, with an average of 11.4 images and 4.2 videos per subject.

	(a)	(b)	(c)
Types of Images:	Frontal, Cooperative subject, Controlled environment	Near frontal, uncooperative, minimal environment variations (e.g., LFW)	Full variation in pose, illumination, environment
Automated detection ability:	Human performance	Near human performance	Cannot detect consistently
Automated recognition ability:	Human performance	Near human performance	Cannot recognize

Figure 1: (a) Face recognition accuracy on frontal images is considered to be a nearly solved problem. (b) Accuracy has greatly improved on face images captured in unconstrained settings that can be detected with a commodity face detector. (c) Example of an image in the IARPA Janus Benchmark A; this is believed to be the most unconstrained face dataset to date.



(a) (b) (c)

Figure 2: Examples of the faces in the IJB-A dataset. These images and video frames highlight many of the key characteristics of this publicly available dataset, including full pose variation, a mixture of images and videos, and a wide variation in imaging conditions and geographic origin.

Dataset	# subjects	# images	# img/subj	# videos	# vid/subj
IJB-A	500	5,712	11.4	2,085	4.2
LFW	5,749	13,233	2.3	0	0
PubFig	200	58,797	294.0	0	0
YTF	1,595	0	0	3,425	2.1
PaSC	293	9,376	32	2,802	9.6

Figure 3: A comparison of key statistics of the proposed IJB-A dataset and seminal unconstrained face recognition datasets. The faces in IJB-A were manually localized and have full pose variation.

Protocol	Applications	Accuracy Metrics
Compare	1:1 match; Access control; Re-identification	TAR @ FAR of 0.1, 0.01, and 0.001; ROC plot (TAR vs. FAR)
Search	De-duplication; Watch list; Forensic	FNIR @ FPIR of 0.1 and 0.01; Rank 1 and 5 accuracy; CMC plot; DET plot (FNIR vs. FPIR)

Figure 4: Overview of the IJB-A recognition protocols. The search applications is measured using open-set identification.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.