NIST Technical Note 1993

NIST Special Database 300

Uncompressed Plain and Rolled Images from Fingerprint Cards

Gregory Fiumara
Patricia Flanagan
John Grantham
Bruce Bandini
Kenneth Ko
John Libert

This publication is available free of charge from:
https://doi.org/10.6028/NIST.TN.1993

National Institute of
Standards and Technology
U.S. Department of Commerce

NIST Technical Note 1993

NIST Special Database 300

Uncompressed Plain and Rolled Images from Fingerprint Cards

Gregory Fiumara
Patricia Flanagan
John Grantham
Bruce Bandini
Kenneth Ko
John Libert
Information Access Division
Information Technology Laboratory

This publication is available free of charge from:
https://doi.org/10.6028/NIST.TN.1993

June 2018
INCLUDES UPDATES AS OF 01-10-2020; SEE APPENDIX A

U.S. Department of Commerce
Wilbur L. Ross, Jr., Secretary

National Institute of Standards and Technology
Walter Copan, NIST Director and Under Secretary of Commerce for Standards and Technology

National Institute of Standards and Technology Technical Note 1993
Natl. Inst. Stand. Technol. Tech. Note 1993, 15 pages (June 2018)
CODEN: NTNOEF

This publication is available free of charge from:
https://doi.org/10.6028/NIST.TN.1993

Abstract

A new collection of legacy inked rolled and plain fingerprint card scans are being released to the public. The cards were scanned at three resolutions in the 8-bit grayscale color space. The data is available as lossless images, and can be freely downloaded on the National Institute of Standards and Technology's website.

Key words

biometrics; data; fingerprints.

Not Human Subjects Research

The National Institute of Standards and Technology Human Subjects Protection Office reviewed the protocol for this project and determined it used only specimens and/or data that met the criteria for *not human subjects research* as defined in Department of Commerce Regulations, 15 CFR 27, also known as the Common Rule (45 CFR 46, Subpart A), for the Protection of Human Subjects.

Acknowledgments

Thank you to the Federal Bureau of Investigation for their donation of these arrest cards and their willingness to fund the National Institute of Standards and Technology in order to provide digitized versions of these cards to the biometrics community.

Table of Contents

List of Tables

List of Figures

1. Introduction

Publicly-available and shareable data is critical to the biometrics research community. Such data makes it possible to validate biometric software packages and quantify algorithm performance in a reproducible manner. Public data is used to enable new technologies for biometric image compression, quality calculation, and matching performance. These tasks assist the biometric community, and are simply not possible without a free, shared repository of data.

The U.S. Government and governments around the world have millions of "legacy" biometric records, including inked fingerprints. Automated fingerprint matching algorithms in use in criminal applications today must continue to be able to match against legacy data. As such, new public datasets of legacy fingerprint data are important to fingerprint matching algorithm development.

Many public databases are collected from complicit participants in ideal environments with the latest technology. In most criminal applications, the subjects being fingerprinted are uncooperative and their fingerprints are captured on the roadside or in the harsh environment of a criminal intake facility. It's important that there be publicly-available biometric datasets representative of such scenarios.

With this in mind, the National Institute of Standards and Technology (NIST) has introduced a new dataset, Special Database (SD) 300, a dataset of digitized early 1900s Federal arrest inked fingerprint card segments. It is a collection of rolled and plain inked fingerprints from nearly 900 subjects, taken at arrest.

2. Acquisition

On behalf of the Federal Bureau of Investigation (FBI), 888 inked fingerprint arrest cards have been digitized. These cards arrived from the FBI in various physical conditions, from pristine to badly damaged and faded. These cards were operational in nature, meaning they were collected during a law enforcement professional's duties. Subjects were potentially non-cooperative and the collection environments were potentially harsh. All subjects whose biometrics appeared on these arrest cards were confirmed to be deceased by the FBI at the time of receipt by NIST.

3. Digitization

3.1 Scanning

All cards were digitized by members of the NIST Image Group. Each card was placed on the scanbed of an Epson Perfection 4990 or an Epson V700 photo scanner. Due to the fragility of the cards, automated document feeders were not used. After positioning the card in a representative corner of the scanbed, NIST-developed scanner software, NISTscan [1], was used under Windows 7 to automate scanning at three different resolutions: 196.85 pixels per centimeter (PPCM) (500 pixels per inch, or PPI), 393.7 PPCM (1000 PPI), and 787.4 PPCM (2000 PPI). All three resolutions were scanned in grayscale at a depth of 8 bits per pixel.

3.2 Card Segmentation

Although all ten-print cards came from the FBI, there was a mixture of card formats used. All cards were similar in nature to the blank card depicted in Fig. 1. In total, there are 35 distinct varieties of arrest card featured in SD 300. In order to protect the identities of the subjects in this dataset, metadata, such as names and arrest reasons written on the arrest card, was not permitted for release and was not retained after digitization. This meant that all fingerprints needed to be segmented away from the rest of the card. Up to 14 segments were chosen: 10 for each rolled print (ANSI/NIST-ITL [2] friction ridge generalized positions 1 to 10), 1 each for simultaneous right and left four-finger capture (friction ridge generalized positions 13 and 14), and 1 for each plain right and left thumb (friction ridge generalized positions 11 and 12). See Table 4 for a tabular version of friction ridge generalized positions mappings.

To accomplish this segmentation, a set of rectangular coordinates on the 196.85 PPCM (500 PPI) card scans were manually recorded for each variety of arrest card to create a "bounding box." As cards were scanned,

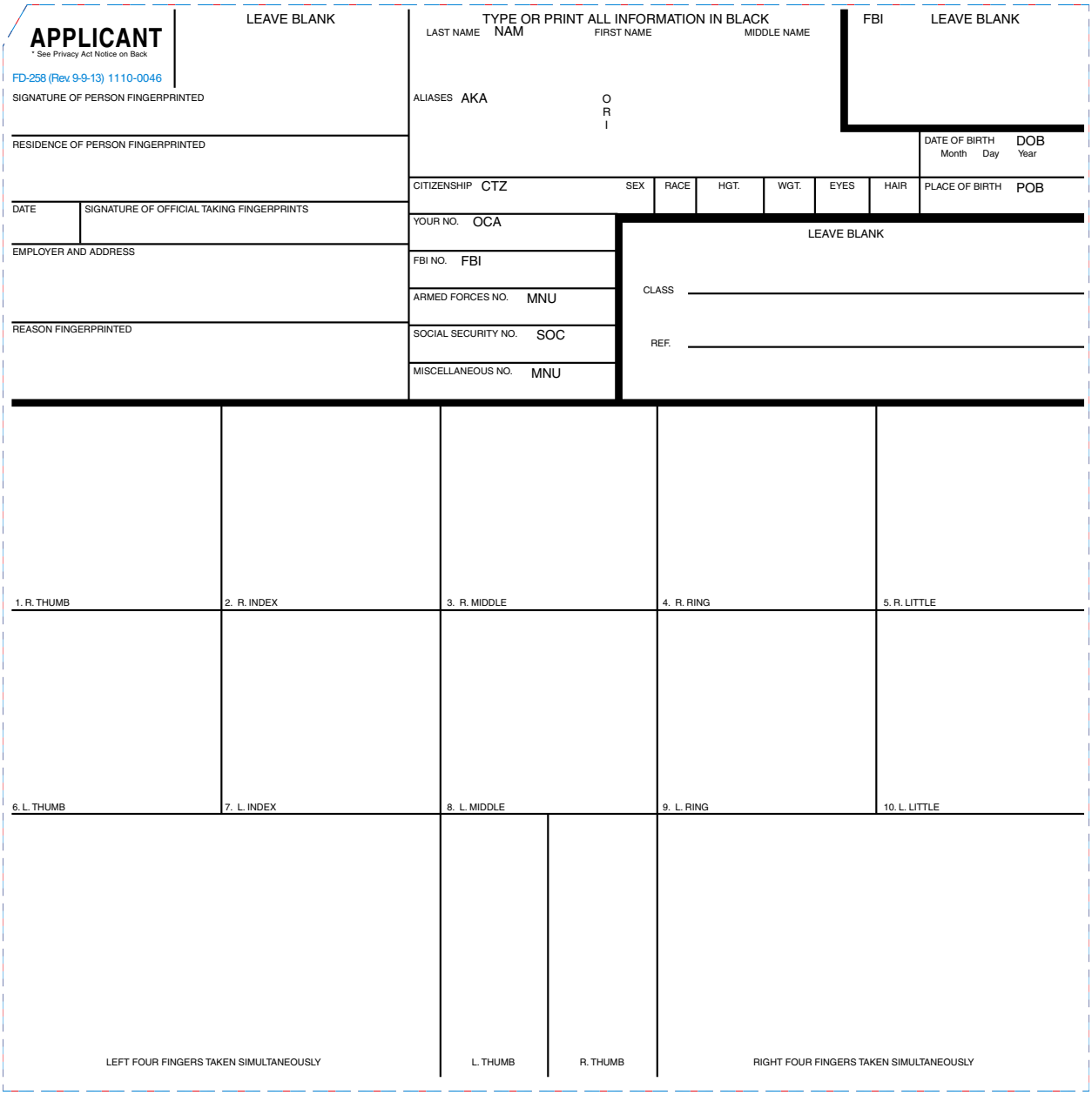

Fig. 1. A representative example of an ink fingerprint card [3], similar to those scanned to create this dataset. In total, there are 35 different varieties of fingerprint cards depicted in this dataset.

Trait	Value
Subjects	888
Rolled Images	8 871
Segmented Plain Images	8 787
Capture Card Varieties	35
Mean Rolled NFIQ 2.0	40
Mean Segmented Plain NFIQ 2.0	42

Table 1. An inventory of and some statistics about the data available in SD 300. The NIST Fingerprint Image Quality (NFIQ) 2.0 algorithm is not yet trained on rolled images, and as such, these NFIQ 2.0 rolled values are unofficial.

they were tagged with the type of arrest card. Once all cards were scanned, the 14 bounding boxes were losslessly sliced from the larger image containing the entire scan of the fingerprint card. These coordinates were then doubled and quadrupled to perform the same slicing in the 393.7 PPCM (1000 PPI) and 787.4 PPCM (2000 PPI) scans, respectively.

A number of captured fingerprints did not fit completely within the bounding box slices defined on the ink card scans. Because this is typically *not* corrected when digitizing inked cards in bulk due to feasibility, it was also not corrected in SD 300, as the dataset was designed to be representative of operational criminal datasets. In many cases, a large percentage of the ridge structure is present and is still enough data from which to extract features.

Some adjustments to images were made when non-fingerprint information was visible in the sliced images. In the event this information, such as arrest circumstances or signatures, was visible, the image was cropped or omitted from SD 300.

3.3 Slap Segmentation

For image slices depicting simultaneous finger captures (friction ridge generalized positions 13 and 14), the `nfseg` fingerprint segmenter, distributed with NIST Biometric Image Software (NBIS) [4], was used to create rectangular polygons around the 1 to 4 individual fingers present in the slice. Each set of segmentation coordinates was visually inspected for accuracy and adjusted if necessary. These coordinates were used by another tool, `slapcrop` [5], to segment the simultaneous captures into individual images. The coordinates are provided as part of SD 300.

`slapcrop` also rotated the segmented images to be in an approximately upright position, using rotation angles that were calculated by `nfseg` and visually inspected for accuracy. Although image rotation interpolates pixels and thus reduces quality, most ten-print fingerprint matching algorithms require fingerprints in an upright position in order to ensure accuracy when matching. The rotation angle is also provided as a part of SD 300.

4. Data

4.1 Inventory

An inventory of the data available in SD 300 is shown in Table 1.

4.2 Ground Truth

Several commercial feature extraction and matching algorithms were used to perform 1:1 matching of the segmented plain (friction ridge generalized positions 11 to 14) captures to the rolled captures from the same subject. High-scoring non-mated pairs and low-scoring mated pairs in common between the majority of the algorithms were manually inspected to check for errors. Example errors included incorrect finger position labeling, bad image slicing, presence of non-fingerprint information, deformations of the source

card, and/or otherwise poor image quality. Obvious labeling, non-fingerprint information, and gratuitous image slicing errors were corrected where possible, and omitted otherwise.

4.3 Image Quality

A summary of image quality is presented in tabular form in Tables 2 and 3, using both the original NIST Fingerprint Image Quality (NFIQ) [6] algorithm as well as NFIQ 2.0 [7]. These values are additionally visualized in Figs. 2 and 3. The NFIQ 2.0 algorithm is not yet trained on rolled images, and as such, the NFIQ 2.0 values in Table 2 and Fig. 3 are unofficial.

NFIQ	Count
1	3 971
2	554
3	2 939
4	529
5	878

NFIQ 2.0	Count
0 to 10	48
11 to 20	417
21 to 30	1 372
31 to 40	2 468
41 to 50	3 341
51 to 60	1 154
61 to 70	67
71 to 80	4
81 to 90	0
91 to 100	0

Table 2. Distribution of NFIQ and NFIQ 2.0 quality values for rolled captures. The NFIQ 2.0 algorithm is not yet trained on rolled images, and as such, these NFIQ 2.0 rolled values are unofficial.

NFIQ	Count
1	2 828
2	2 068
3	2 963
4	253
5	675

NFIQ 2.0	Count
0 to 10	692
11 to 20	908
21 to 30	1 171
31 to 40	1 275
41 to 50	1 411
51 to 60	1 345
61 to 70	1 255
71 to 80	584
81 to 90	140
91 to 100	6

Table 3. Distribution of NFIQ and NFIQ 2.0 quality values for segmented plain captures.

4.4 Number of Minutiae

Of the 155 total quality features tested during development of the NFIQ 2.0 algorithm, minutiae counts were selected as one of the final 14 features incorporated into the overall quality score. The count of high-quality minutiae found for images in this dataset, as discovered by *FingerJet Feature eXtractor (FX) Open Source Edition (OSE)* via NFIQ 2.0, are presented in Fig. 4. These values were derived by multiplying the `FingerJetFX_MinutiaeCount` NFIQ 2.0 feature value by the `FJFXPos_OCL_MinutiaeQuality_80` NFIQ 2.0 feature value.

5. Obtaining and Using Special Database 300

The dataset can be downloaded from the Internet for free by visiting our website, https://www.nist.gov/itl/iad/image-group/special-database-300. Before downloading, researchers must agree to the terms and conditions of SD 300 that are listed on the webpage.

The directory structure of SD 300 after expanding the downloaded archive can be found in Fig. 5. This directory structure was chosen to allow for NIST to easily deliver future versions of the same images in different scanning resolutions and file formats alongside this distribution.

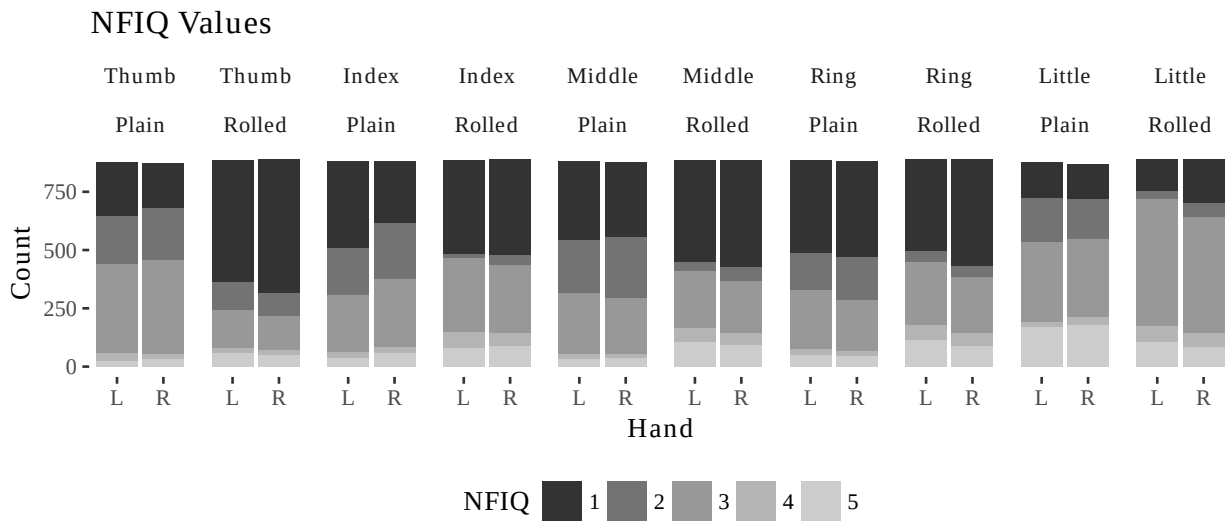

Fig. 2. NFIQ values for plain and rolled captures, separated by hand and friction ridge generalized position. For NFIQ 2.0 values, see Fig. 3.

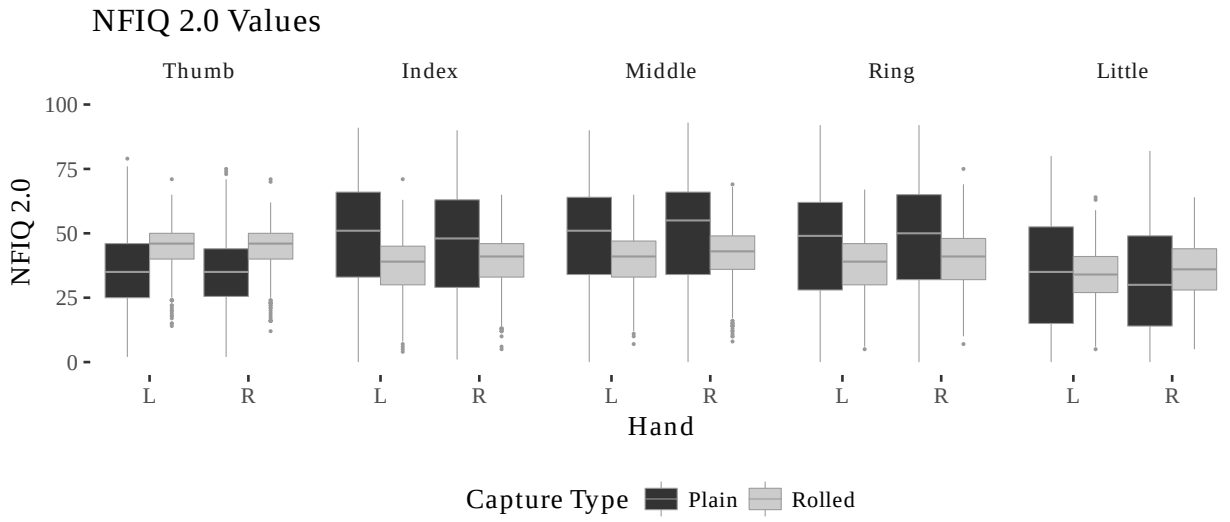

Fig. 3. NFIQ 2.0 values for plain and rolled captures, separated by hand and friction ridge generalized position. For NFIQ values, see Fig. 2. The NFIQ 2.0 algorithm is not yet trained on rolled images, and as such, these NFIQ 2.0 rolled values are unofficial.

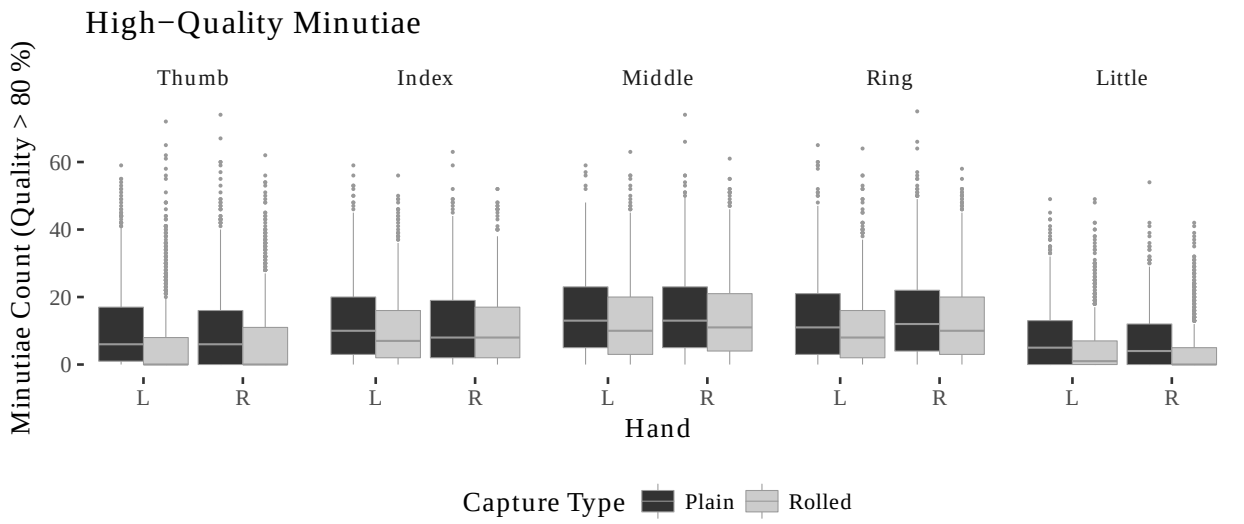

Fig. 4. Count of high-quality minutiae in plain and rolled captures discovered by FingerJet FX OSE via NFIQ 2.0. The NFIQ 2.0 algorithm is not yet trained on rolled images, and as such, these NFIQ 2.0 rolled values are unofficial.

The topmost directory contains a single directory, `images`. That directory contains a directory for each resolution in which the fingerprint cards were scanned. Each resolution directory contains a directory of image file formats. Inside the file format directories are directories for `plain` and `roll` impression types. The impression type directories contain images named in the form **`SUBJECT_IMPRESSION_PPI_FRGP.EXT`**, where:

`SUBJECT` Unique identifier for this individual's fingerprints.

`IMPRESSION` Impression type for each fingerprint capture.

`PPI` The scanning resolution of the image, in pixels per inch.

`FRGP` The ANSI/NIST-ITL 1-2011 Update:2015 friction ridge generalized position code (Table 4).

`EXT` File format extension.

The images in the `roll` directory contain friction ridge generalized positions 1 to 10, as scanned from the top of the card. Within the `plain` directory are scans of the simultaneous capture plain images, as well as the segmented and rotated versions of the distal joints of the fingers within the simultaneous capture plain images. A comma-separated values (CSV) file, `segmentation_coordinates_`**`PPI`**`.csv`, is included, and contains the four rectangular coordinates and rotation angle (in degrees) used to create these images.

SD 300 is a series of distributions, each containing a particular image encoding and resolution. For example, SD 300a contains 196.85 PPCM (500 PPI) Portable Network Graphics (PNG) images.

A CSV file, `checksum_`**`PPI_EXT_IMPRESSION`**`.csv`, accompanies every directory of images. Contained in this file are the SHA 256 checksums of the files contained within the named directory.

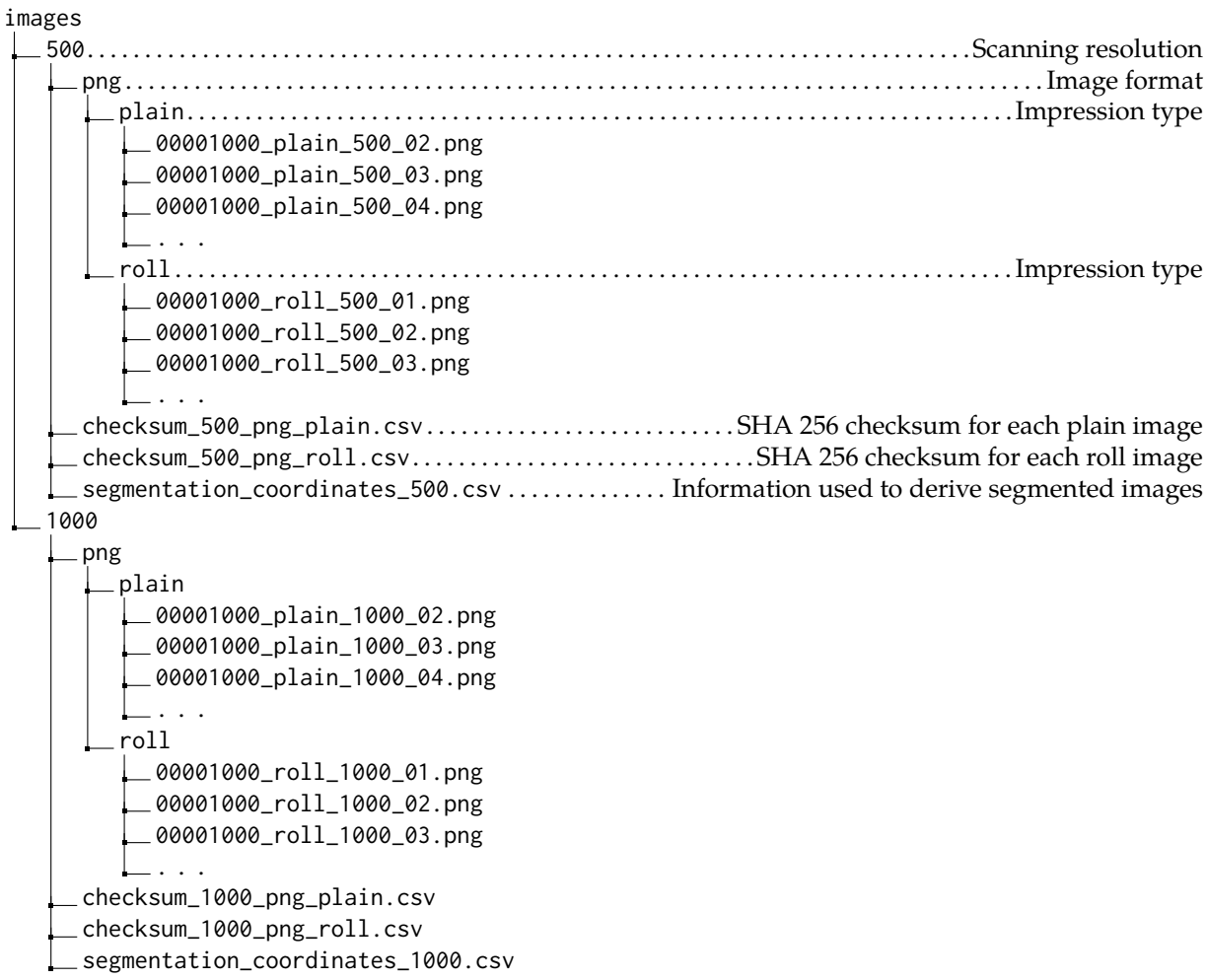

Fig. 5. Directory listing of SD 300. For an explanation of filenames, refer to Section 5.

FRGP	Description
1	Right Thumb
2	Right Index
3	Right Middle
4	Right Ring
5	Right Little
6	Left Thumb
7	Left Index
8	Left Middle
9	Left Ring
10	Left Little
11	Plain Right Thumb
12	Plain Left Thumb
13	Plain Left Four Fingers
14	Plain Right Four Fingers

Table 4. Friction ridge generalized position mapping, reproduced from ANSI/NIST-ITL 1-2011 Update:2015, Table 9 [2].

References

[1] Grantham J (2017) NISTscan, https://github.com/usnistgov/NISTscan. [Online; accessed 14 June 2018].

[2] American National Standard for Information Systems (2016) Information Technology: ANSI/NIST-ITL 1-2011 Update 2015 — Data Format for the Interchange of Fingerprint, Facial & Other Biometric Information. *NIST Special Publication 500-290e3* https://doi.org/10.6028/NIST.SP.500-290e3

[3] Federal Bureau of Investigation (2017) Standard Fingerprint Form (FD-258), https://www.fbi.gov/file-repository/standard-fingerprint-form-fd-258-1.pdf/view. [Online; accessed 14 June 2018].

[4] Watson C, et al. (2007) User's Guide to Export Controlled Distribution of NIST Biometric Image Software (NBIS-EC). *NIST Interagency Report 7391* https://doi.org/10.6028/NIST.IR.7391

[5] Salamon W, Fiumara G (2017) Biometric Evaluation Framework Tools, https://github.com/usnistgov/BiometricEvaluation. [Online; accessed 14 June 2018].

[6] Tabassi E, Wilson CL, Watson CI (2004) Fingerprint Image Quality. *NIST Interagency Report 7151* https://doi.org/10.6028/NIST.IR.7151

[7] Tabassi E (2016) Development of NFIQ 2.0, https://www.nist.gov/services-resources/software/development-nfiq-20. [Online; accessed 14 June 2018].

A. Change Log

Release 1: 10 January 2020

- A note in the front matter originally indicated that the protocol reviewed by the NIST Human Subjects Protection Office (now Research Protections Office) was ***exempt*** *human subjects research*. This was incorrect. The determination was ***not*** *human subjects research*.
- The last sentence in Section 2 stated that all subjects whose biometrics appeared in the arrest cards processed to create NIST SD 300 were deceased at the time NIST TN 1993 was published. While this was true, the sentence has been updated to clarify that the subjects were deceased at the time the arrest cards were received by NIST.