

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17



Face In Video Evaluation (FIVE)

Face In Video Evaluation (FIVE)
Concept, Evaluation Plan, and API
Version 0.7

Patrick Grother and Mei Ngan

Image Group
Information Access Division
Information Technology Laboratory

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

November 19, 2014

18

Timeline of the FIVE Evaluation

| Phase | Date | External actions, deadlines |
|---------|------------|---|
| Phase 0 | 2014-07-15 | Web site up, announce schedule |
| | 2014-08-15 | First draft Evaluation Plan and API |
| | 2014-08-31 | Public comments on first drafts due |
| | 2014-10-01 | Second draft Evaluation Plan and API |
| | 2014-10-15 | Public comments on second drafts due |
| | 2014-11-04 | Third draft Evaluation Plan and API. Draft five.h available. |
| | 2014-11-11 | Public comments on third drafts due |
| | 2014-11-17 | Final Evaluation Plan and API available. Final five.h available |
| | 2014-11-17 | FIVE validation package available |
| Phase 1 | 2014-11-17 | Opening of Phase 1 submission period |
| | 2015-02-08 | Deadline for submission for inclusion of results in first interim report card |
| | 2015-04-03 | First interim report card released to submitting participants |
| Phase 2 | 2015-04-06 | Opening of Phase 2 submission period |
| | 2015-06-12 | Deadline for submission for inclusion of results in second interim report card. |
| | 2015-08-07 | Second interim report card released to submitting participants |
| Phase 3 | 2015-08-10 | Opening of Phase 3 |
| | 2015-10-19 | Deadline for submission of algorithms to Phase 3 |

19

20

| November 2014 | | | | | | | December 2014 | | | | | | | January 2015 | | | | | | | February 2015 | | | | | | | March 2015 | | | | | | | April 2015 | | | | | | | |
|---------------|----|----|----|----|----|----|---------------|----|----|----|----|----|----|--------------|----|----|----|----|----|----|---------------|----|----|----|----|----|----|------------|----|----|----|----|----|----|------------|----|----|----|----|----|----|---|
| Su | Mo | Tu | We | Th | Fr | Sa | Su | Mo | Tu | We | Th | Fr | Sa | Su | Mo | Tu | We | Th | Fr | Sa | Su | Mo | Tu | We | Th | Fr | Sa | Su | Mo | Tu | We | Th | Fr | Sa | Su | Mo | Tu | We | Th | Fr | Sa | |
| | | | | | | 1 | 1 | 2 | 3 | 4 | 5 | 6 | | | | | | | 1 | 2 | 3 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | |
| 9 | 10 | 11 | 12 | 13 | 14 | 15 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | |
| 16 | 17 | 18 | 19 | 20 | 21 | 22 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | |
| 23 | 24 | 25 | 26 | 27 | 28 | 29 | 28 | 29 | 30 | 31 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | | | | | | | | 29 | 30 | 31 | 26 | 27 | 28 | 29 | 30 | | | | | | | | | | |
| 30 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

21

22 Major API Changes since FRVT 2013 Class V

23 The header/source files for the API will be made available to implementers at <http://nigos.nist.gov:8080/five>.

- 24 — The structures ONEFACE (see Table 12) and MULTIFACE (see Table 13) have been changed to classes.
- 25 — The MULTIFACE class contains a new “description” member variable and valid values are specified in Table 11.
- 26 — The labels for describing types of still images have been updated (see Table 10).
- 27 — The ONEVIDEO (see Table 15) class contains a new “peopleDensity” member variable and valid values are specified in
- 28 Table 14.

29

30 **Table of Contents**

31 1. FIVE 6

32 1.1. Scope 6

33 1.2. Audience 6

34 1.3. Market drivers 7

35 1.4. Offline testing 7

36 1.5. Phased testing 7

37 1.6. Interim reports 7

38 1.7. Final reports..... 7

39 1.8. Application scenarios..... 8

40 1.9. Image source labels 8

41 1.10. Rules for participation 8

42 1.11. Number and schedule of submissions 8

43 1.12. Use of multiple images per person 9

44 1.13. Core accuracy metrics 9

45 1.14. Generalized accuracy metrics..... 10

46 1.15. Reporting template size..... 10

47 1.16. Reporting computational efficiency 10

48 1.17. Exploring the accuracy-speed trade-space 10

49 1.18. Hardware specification 10

50 1.19. Operating system, compilation, and linking environment 11

51 1.20. Software and documentation..... 11

52 1.21. Runtime behavior 13

53 1.22. Threaded computations 13

54 1.23. Time limits 14

55 1.24. Test datasets..... 14

56 1.25. Ground truth integrity 15

57 2. Data structures supporting the API 15

58 2.1. Overview 15

59 2.2. Requirement 16

60 2.3. File formats and data structures 16

61 2.4. File structures for enrolled template collection 22

62 3. API Specification 22

63 3.1. Definitions 22

64 3.2. 1:N Identification 24

65 3.3. Interfaces..... 25

66 4. References 40

67 Annex A Submission of Implementations to the FIVE 41

68 A.1 Submission of implementations to NIST 41

69 A.2 How to participate..... 41

70 A.3 Implementation validation 42

71

72 **List of Figures**

73 Figure 1 – Organization and documentation of the FIVE 6

74 Figure 2 – Examples of pose angles and their encodings (yaw, pitch) 17

75

76 **List of Tables**

77 Table 1 – Abbreviations..... 5

78 Table 2 – Subtests supported under the FIVE activity..... 8

79 Table 3 – Cumulative total number of algorithms 8

80 Table 4 – Summary of accuracy metrics..... 9

81 Table 5 – Implementation library filename convention 12

FIVE

82 Table 6 – Number of threads allowed for each function 13

83 Table 7 – Processing time limits in milliseconds 14

84 Table 8 – Main video corpora (others will be used) 14

85 Table 9 – Labels describing types of images..... 16

86 Table 10 – Labels describing **categories** of MULTIFACES..... 17

87 Table 11 – **ONEFACE** class 18

88 Table 12 – **MULTIFACE** class 18

89 Table 13 – Labels describing the density of people in the video frames 18

90 Table 14 – **ONEVIDEO** Class 19

91 Table 15 – **EYEPAIR** Class 19

92 Table 16 – PersonTrajectory typedef 20

93 Table 17 – **PERSONREP** Class 20

94 Table 18 – **CANDIDATE** Class 20

95 Table 19 – **CANDIDATELIST** typedef 21

96 Table 20 – ReturnCode class 21

97 Table 21 – Enrollment dataset template manifest 22

98 Table 22 – API implementation requirements for FIVE 22

99 Table 23 – Procedural overview of the identification test 24

100 Table 24 – VideoEnrollment::initialize 26

101 Table 25 – VideoEnrollment::generateEnrollmentTemplate 26

102 Table 26 – ImageEnrollment::initialize 27

103 Table 27 – ImageEnrollment::generateEnrollmentTemplate..... 28

104 Table 28 – Finalize::finalize 30

105 Table 29 – VideoFeatureExtraction::initialize 32

106 Table 30 – VideoFeatureExtraction::generateIdTemplate 32

107 Table 31 – ImageFeatureExtraction::initialize 33

108 Table 32 – ImageFeatureExtraction::generateIdTemplate 33

109 Table 33 – Search::initialize 34

110 **Table 34 – VideoToVideoSearch::identify** 35

111 **Table 35 – StillToVideoSearch::identify** 36

112 **Table 36 – VideoToStillSearch::identify** 37

113

114 **Acknowledgements**

115 — The authors are grateful to the experts who made extensive comments on the first version of this document.

116 **Project History**

117 — 2012 – 2014 – The FRVT 2013 program included a video track (class V) that evaluated face recognition from video.
 118 The FIVE program supersedes the FRVT work but proceeds in an almost identical manner.

119 — August 15, 2014 - Release of first public draft of the Face In Video Evaluation (FIVE) – Concept, Evaluation Plan and
 120 API v0.1.

121 **Terms and definitions**

122 The abbreviations and acronyms of Table 1 are used in many parts of this document.

123 **Table 1 – Abbreviations**

| | |
|---------------|--|
| FNIR | False negative identification rate |
| FPIR | False positive identification rate |
| FIVE | NIST’s Face In Video Evaluation program |
| FRVT | NIST’s Face Recognition Vendor Test program |
| FTA | Failure to acquire a search sample |
| FTE | Failure to extract features from an enrollment image |
| DET | Detection error tradeoff characteristic: For identification this is a plot of FNIR vs. FPIR. |
| INCITS | InterNational Committee on Information Technology Standards |
| ISO/IEC 19794 | ISO/IEC 19794-5: Information technology — Biometric data interchange formats — Part 5:Face image data. First edition: 2005-06-15. (See Bibliography entry). |
| MBE | NIST’s Multiple Biometric Evaluation program |
| NIST | National Institute of Standards and Technology |
| SDK | The term Software Development Kit refers to any library software submitted to NIST. This is used synonymously with the terms "implementation" and "implementation under test". |

124

125 **1. FIVE**

126 **1.1. Scope**

127 The Face In Video Evaluation (FIVE) is being conducted to assess the capability of face recognition algorithms to correctly
 128 identify or ignore persons appearing in video sequences – i.e. the open-set identification problem. Both comparative and
 129 absolute accuracy measures are of interest, given the goals to determine which algorithms are most effective and
 130 whether any are viable for the following primary operational use-cases:

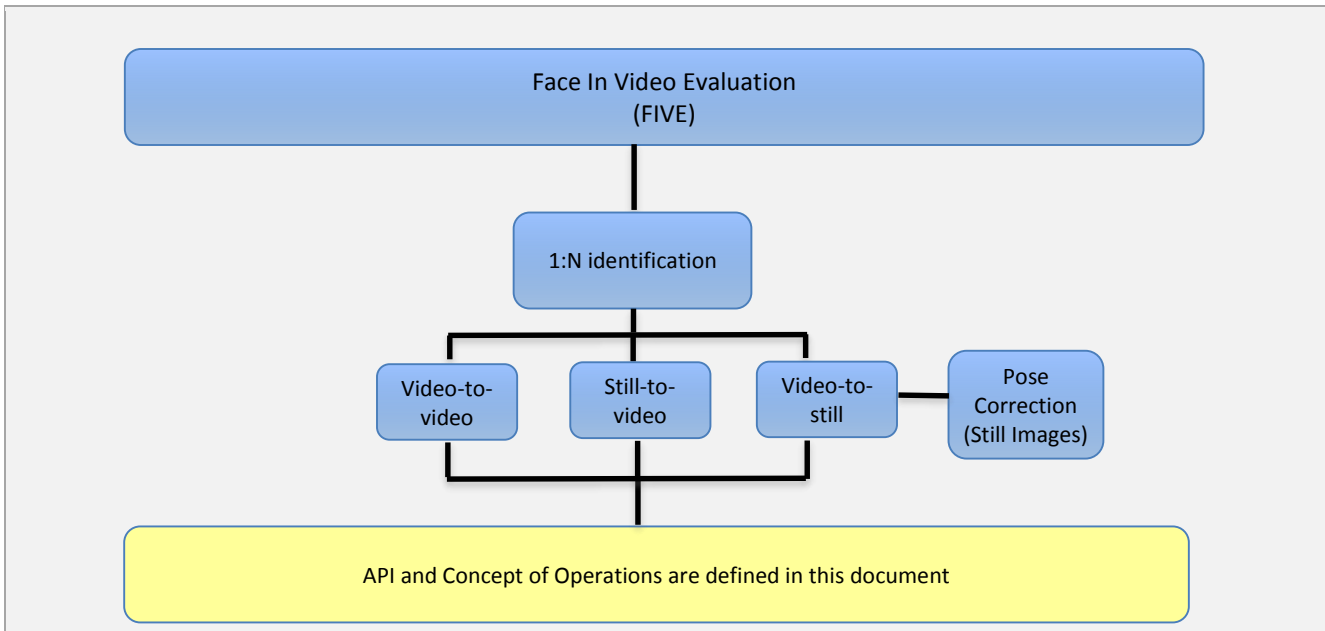
- 131 1. High volume screening of persons in the crowded spaces (e.g. an airport)
- 132 2. Low volume forensic examination of footage from a crime scene (e.g. a convenience store)
- 133 3. Persons in business meetings (e.g. for video-conferencing)
- 134 4. Persons appearing in television footage

135
 136
 137 These applications differ in their tolerance of false positives, whether a human examiner will review outputs, the prior
 138 probabilities of mate vs. non-mate presence, and the cost of recognition errors.

139
 140 **Out of scope:** Areas that are out of scope for this evaluation and will not be studied include: gait, iris and voice
 141 recognition; recognition across multiple views (e.g. via stereoscopic techniques); tracking across sequential cameras (re-
 142 identification); anomaly detection; detection of evasion.

143
 144 This document establishes a concept of operations and an application programming interface (API) for evaluation of face
 145 recognition in video implementations submitted to NIST's Face In Video Evaluation. See
 146 <http://www.nist.gov/itl/iad/ig/five.cfm> for all FIVE documentation.

147



148 **Figure 1 – Organization and documentation of the FIVE**

149 **1.2. Audience**

150 Universities and commercial entities with capabilities in detection and identification of faces in video sequences **and/or**
 151 **pose correction capabilities on still images** are invited to participate in the FIVE Video test.

152 Organizations will need to implement the API defined in this document. Participation is open worldwide. There is no
 153 charge for participation. While NIST intends to evaluate technologies that could be readily made operational, the test is
 154 also open to experimental, prototype and other technologies.

155 **1.3. Market drivers**

156 This test is intended to support a plural marketplace of face recognition in video systems. There is considerable interest
 157 in the potential use of face recognition for identification of persons in videos.

158 **1.4. Offline testing**

159 While this set of tests is intended as much as possible to mimic operational reality, this remains an offline test executed
 160 on databases of images. The intent is to assess the core algorithmic capability of face recognition in video algorithms. This
 161 test will be conducted purely offline - it does not include a live human-presents-to-camera component. Offline testing is
 162 attractive because it allows uniform, fair, repeatable, and efficient evaluation of the underlying technologies. Testing of
 163 implementations under a fixed API allows for a detailed set of performance related parameters to be measured.

164 **1.5. Phased testing**

165 To support research and development efforts, this testing activity will embed multiple rounds of testing. These test
 166 rounds are intended to support improved performance. Once the test commences, NIST will evaluate implementations
 167 on a first-come-first-served basis and will return results to providers as expeditiously as possible. Providers may submit
 168 revised SDKs to NIST only after NIST provides results for the prior SDK and invites further submission. The frequency with
 169 which a provider may submit SDKs to NIST will depend on the times needed for developer preparation, transmission to
 170 NIST, validation, execution and scoring at NIST, and developer review and decision processes.

171 For the schedule and number of SDKs of each class that may be submitted, see sections 1.10 and 1.11.

172 **1.6. Interim reports**

173 The performance of each SDK will be reported in a "score-card". This will be provided to the participant. While the score
 174 cards may be used by the provider for arbitrary purposes, they are intended to facilitate development. Score cards will

- 175 – be machine generated (i.e. scripted),
- 176 – be provided to participants with identification of their implementation,
- 177 – include timing, accuracy and other performance results,
- 178 – include results from other implementations, but will not identify the other providers,
- 179 – be expanded and modified as revised implementations are tested, and as analyses are implemented,
- 180 – be generated and released asynchronously with SDK submissions,
- 181 – be produced independently of the other status of other providers' implementations,
- 182 – be regenerated on-the-fly, usually whenever any implementation completes testing, or when new analysis is added.

183 NIST does not intend to release these interim test reports publicly. NIST may release such information to the U.S.
 184 Government test sponsors. While these reports are not intended to be made public, NIST can only request that agencies
 185 not release this content.

186 **1.7. Final reports**

187 NIST will publish one or more final public reports. NIST may also

- 188 – publish additional supplementary reports (typically as numbered NIST Interagency Reports),
- 189 – publish in other academic journals,
- 190 – present results at conferences and workshops (typically PowerPoint).

191 Our intention is that the final test reports will publish results for the best-performing implementation from each
 192 participant. Because “best” is ill-defined (accuracy vs. time vs. template size, for example), the published reports may
 193 include results for other implementations. The intention is to report results for the most capable implementations (see
 194 section 1.13, on metrics). Other results may be included (e.g. in appendices) to show, for example, examples of progress
 195 or tradeoffs. IMPORTANT: Results will be attributed to the providers.

196 **1.8. Application scenarios**

197 This test will include one-to-many identification tests for video sequences. As described in Table 2, the test is intended to
 198 represent identification applications for face recognition in video.

199 **Table 2 – Subtests supported under the FIVE activity**

| # | | Video-to-Video | Video-to-Still | Still-to-Video | Pose Correction |
|----|--|---|---|---|---|
| 1. | Aspect | 1:N identification of video-to-video | 1:N identification of video-to-still | 1:N identification of still-to-video | Frontal reconstruction of off-angle still facial images |
| 2. | Enrollment dataset | N enrolled video sequences | N enrolled stills | N enrolled video sequences | |
| 3. | Prior NIST test references | Equivalent to 1 to N matching in [FRVT 2013] | | | Class F from [FRVT 2013] |
| 4. | Example Application | Identity Clustering, Re-identification | Watch-list Surveillance | Media Search, Asylum re-identification | Pose correction of facial images prior to enrollment into gallery |
| 5. | Score or feature space normalization support | Any score or feature based statistical normalization techniques-are applied against enrollment database | | | NA |
| 6. | Intended number of subjects in gallery, N | Expected $O(10^2)$ - $O(10^4)$ | | | |
| 7. | Number of gallery images per individual | N/A | Sometimes fixed at 1, otherwise variable, see section 1.12. | Sometimes fixed at 1, otherwise variable, see section 1.12. | Sometimes fixed at 1, otherwise variable, see section 1.12 |
| 8. | Number of search individuals | Large numbers of searches will be conducted. A search generally includes imagery containing more than one individual, except in the case of a still-to-video search where only one individual will appear. Some searches have an enrolled mate, and the remainder do not; the relative proportions are not disclosed. | | | Pose correction not applied except to gallery images in video-to-still searches |

200 **1.9. Image source labels**

201 NIST may mix images from different sources in an enrollment set. For example, NIST could combine frontal images and
 202 images with varying poses into a single enrollment dataset. For this reason, in the data structure defined in clause 2.3.4,
 203 each image is accompanied by a "label" which identifies the set-membership images. Legal values for labels are in clause
 204 2.3.2.

205 **1.10. Rules for participation**

206 **There is no charge to participate in FIVE.** A participant must properly follow, complete and submit a participation
 207 agreement (see Annex A). This must be done once, not before November 17, 2014. It is not necessary to do this for each
 208 submitted SDK. All submitted SDKs must meet the API requirements as detailed in section 3.

209 **1.11. Number and schedule of submissions**

210 The test is conducted in three phases, as scheduled on page 2. The maximum total (i.e. cumulative) number of
 211 submissions is regulated in Table 3.

212 **Table 3 – Cumulative total number of algorithms**

| # | Phase 1 | Total over Phases 1 + 2 | Total over Phases 1 + 2 + 3 |
|---|---------|-------------------------|-----------------------------|
|---|---------|-------------------------|-----------------------------|

| | | | |
|---|---|---|---|
| Cumulative total number of video recognition submissions | 2 | 3 | 5 if at least 1 was successfully executed by end Phase 2 2 if zero had been successfully executed by end Phase 2 |
| Cumulative total number of image (pose) correction algorithms | 1 | 2 | 3 |

213 The numbers above may be increased as resources allow.

214 Participation in Phase 1 is not required for Phase 2.

215 NIST cannot conduct surveys over runtime parameters because NIST must limit the extent to which participants are able
216 to train on the test data.

217 **1.12. Use of multiple images per person**

218 Some of the proposed datasets includes $K > 2$ images per person for some persons. For video-to-still recognition in this
219 test, NIST will enroll $K \geq 1$ images under each identity. For still-to-video, the probe will consist of $K \geq 1$ images. Normally
220 the probe will consist of a single image, but NIST may examine the case that it could consist of multiple images. The
221 method by which the face recognition implementation exploits multiple images is not regulated: The test seeks to
222 evaluate developer provided technology for multi-presentation fusion. This departs from some prior NIST tests in which
223 NIST executed fusion algorithms (e.g. [FRVT2002b]), and sum score fusion, for example, [MINEX]).

224 This document defines a template to be the result of applying feature extraction to a set of $K \geq 1$ images or $K \geq 1$ video
225 frames. That is, a template contains the features extracted from one or more images or video frames, not generally just
226 one. An SDK might internally fuse K feature sets into a single representation or maintain them separately - In any case the
227 resulting proprietary template is contained in a contiguous block of data. All identification functions operate on such
228 multi-image or multi-frame templates.

229 The number of images per person will depend on the operationally deployed application area:

- 230 – In civil identity credentialing (e.g. passports, driving licenses) the images will be acquired approximately uniformly
231 over time (e.g. five to ten years for a Canadian passport). While the distribution of dates for such images of a person
232 might be assumed uniform, a number of factors might undermine this assumption¹.
- 233 – In criminal applications the number of images would depend on the number of arrests². The distribution of dates for
234 arrest records for a person (i.e. the recidivism distribution) has been modeled using the exponential distribution, but
235 is recognized to be more complicated. NIST currently estimates that the number of images will never exceed 100.

236 **1.13. Core accuracy metrics**

237 For identification testing, the test will target open-universe applications such as benefits-fraud, watch-lists, and forensic
238 investigation. It will not address the closed-set task because it is operationally uncommon.

239 While some one-to-many applications operate with purely rank-based metrics, this test will primarily target score-based
240 identification metrics. Metrics are defined in Table 4. The analysis will survey over various rank and thresholds³. Plots of
241 the two error rates, parametric on threshold, will be the primary reporting mechanism.

242 **Table 4 – Summary of accuracy metrics**

| Application | Metric |
|-------------|--------|
|-------------|--------|

¹ For example, a person might skip applying for a passport for one cycle (letting it expire). In addition, a person might submit identical images (from the same photography session) to consecutive passport applications at five year intervals.

² A number of distributions have been considered to model recidivism, see "Random parameter stochastic process models of criminal careers." In Blumstein, Cohen, Roth & Visher (Eds.), Criminal Careers and Career Criminals, Washington, D.C.: National Academy of Sciences Press, 1986.

³ Threshold and rank limits are established operationally to limit human labor requirements: On the one side, in a low volume forensic application e.g. investigation of video collected in a convenience store hold-up, or in looking at videos of passengers dis-embarking flights to document an asylum claim, an examiner might be willing to adjudicate $R \gg 1$ candidates with threshold, $T = 0$. At the other end, a high volume watch-list application in which crowded airport concourses are surveilled for bad actors, a high threshold would be used to limit false positive outcomes. In that case, candidate lists will often have zero length. NIST will report metrics appropriate to the "human-automated" hybrid application, and the "lights-out" hits-are-rare use case.

| | |
|--------------------|--|
| 1:N Identification | FPIR = The rate at which unknown subjects are incorrectly associated with any of N enrolled identities. The association will be parameterized on a continuous threshold T. |
| | FNIR = The rate at which known subjects are incorrectly not associated with the correct enrolled identities. The association will be parameterized on a continuous threshold T, and a candidate rank, R. |

243
244 FPIR will be estimated using probe images or video clips for which there is no enrolled mate. The stability of FPIR at a
245 fixed threshold under changes to image properties or demographics will be reported.

246 NIST will extend the analysis in other areas, with other metrics, and in response to the experimental data and results.

247 **1.14. Generalized accuracy metrics**

248 Under the ISO/IEC 19795-1 biometric testing and reporting standard, a test must account for "failure to acquire" (FTA)
249 and "failure to enroll" (FTE) events (e.g. elective refusal to make a template, or fatal errors). The way these are treated is
250 application-dependent.

251 For identification, the appropriate metrics reported in FIVE will be generalized to include FTA and FTE events.

252 **1.15. Reporting template size**

253 Because template size is influential on storage requirements and computational efficiency, this API supports
254 measurement of template size. NIST will report statistics on the actual sizes of templates produced by face recognition
255 implementations submitted to FIVE. NIST may report statistics on runtime memory usage. Template sizes were reported
256 in the FRVT 2013 test⁴, IREX III test⁵, and the MBE-STILL 2010 test⁶.

257 **1.16. Reporting computational efficiency**

258 As with other tests, NIST will compute and report recognition accuracy. In addition, NIST will also report timing statistics
259 for all core functions of the submitted SDK implementations. This includes feature extraction and 1:N recognition. For an
260 example of how efficiency can be reported, see the final report of the FRVT 2013 test, IREX III test, and the MBE-STILL
261 2010 test.

262 **1.17. Exploring the accuracy-speed trade-space**

263 NIST will explore the accuracy vs. speed tradeoff for face recognition algorithms running on a fixed platform. NIST will
264 report both accuracy and speed of the implementations tested. While NIST cannot force submission of "fast vs. slow"
265 variants, participants may choose to submit variants on some other axis (e.g. "experimental vs. mature")
266 implementations. NIST encourages "fast-less-accurate vs. slow-more-accurate" with a factor of three between the speed
267 of the fast and slow versions.

268 **1.18. Hardware specification**

269 NIST intends to support high performance by specifying the runtime hardware beforehand. There are several types of
270 computer blades that may be used in the testing. The blades are labeled as Dell M605, M905, M610, and M910. The
271 following list gives some details about the hardware of each blade type:

- 272 • ~~Dell M605 - Dual Intel Xeon E5405 2 GHz CPUs (4 cores each)~~
- 273 • Dell M905 - Quad AMD Opteron 8376HE 2 GHz CPUs (4 cores each)
- 274 • Dell M610 - Dual Intel Xeon X5680 3.3 GHz CPUs (6 cores each)
- 275 • Dell M910 - Dual Intel Xeon X7560 2.3 GHz CPUs (8 cores each)

⁴ See the FRVT 2013 test report: NIST Interagency Report 8009, linked from <http://face.nist.gov/frvt>

⁵ See the IREX III test report: NIST Interagency Report 7836, linked from <http://iris.nist.gov/irex>

⁶ See the MBE-STILL 2010 test report, NIST Interagency Report 7709, linked from <http://face.nist.gov/mbe>

276 Each CPU has 512K cache. The bus runs at 667 Mhz. The main memory is 192 GB Memory as 24 8GB modules. We
277 anticipate that 16 processes can be run without time slicing.

278 The minimum instruction set across all processors used in the evaluation is specified here⁷. Dependence on instructions
279 not included in the minimum instruction set is prohibited.

280 NIST is requiring use of 64 bit implementations throughout. This will support large memory allocation to support 1:N
281 identification task with image and video frame counts in the millions. For still images, if all templates were to be held in
282 memory, the 192GB capacity implies a limit of ~19KB per template, for a 10 million **one-image-per-identity** enrollment.
283 For video, given the data expectations and the occurrence of faces in the imagery, we anticipate the developers will have
284 sufficient memory for video templates. Note that while the API allows read access of the disk during the 1:N search, the
285 disk is, of course, relatively slow.

286 Some of the section 3 API calls allow the implementation to write persistent data to hard disk. The amount of data shall
287 not exceed 200 kilobytes per enrolled **identity**. NIST will respond to prospective participants' questions on the hardware,
288 by amending this section.

289 **1.19. Operating system, compilation, and linking environment**

290 The operating system that the submitted implementations shall run on will be released as a downloadable file accessible
291 from <http://nigos.nist.gov:8080/evaluations/>, which is the 64-bit version of **CentOS 7.0 running Linux kernel 3.10.0**.

292 For this test, Windows machines will not be used. Windows-compiled libraries are not permitted. All software must run
293 under Linux.

294 NIST will link the provided library file(s) to our C++ language test drivers. Participants are required to provide their library
295 in a format that is linkable using **g++ version 4.8.2**. The standard libraries are:

296 `/usr/lib64/libstdc++.so.6.0.19 lib64/libc.so.6 -> libc-2.17.so lib64/libm.so.6 -> libm-2.17.so`

297 A typical link line might be

298 `g++ -l. -Wall -m64 -o fivetest fivetest.cpp -L. -lfive_Enron_A_07`

299 The Standard C++ library should be used for development of the SDKs. The prototypes from the API of this document will
300 be written to a file "five.h" which will be included via

```
#include <five.h>
```

301 The header files will be made available to implementers at <http://nigos.nist.gov:8080/five>.

302 NIST will handle all input of images via the JPEG and PNG libraries, sourced, respectively from <http://www.iijg.org/> and
303 <http://libpng.org>.

304 All compilation and testing will be performed on x86 platforms. Thus, participants are strongly advised to verify library-
305 level compatibility with g++ (on an equivalent platform) prior to submitting their software to NIST to avoid linkage
306 problems later on (e.g. symbol name and calling convention mismatches, incorrect binary file formats, etc.).

307 Dependencies on external dynamic/shared libraries such as compiler-specific development environment libraries are
308 discouraged. If absolutely necessary, external libraries must be provided to NIST upon prior approval by the Test Liaison.

309 **1.20. Software and documentation**

310 **1.20.1. SDK Library and platform requirements**

311 Participants shall provide NIST with binary code only (i.e. no source code). Header files (".h") are allowed, but these shall
312 not contain intellectual property of the company nor any material that is otherwise proprietary. It is preferred that the
313 SDK be submitted in the form of a single static library file. However, dynamically linked shared library files are permitted.

⁷ `cat /proc/cpuinfo` returns `fpu vmx de pse tsc msr pae mce cx8 apic sep mtrr pge mca cmov pat pse36 clflush mmx fxsr sse sse2 ht syscall nx mmxext fxsr_opt pdpe1gb rdtscp lm 3wext 3dnow constant_tsc nonstop_tsc pni cx16 popcnt lahf_lm cmp_legacy svm extapic cr8_legacy altmovcr8 abm sse4a misalignsse 3dnowprefetch osvw`

314 The core library shall be named according to Table 5. Additional shared object library files may be submitted that support
 315 this “core” library file (i.e. the “core” library file may have dependencies implemented in these other libraries).

316 Intel Integrated Performance Primitives (IPP) libraries are permitted if they are delivered as a part of the developer-
 317 supplied library package. It is the provider’s responsibility to establish proper licensing of all libraries. The use of IPP
 318 libraries shall not inhibit the SDK’s ability to run on CPUs that do not support IPP. Please take note that some IPP
 319 functions are multithreaded and threaded implementations may complicate comparative timing.

320 Access to any GPUs is not permitted.

321 **Table 5 – Implementation library filename convention**

| Form | libFIVE_provider_class_sequence.ending | | | | |
|--|--|---|---|--|------------------|
| Underscore delimited parts of the filename | libFIVE | provider | class | sequence | ending |
| Description | First part of the name, required to be this. | Single word name of the main provider EXAMPLE: Acme | “V” for video submissions “PC” for pose correction submissions | A two digit decimal identifier to start at 00 and increment by 1 every time any SDK is sent to NIST. EXAMPLE: 07 | Either .so or .a |
| Example | libFIVE_Acme_V_07.a | | | | |

322
 323 NIST will report the size of the supplied libraries.

324 **1.20.2. Configuration and developer-defined data**

325 The implementation under test may be supplied with configuration files and supporting data files. The total size of the
 326 SDK, that is all libraries, include files, data files and initialization files shall be less than or equal to 1 073 741 824 bytes =
 327 1024³ bytes.

328 NIST will report the size of the supplied configuration files.

329 **1.20.3. Installation and Usage**

330 The SDK must install easily (i.e. one installation step with no participant interaction required) to be tested, and shall be
 331 executable on any number of machines without requiring additional machine-specific license control procedures or
 332 activation.

333 The SDK shall be installable using simple file copy methods. It shall not require the use of a separate installation program.

334 The SDK shall neither implement nor enforce any usage controls or limits based on licenses, number of executions,
 335 presence of temporary files, etc. The submitted implementations shall remain operable with no expiration date.

336 Hardware (e.g. USB) activation dongles are not acceptable.

337 **1.20.4. Hard disk space**

338 FIVE participants should inform NIST if their implementations require more than 200K of persistent storage, per enrolled
 339 identity on average.

340 **1.20.5. Documentation**

341 Participants shall provide complete documentation of the SDK and detail any additional functionality or behavior beyond
 342 that specified here. The documentation must define all (non-zero) developer-defined error or warning return codes.

343 **1.20.6. Modes of operation**

344 Individual SDKs provided shall not include multiple “modes” of operation, or algorithm variations. No switches or options
 345 will be tolerated within one library. For example, the use of two different “coders” by a feature extractor must be split
 346 across two separate SDK libraries, and two separate submissions.

347 **1.21. Runtime behavior**

348 **1.21.1. Interactive behavior**

349 The SDK will be tested in non-interactive “batch” mode (i.e. without terminal support). Thus, the submitted library shall
 350 not use any interactive functions such as graphical user interface (GUI) calls, or any other calls which require terminal
 351 interaction e.g. reads from “standard input”.

352 **1.21.2. Error codes and status messages**

353 The SDK will be tested in non-interactive “batch” mode, without terminal support. Thus, the submitted library shall run
 354 quietly, i.e. it should not write messages to "standard error" and shall not write to “standard output”. An SDK may write
 355 debugging messages to a log file - the name of the file must be declared in documentation.

356 **1.21.3. Exception Handling**

357 The application should include error/exception handling so that in the case of a fatal error, the return code is still
 358 provided to the calling application.

359 **1.21.4. External communication**

360 Processes running on NIST hosts shall not side-effect the runtime environment in any manner, except for memory
 361 allocation and release. Implementations shall not write any data to external resource (e.g. server, file, connection, or
 362 other process), nor read from such. If detected, NIST will take appropriate steps, including but not limited to, cessation of
 363 evaluation of all implementations from the supplier, notification to the provider, and documentation of the activity in
 364 published reports.

365 **1.21.5. Stateless behavior**

366 All components in this test shall be stateless, except as noted. This applies to face detection, feature extraction and
 367 matching. Thus, all functions should give identical output, for a given input, independent of the runtime history. NIST
 368 will institute appropriate tests to detect stateful behavior. If detected, NIST will take appropriate steps, including but not
 369 limited to, cessation of evaluation of all implementations from the supplier, notification to the provider, and
 370 documentation of the activity in published reports.

371 **1.22. Threaded computations**

372 Table 6 shows the limits on the numbers of threads a face recognition implementation may use. **Threading is prohibited**
 373 **for feature extraction and search**. NIST will parallelize the test by dividing the workload across many cores and many
 374 machines. For the functions where we allow multi-threading, NIST requires the provider to disclose the maximum
 375 number of threads to us.

376 **Table 6 – Number of threads allowed for each function**

| Function | Video-to-Video | Video-to-Still | Still-to-Video | Pose Correction |
|---|----------------|----------------|----------------|-----------------|
| Feature extraction for enrollment | 1 ≤ T ≤ 16 | 1 | 1 ≤ T ≤ 16 | 1 |
| Finalize enrollment offline, before 1:N identification | 1 ≤ T ≤ 16 | 1 ≤ T ≤ 16 | 1 ≤ T ≤ 16 | |
| Feature extraction for identification | 1 ≤ T ≤ 16 | 1 ≤ T ≤ 16 | 1 | |
| Identification | 1 ≤ T ≤ 16 | 1 ≤ T ≤ 16 | 1 ≤ T ≤ 16 | |

377 To expedite testing NIST will run up to P >> 1 processes concurrently. **We will reduce P when threading is in use**. NIST's
 378 calling applications are single-threaded.

379 **1.23. Time limits**

380 The elemental functions of the implementations shall execute under the time constraints of Table 7. These time limits
 381 apply to the function call invocations defined in section 3. Assuming the times are random variables, NIST cannot regulate
 382 the maximum value, so the time limits are 90-th percentiles. This means that 90% of all operations should take less than
 383 the identified duration.

384 The time limits apply per image or video frame. When K images of a person are present or K frames are in a video clip,
 385 the time limits shall be increased by a factor K.

386 **Table 7 – Processing time limits in milliseconds**

| Function | Video-to-Video | Video-to-Still | Still-to-Video |
|---------------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|
| Feature extraction for enrollment | 5 * 1500 per video frame (1 core) | 1500 per image (1 core) | 5 * 1500 per video frame (1 core) |
| Feature extraction for identification | 5 * 1500 per video frame (1 core) | 5 * 1500 per video frame (1 core) | 1500 per image (1 core) |

387 For video: the multiple of K=5 is a notional average of the number of persons expected in any given frame. This figure is
 388 proportionally unreliable for any given sample.

389 While there is no time limit for the enrollment finalization procedure, NIST will report the execution duration.

390 **1.24. Test datasets**

391 This section is under development. The data has, in some cases, been estimated from initial small partitions. The
 392 completion of this section depends on further work. The information is subject to change. We intend to update this
 393 section as fully as possible.

394 NIST is likely to use other datasets, in addition.

395 **Table 8 – Main video corpora (others will be used)**

| | Dataset P | Dataset T | Dataset B | Other datasets - Undisclosed |
|--|---|--------------------------|--|------------------------------|
| Collection, environment | Indoor public space with individuals walking mostly toward cameras as could occur on a transit terminal | | Television footage, indoor and outdoor | |
| Number of individuals in field of view | Multiple, usually below 20 many not fully visible but usually more than 1. | | Few, most often 1, occasionally others in background | |
| View angle | Various pitch due to different heights of camera installation, some yaw also due to subject behavior | | Pitch variation present, but yaw angles vary more due to subject behavior | |
| Video frame size | 1920 x 1080 | Various | Various | |
| Eye to eye distance | 10-100 pixels | 10-150 pixels | 10-120 | |
| | The above values are guidelines; exceptions will inevitably occur in large datasets. | | | |
| Camera properties | Consumer-grade video | Professional-grade video | Professional-grade video cameras | |
| Camera motion | Fixed geometry, fixed optics | | Usually camera is still or slowly panning or zooming | |
| Frames per second | 24 | Up to 30 | Up to 30 | |
| Similar composition to | Compare to the iLids data but with higher spatial resolution on the face. | | Similar to YouTubeFaces in that typically one subject is present and in the foreground | |
| Accompanying stills | Yes, for video-to-still and still-to-video searches, high-resolution stills approximating ISO/IEC 19794-5 are available. In addition, off-angle images exist with many combinations of pitch and yaw. In addition, less formal "social-media" like stills are available | | Stills usually resemble frames from the video. ISO/IEC 19794-5 images are not usually available. | |

| | | | |
|--|---|--|--|
| | <p>also. Various galleries will be formed from these images.</p> <p>Images for which interocular distance exceeds 240 pixels will be downsized.</p> | | |
|--|---|--|--|

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

NIST does not know the minimum and maximum numbers of persons appearing in video sequences. Moreover, NIST will apply the algorithms to other databases. The maximum number of frames in a video sequence will be limited by the duration of the sequence. NIST expects to use sequences whose duration extends from a few seconds to a few minutes

Some notes regarding the video data:

- NIST does not anticipate using interlaced video.
 - The videos are contiguous in time, without interruptions. The videos will not cross shot-boundaries, such that the background scene does not change abruptly.
 - Some sequences exist at much higher frame rates. NIST will examine whether this offers benefit.
 - Some of the datasets were collected using consumer-grade cameras capturing video in standard formats while others were collected using professional-grade cameras captured in modern proprietary video codecs.
- In some videos, the scenes capture people walking towards the camera. Occasionally, there are people walking in various transverse directions including people walking away from the camera. The cameras have varying pitch angles ranging from 0 degrees (frontal) to higher values. The depth of scene varies between the cameras such that the sizes of the faces vary, with the following:
- Eye-to-eye distances range from approximately 10 pixels to 120 pixels
 - Amount of time a face is fully visible in a scene can vary from approximately 0 to 30 seconds
 - Some of the captures include non-uniform lighting due to light coming through adjacent windows

Please note that the properties stated above may not hold for all datasets that might be employed in FIVE.

1.25. Ground truth integrity

Some of the test databases will be derived from operational systems. They may contain ground truth errors in which

- a person may appear in a video but not be tagged as being in the video, or
- a person may be tagged as being in a video but doesn't actually appear in the video, or
- a single person is present under two different identifiers, or
- two persons are present under one identifier, or
- a face is not present in the image.

If these errors are detected, they will be removed. NIST will use aberrant scores (high impostor scores, low genuine scores) to detect such errors. This process will be imperfect, and residual errors are likely. For comparative testing, identical datasets will be used and the presence of errors should give an additive increment to all error rates. For very accurate implementations this will dominate the error rate. NIST intends to attach appropriate caveats to the accuracy results. For prediction of operational performance, the presence of errors gives incorrect estimates of performance.

2. Data structures supporting the API

2.1. Overview

This section describes the API for the face recognition in video applications described in section 1.8. All SDK's submitted to FIVE shall implement the functions required in Section 3.

432 **2.2. Requirement**

433 FIVE participants shall submit an SDK which implements the relevant C++ prototyped interfaces of clause 3. C++ was
 434 chosen in order to make use of some object-oriented features.

435 **2.3. File formats and data structures**

436 **2.3.1. Overview**

437 In this test, an individual is represented by $K \geq 1$ two-dimensional still facial images, and by subject and image-specific
 438 metadata.

439 **2.3.2. Struct representing pose information for images**

440 **Table 9 – Pose struct**

| | C++ code fragment | Remarks |
|----|---------------------|---|
| 1. | typedef struct { | |
| 2. | bool isAssigned; | This boolean is set to true if the yaw and pitch values are assigned for an image. For uncontrolled images or images with unknown pose information, this boolean will be set to false, and the yaw and pitch values will not be assigned and shall not be parsed. |
| 3. | int yaw; | Yaw value on the range [-90,90]. See Figure 2 for pose angle encodings. |
| 4. | int pitch; | Pitch value on the range [-40, 40]. See Figure 2 for pose angle encodings. |
| 5. | } Pose; | |

441

442 For mugshot or visa images, the following values could be assigned:

443 `isAssigned=true; yaw=0; pitch=0;`

444 For off-angle images such as from the FERET b-series, the following values could be assigned:

445 `isAssigned=true; yaw=-25; pitch=0;`

446 For “in the wild” images, the following values could be assigned to the Pose struct:

447 `isAssigned=false; yaw=<unassigned>; pitch=<unassigned>;`

448

449 Note: The pose values assigned to images are nominal values and are not exact. Exact pose measurements would require
 450 specialized apparatus and to the best of our knowledge, was not used during image collection.

451

452 **2.3.3. Dictionary of terms describing images and MULTIFACES**

453 Still facial images will be accompanied by one of the labels given in Table 10. Face recognition implementations
 454 submitted to FIVE should tolerate images of any category.

455 **Table 10 – Labels describing types of images**

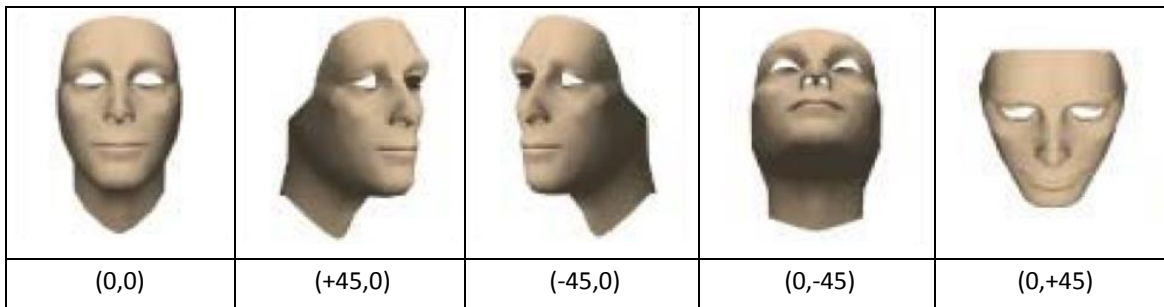
| | Label as C++ enumeration | Meaning | Yaw (degrees) | Pitch (degrees) |
|----|--------------------------|--------------|---------------|-----------------|
| | typedef enum { | | | |
| 1. | FF=0, | Full frontal | 0 | 0 |
| 2. | FD=1, | Face down | 0 | 10 to 40 |
| 3. | FU=2, | Face up | 0 | -10 to -40 |
| 4. | QL=3, | Quarter left | -10 to -35 | 0 |

FIVE

| | | | | |
|-----|---------------------|--|----------|----------|
| 5. | QR=4, | Quarter right | 10 to 35 | 0 |
| 6. | HL=5, | Half left | 36 to 70 | 0 |
| 7. | HR=6, | Half right | 36 to 70 | 0 |
| 8. | PL=7, | Profile left | 71 to 90 | 0 |
| 9. | PR=8, | Profile right | 71 to 90 | 0 |
| 10. | QLU=9, | Quarter left up | 10 to 35 | 10 to 40 |
| 11. | QRU=10, | Quarter right up | 10 to 35 | 10 to 40 |
| 12. | QLD=11, | Quarter left down | 10 to 35 | 10 to 40 |
| 13. | QRD=12, | Quarter right down | 10 to 35 | 10 to 40 |
| 14. | HLU=13, | Half left up | 36 to 70 | 10 to 40 |
| 15. | HRU=14, | Half right up | 36 to 70 | 10 to 40 |
| 16. | HLD=15, | Half left down | 36 to 70 | 10 to 40 |
| 17. | HRD=16, | Half right down | 36 to 70 | 10 to 40 |
| 18. | IMG_UNKNOWN=17, | Either the label is unknown or unassigned, | | |
| 19. | IMG_UNCONTROLLED=18 | Any illumination, pose is unknown and could be frontal | | |
| 20. | } ImageLabel; | | | |

456 **Figure 2** provides examples of pose angles and their encoding (yaw, pitch) as specified in the ISO/IEC 19794-5 [ISO], with
 457 yaw angle defined as the rotation in degrees about the y-axis (vertical axis) and pitch angle defined as the rotation in
 458 degrees about the x-axis (horizontal axis).

459



460

Figure 2 – Examples of pose angles and their encodings (yaw, pitch)

461

NOTE 1: We do not intend to deliberately include non-face images in this test.

462

NOTE 2: **MULTIFACES** will contain face images of only one person.

463

464

A **MULTIFACE** (see Table 13) will be accompanied by one of the labels given in Table 11. Face recognition

465

implementations submitted to FIVE should tolerate **MULTIFACES** of any category.

466

Table 11 – Labels describing categories of MULTIFACES

| | Label as C++ enumeration | Meaning |
|----|----------------------------------|---|
| | <code>typedef enum {</code> | |
| 1. | <code>FRONTAL=0,</code> | All ONEFACES contain nominally frontal images |
| 2. | <code>MULTIPOSE=1,</code> | Contains off-angle ONEFACES . Can contain nominally frontal images. |
| 3. | <code>UNCONTROLLED=2</code> | All ONEFACES have uncontrolled or unknown pose information (i.e., Pose.isAssigned=false). |
| 4. | <code>MULTIFACE_UNKNOWN=3</code> | This label is assigned to MULTIFACES that are not FRONTAL , MULTIPOSE or INFORMAL . |
| 5. | <code>} MultifaceLabel;</code> | |

467

468 **2.3.4. Data structures for encapsulating multiple still images**

469 The standardized formats for facial images are the ISO/IEC 19794-5:2005 and the ANSI/NIST ITL 1-2011 type 10 record.
 470 The ISO record can store multiple images of an individual in a standalone binary file. In the ANSI/NIST realm, K images of
 471 an individual are usually represented as the concatenation of one Type 1 record + K Type 10 records. The result is usually
 472 stored as an EFT file.

473 An alternative method of representing K images of an individual is to define a structure containing an image filename and
 474 metadata fields. Each file contains a standardized image format, e.g. PNG (lossless) or JPEG (lossy).

475 **2.3.5. Class for encapsulating a single face image**

476 **Table 12 – ONEFACE class**

| | C++ code fragment | Remarks |
|-----|--|---|
| 1. | <code>class ONEFACE</code> | |
| 2. | <code>{</code> | |
| | <code>private:</code> | |
| 3. | <code>uint16_t imageWidth;</code> | Number of pixels horizontally |
| 4. | <code>uint16_t imageHeight;</code> | Number of pixels vertically |
| 5. | <code>uint8_t imageDepth;</code> | Number of bits per pixel. Legal values are 8 and 24. |
| 6. | <code>uint8_t format;</code> | Flag indicating native format of the image as supplied to NIST 0x01 = JPEG (i.e. compressed data) 0x02 = PNG (i.e. never compressed data) |
| 7. | <code>const uint8_t *data;</code> | Pointer to raster scanned data. Either RGB color or intensity. If image_depth == 24 this points to 3WH bytes RGBRGBRGB... If image_depth == 8 this points to WH bytes I I I I I I I I I |
| 8. | <code>Pose pose;</code> | Pose information for the image. See 2.3.2. |
| 9. | <code>public:</code> <code>//getter/setter methods</code> | |
| 10. | <code>};</code> | |

477 **2.3.6. Class for encapsulating a set of face images from a single person**

478 **Table 13 – MULTIFACE class**

| | C++ code fragment | Remarks |
|----|---|---|
| 1. | <code>class MULTIFACE</code> <code>{</code> | Vector containing F pointers of pre-allocated face images of the same person. The number of items stored in the vector is accessible via the vector::size() function. |
| | <code>private:</code> | |
| 2. | <code>std::vector<const ONEFACE*> faces;</code> <code>MultifaceLabel description;</code> | Single description of the vector of ONEFACES. The possible values for this enumeration are given in Table 11. |
| 3. | <code>public:</code> <code>//getter/setter methods</code> | |
| 4. | <code>};</code> | |

479 **2.3.7. Dictionary of terms describing ONEVIDEOS**

480 A ONEVIDEO will be accompanied by one of the labels given in Table 14, describing the density of people in the video
 481 frames. Face recognition implementations submitted to FIVE should tolerate ONEVIDEOS of any category.

482 **Table 14 – Labels describing the density of people in the video frames**

| | Label as C++ enumeration | Meaning |
|----|-----------------------------|--|
| | <code>typedef enum {</code> | |
| 1. | <code>SINGLE=0,</code> | All of the video frames contain one and only one person. Such video might arise from a TV interview or speech. An algorithm should produce one template from the ONEVIDEO. |

| | | |
|----|--------------------------------|---|
| 2. | <code>DENSITY_UNKNOWN=1</code> | <p>Video frames can contain zero or more people in each frame. Such videos might arise in a surveillance clip. The number of templates to return would be a random variable.</p> <p>In cases where an individual disappears from view, or is temporarily occluded, an algorithm might reasonably return more than one template for the individual. All templates will be searched against the enrolled data; if any result hits a gallery individual the result is either a successful mated search or a false positive in a non-mate search.</p> |
| | <code>}; Density;</code> | |

483 **2.3.8. Class for encapsulating a video sequence**

484 **Table 15 – ONEVIDEO Class**

| | C++ code fragment | Remarks |
|-----|--|--|
| 1. | <code>class ONEVIDEO</code> | |
| 2. | <code>{</code> <code>private:</code> | |
| 3. | <code>uint16_t frameWidth;</code> | Number of pixels horizontally of all frames |
| 4. | <code>uint16_t frameHeight;</code> | Number of pixels vertically of all frames |
| 5. | <code>uint8_t frameDepth;</code> | Number of bits per pixel for all frames. Legal values are 8 and 24. |
| 6. | <code>uint8_t framesPerSec;</code> | The frame rate of the video sequence. If this value is 0, the frames are sampled irregularly and perhaps infrequently from the parent video clip (e.g. manually selected frames, or just the I-frames). Such clips are derived from one camera observing one location over a short time period i.e. one "event", such that both the appearance of the person will change only due to their motion and the background will remain mostly unchanged. |
| 7. | <code>Density peopleDensity;</code> | Single description of the density of people in the video frames. The possible values for this enumeration are given in Table 14. |
| 8. | <code>std::vector<const uint8_t*> data;</code> | Vector of pointers to data from each frame in the video sequence. The number of frames (i.e. size of the vector) can be obtained by calling <code>vector::size()</code> . The i-th entry in data (ie. <code>data[i]</code>) points to <code>frame_width</code> x <code>frame_height</code> pixels of data for the i-th frame. |
| 9. | <code>public:</code> | |
| 10. | <code>//getter/setter methods</code> | |
| 11. | <code>};</code> | |

485 **2.3.9. Class representing a pair of eye coordinates**

486 The data structure for reporting person locations in video appears in Table 16. The coordinates may be useful to NIST for
487 relating spatial location to recognition success during our analysis.

488 **Table 16 – EYEPAIR Class**

| | C++ code fragment | Remarks |
|----|--|--|
| 1. | <code>class EYEPAIR</code> | Note that the left and right eyes are with respect to the subject. |
| 2. | <code>{</code> <code>private:</code> | |
| 3. | <code>bool isSet;</code> | If the eye coordinates have been computed and assigned successfully, this value should be set to true, otherwise it should be set to false. |
| 4. | <code>int16_t xLeft;</code> <code>int16_t yLeft;</code> | X and Y coordinate of the center of the subject's left eye . Out-of-range values (e.g. <code>x < 0</code> or <code>x >= width</code>) indicate the implementation believes the eye center is outside the image. |
| 5. | <code>int16_t xRight;</code> <code>int16_t yRight;</code> | X and Y coordinate of the center of the subject's right eye . Out-of-range values (e.g. <code>x < 0</code> or <code>x >= width</code>) indicate the implementation believes the eye center is outside the image. |

| | | |
|----|--------------------------------------|---|
| 6. | <code>uint16_t frameNum</code> | For video: the frame number that corresponds to the video frame from which the eye coordinates were generated. (i.e., the i-th frame from the video sequence). This field should not be set for eye coordinates for a single still image. |
| 7. | <code>public:</code> | |
| 8. | <code>//getter/setter methods</code> | |
| | <code>};</code> | |

489 **2.3.10. Data type for representing a person’s trajectory via eye coordinates from a video sequence or**
 490 **eye coordinates from image(s)**

491 **Table 17 – PersonTrajectory typedef**

| | C++ code fragment | Remarks |
|----|---|---|
| 1. | <code>typedef std::vector<EYEPAIR> PersonTrajectory;</code> | Vector of EYEPAIR (see 2.3.9) objects for video frames or image(s) where eyes were detected. This data structure should store eye coordinates for each video frame or image where eyes were detected for a particular person. For video frames or image(s) where the person’s eyes were not detected, the SDK shall not add an EYEPAIR to this data structure. If a face can be detected, but not the eyes, the implementation should nevertheless fill this data structure with $(x,y)_{LEFT} == (x,y)_{RIGHT}$ representing some point on the center of the face. |

492 **2.3.11. Class for representing a person from a video sequence or an image**

493 **Table 18 – PERSONREP Class**

| | C++ code fragment | Remarks |
|-----|--|---|
| 1. | <code>class PERSONREP</code> | |
| 2. | <code>{</code> | |
| | <code>private:</code> | |
| 3. | <code>PersonTrajectory eyeCoordinates;</code> | Data structure for capturing eye coordinates |
| 4. | <code>PersonTemplate proprietaryTemplate;</code> | PersonTemplate is a wrapper to a <code>uint8_t*</code> for capturing proprietary template data representing a person from a video sequence or an image. |
| 5. | <code>public:</code> | |
| 6. | <code>PERSONREP(const uint64_t inSize);</code> | The constructor takes a size parameter and allocates memory of <i>inSize</i> . <code>getPersonTemplatePtr()</code> should be called to access the newly allocated memory for SDK manipulation. Please note that this class will take care of all memory allocation and de-allocation of its own memory. The SDK shall not de-allocate memory created by this class. |
| 7. | <code>void pushBackEyeCoord(const EYEPAIR &eyes);</code> | This function should be used to add EYEPAIRs for the video frames or images where eye coordinates were detected. |
| 8. | <code>uint8_t* getPersonTemplatePtr();</code> | This function returns a <code>uint8_t*</code> to the template data. |
| 9. | <code>uint64_t getPersonTemplateSize() const;</code> | This function returns the size of the template data. |
| 10. | <code>//... getter methods, copy constructor, //... assignment operator</code> | |
| 11. | <code>};</code> | |

494 **2.3.12. Class for result of an identification search**

495 All identification searches shall return a candidate list of a NIST-specified length. The list shall be sorted with the most
 496 similar matching entries **listed** first with lowest rank.

497 **Table 19 – CANDIDATE Class**

| | C++ code fragment | Remarks |
|----|---|---------|
| 1. | <code>class CANDIDATE</code> | |
| 2. | <code>{</code> <code>private:</code> | |

| | | |
|----|--|---|
| 3. | <code>bool isSet;</code> | If the candidate is valid, this should be set to true. If the candidate computation failed, this should be set to false. |
| 4. | <code>uint32_t templateId;</code> | The Template ID integer from the enrollment database manifest defined in clause 2.4. |
| 5. | <code>double similarityScore;</code> | Measure of similarity between the identification template and the enrolled candidate. Higher scores mean more likelihood that the samples are of the same person. An algorithm is free to assign any value to a candidate. The distribution of values will have an impact on the appearance of a plot of false-negative and false-positive identification rates. |
| 6. | <code>public: //getter/setter methods</code> | |
| 7. | <code>};</code> | |

498 **2.3.13. Data type for representing a list of results of an identification search**

499 **Table 20 – CANDIDATELIST typedef**

| | C++ code fragment | Remarks |
|----|--|---|
| 1. | <code>typedef std::vector<CANDIDATE> CANDIDATELIST;</code> | A vector containing objects of CANDIDATE s. The CANDIDATE class is defined in section 2.3.12. |

500

501 **2.3.14. Class representing return code values**

502 **Table 21 – ReturnCode class**

| | C++ code fragment | Remarks |
|-----|---|---|
| | <code>class ReturnCode { public:</code> | |
| 1. | <code>typedef enum</code> | |
| 2. | <code>{</code> | |
| 3. | <code>Success=0,</code> | Success |
| 4. | <code>MissingConfig=1,</code> | The configuration data is missing or unreadable |
| 5. | <code>EnrollDirFailed=2,</code> | An operation on the enrollment directory failed |
| 6. | <code>InitNumData=3,</code> | The SDK can't support the number of images or videos |
| 7. | <code>InitBadDesc=4,</code> | The image descriptions are unexpected or unusable |
| 8. | <code>RefuseInput=5,</code> | Elective refusal to process this kind of input (ONEVIDEO or MULTIFACE) |
| 9. | <code>FailExtract=6,</code> | Involuntary failure to extract features |
| 10. | <code>FailTempl=7,</code> | Elective refusal to produce a template |
| 11. | <code>FailParse=8,</code> | Cannot parse input data |
| 12. | <code>FinInputData=9,</code> | Cannot locate input data |
| 13. | <code>FinTemplFormat=10,</code> | One or more template files are in an incorrect format |
| 14. | <code>IdBadTempl=11,</code> | The input template was defective |
| 15. | <code>ImgSizeNotSupported=12,</code> | Size of input image/frame not supported |
| 16. | <code>Vendor=13</code> | Vendor-defined failure |
| 17. | <code>} Status;</code> | |
| 18. | <code>ReturnCode(const Status inStatus);</code> | Constructor that takes an input parameter of a Status enum value. All of the functions that need to be implemented for the Video API return an instantiation of a ReturnCode object with a valid status value passed in as a parameter. |
| 19. | <code>Status getStatus() const;</code> | Getter method to return status value |
| 20. | <code>private:</code> | |
| 21. | <code>Status status;</code> | Member variable for storing status |
| 22. | <code>};</code> | |

503 **2.4. File structures for enrolled template collection**

504 For still image enrollment, an SDK converts a **MULTIFACE** into a template using the
 505 ImageEnrollment::generateEnrollmentTemplate() function of section 3.3.2.2. For video enrollment, an SDK converts a
 506 **ONEVIDEO** into one or more templates, using the VideoEnrollment::generateEnrollmentTemplate() of section 3.3.1.2. To
 507 support the identification functions, NIST will concatenate enrollment templates into a single large file. This file is called
 508 the EDB (for enrollment database). The EDB is a simple binary concatenation of proprietary templates. There is no
 509 header. There are no delimiters. The EDB may extend to hundreds of gigabytes in length.

510 This file will be accompanied by a manifest; this is an ASCII text file documenting the contents of the EDB. The manifest
 511 has the format shown as an example in Table 22. If the EDB contains N templates, the manifest will contain N lines. The
 512 fields are space (ASCII decimal 32) delimited. There are three fields, all containing numeric integers. Strictly speaking, the
 513 third column is redundant.

514 **Table 22 – Enrollment dataset template manifest**

| Field name | Template ID | Template Length | Position of first byte in EDB |
|---|--------------------------|--------------------------|-------------------------------|
| Datatype required | Unsigned decimal integer | Unsigned decimal integer | Unsigned decimal integer |
| Datatype length required | 4 bytes | 4 bytes | 8 bytes |
| Example lines of a manifest file appear to the right. Lines 1, 2, 3 and N appear. | 90201744 | 1024 | 0 |
| | 163232021 | 1536 | 1024 |
| | 7456433 | 512 | 2560 |
| | ... | | |
| | 183838 | 1024 | 30720000 |

515
 516 The EDB scheme avoids the file system overhead associated with storing millions of individual files.

517 **3. API Specification**

518 **3.1. Definitions**

519 As shown in Table 23, the API supports 1:N identification of video-to-video, video-to-still image, and still image-to-video,
 520 **and pose correction on still images**. The following hold:

- 521 – A still image is a picture of one and only one person. One or more such images are presented to the implementation
 522 using a **MULTIFACE** data structure.
- 523 – A video is a sequence of $F \geq 1$ frames.
- 524 – A frame is 2D still image containing $P \geq 0$ persons.
- 525 – Any person might be present in $0 \leq f \leq F$ frames, and their presence may be non-contiguous (e.g. due to occlusion).
- 526 – Different videos contain different numbers of frames and people.
- 527 – A **ONEVIDEO** container is used to represent a video. It contains a small header and pointers to F frames.
- 528 – Any person found in a video is represented by proprietary template (feature) data contained with a **PERSONREP** data
 529 structure. A proprietary template contains information from one or more frames. Internally, it might embed multiple
 530 traditional still-image templates, or it might integrate feature data by tracking a person across multiple frames.
- 531 – A **PERSONREP** structure additionally contains a trajectory indicating the location of the person in each frame.

532
 533 All of the code for the classes needed to implement the video API will be provided to implementers at
 534 <http://nigos.nist.gov:8080/five>. A single sample video has been made available at the same link. The sample video is
 535 only approximately representative of the scene and is not an extraction from the actual video data that will be used in the
 536 evaluation. It is only intended to illustrate similarities in terms of camera placement relative to the subject and people
 537 behavior. It is not intended to represent the optical properties of the actual imaging systems, particularly the spatial
 538 sampling rate, nor the compression characteristics.

539
 540 **Table 23 – API implementation requirements for FIVE**

| Function | Video-to-video | Still-to-video | Video-to-still | Pose Correction (Still images only) |
|---------------------------|---|---|---|--|
| Enroll | Videos | Videos | Stills | |
| Enrollment input datatype | ONEVIDEO | ONEVIDEO | MULTIFACE | |
| Enrollment datatype | PERSONREP | PERSONREP | PERSONREP | |
| Search | Video | Still | Video | |
| Search input datatype | ONEVIDEO | MULTIFACE | ONEVIDEO | |
| Search datatype | PERSONREP | PERSONREP | PERSONREP | |
| Search result | CANDIDATELIST | CANDIDATELIST | CANDIDATELIST | |
| API requirements | 3.3.1 + 3.3.3.2 + 3.3.4 + 3.3.6.2 | 3.3.1 + 3.3.3.2 + 3.3.5 + 3.3.6.3 | 3.3.2 + 3.3.3.3 + 3.3.4 + 3.3.6.4 | 3.3.7 |

541 **3.1.1. Video-to-video**

542 Video-to-video identification is the process of enrolling N videos and then searching the enrollment database with a
 543 search video. During identification, the SDK shall return a set of indices of candidate videos that contain people who
 544 appear in the search video.

- 545 – N templates will be generated from M enrollment videos. If no people appear in the videos, N will be 0. If many
 546 people appear in each video, we'd expect $N > M$.
- 547 – The N templates will be concatenated and finalized into a proprietary enrollment data structure.
- 548 – A **ONEVIDEO** will be converted to $S \geq 0$ identification template(s) based on the number of people detected in the
 549 video.
- 550 – Each identification template generated will be searched against the enrollment database of templates generated
 551 from the M input videos.
- 552 – We anticipate that the same person may appear in more than one enrolled video.

553 **3.1.2. Still image-to-video**

554 Still image-to-video identification is the process of enrolling N videos and then searching the enrollment database with a
 555 template produced from a **MULTIFACE** as follows:

- 556 – N templates will be generated from $1 < M \leq N$ enrollment videos.
- 557 – The N templates will be concatenated and finalized into a proprietary enrollment data structure.
- 558 – A **MULTIFACE** (still image) will be converted to an identification template.
- 559 – The identification template will be searched against the enrollment database of N templates.
- 560 – We anticipate that the same person may appear in more than one enrolled video.

561 **3.1.3. Video-to-still image**

562 Video-to-still image identification is the process of enrolling N **MULTIFACES** (see Table 13) and then searching the
 563 enrollment database with templates from persons found in a video as follows:

- 564 – N templates will be generated from N still-image **MULTIFACES**.
- 565 – The N templates will be concatenated and finalized into a proprietary enrollment data structure.
- 566 – A **ONEVIDEO** will be converted to $S \geq 0$ identification template(s) based on the number of people detected in the
 567 video.
- 568 – Each of the S identification templates will be searched separately against the enrollment database of N templates.

569 **3.1.4. Pose Correction**

570 Pose correction is the process rendering off-angle facial images to frontal facial images.

- 571 – The pose correction function maps $K \geq 1$ input faces to L frontal faces. When $L = 1$, the algorithm should render a
 572 frontal image as close as possible to ISO/IEC 19794-5 Token image geometry [ISO]. When $L > 1$, the implementation
 573 should render non-degenerate faces around Token geometry. The non-degenerate aspect is supplier-defined, but
 574 should be intended to be of utility to downstream recognition algorithms.

- 575 – Pose correction will only be applied to still images.
- 576 – Participants with pose correction capability may submit a pose correction-only SDK to FIVE.
- 577 – Pose correction implementations must link and run on the specified platform detailed in Section 1.19.
- 578 – NOTE: This API naturally supports image correction techniques unrelated to pose correction (e.g. illumination
- 579 correction, expression correction, etc). Submissions for such techniques are encouraged and welcome.

580 **3.2. 1:N Identification**

581 **3.2.1. Overview**

582 The 1:N application proceeds in two phases, enrollment and identification. The identification phase includes separate
 583 pre-search feature extraction stage, and a search stage.

584 The design reflects the following *testing* objectives for 1:N implementations.

- support distributed enrollment on multiple machines, with multiple processes running in parallel
- allow recovery after a fatal exception, and measure the number of occurrences
- allow NIST to copy enrollment data onto many machines to support parallel testing
- respect the black-box nature of biometric templates
- extend complete freedom to the provider to use arbitrary algorithms
- support measurement of duration of core function calls
- support measurement of template size

585 **Table 24 – Procedural overview of the identification test**

| Phase | # | Name | Description | Performance Metrics to be reported by NIST |
|------------|----|---------------------|--|--|
| Enrollment | E1 | Initialization | <p>For still image enrollment, give the implementation advance notice of the number of individuals and images that will be enrolled.</p> <p>Give the implementation the name of a directory where any provider-supplied configuration data will have been placed by NIST. This location will otherwise be empty.</p> <p>The implementation is permitted read-write-delete access to the enrollment directory during this phase. The implementation is permitted read-only access to the configuration directory.</p> <p>After enrollment, NIST may rename and relocate the enrollment directory - the implementation should not depend on the name of the enrollment directory.</p> | |
| | E2 | Parallel Enrollment | <p>For still image enrollment, for each of N individuals, pass multiple images to the implementation for conversion to a combined template. For video enrollment, for each of M video clips, pass multiple video frames to the implementation for generation of N templates, based on the number of people detected in the videos. The implementation will return a template to the calling application.</p> <p>The implementation is permitted read-only access to the enrollment directory during this phase. NIST's calling application will be responsible for storing all templates as binary files. These will not be available to the implementation during this enrollment phase.</p> <p>Multiple instances of the calling application may run simultaneously or sequentially. These may be executing on different computers. For still image enrollment, the same person will not be enrolled twice.</p> | <p>Statistics of the times needed to enroll an individual or video clip.</p> <p>Statistics of the sizes of created templates.</p> <p>The incidence of failed template creations.</p> |

| | | | | |
|------------|----|----------------------|---|--|
| | E3 | Finalization | <p>Permanently finalize the enrollment directory. This supports, for example, dis-interleaving of internal feature representations, writing of a manifest, indexing, tree construction, computation of statistical information over the enrollment dataset, and adaptation of the representation.</p> <p>The implementation is permitted read-write-delete access to the enrollment directory during this phase.</p> | <p>For still image enrollment, size of the enrollment database as a function of population size N and the number of images.</p> <p>Duration of this operation. The time needed to execute this function shall be reported with the preceding enrollment times.</p> |
| Pre-search | S1 | Initialization | <p>Tell the implementation the location of an enrollment directory. The implementation could look at the enrollment data.</p> <p>The implementation is permitted read-only access to the enrollment directory during this phase.</p> | <p>Statistics of the time needed for this operation.</p> |
| | S2 | Template preparation | <p>For each probe, create a template from a set of input images or one or more templates from a set of video clips. This operation will generally be conducted in a separate process invocation to step S2.</p> <p>The implementation is permitted no access to the enrollment directory during this phase.</p> <p>The result of this step is a search template.</p> | <p>Statistics of the time needed for this operation.</p> <p>Statistics of the size of the search template(s).</p> |
| Search | S3 | Initialization | <p>Tell the implementation the location of an enrollment directory. The implementation should read all or some of the enrolled data into main memory, so that searches can commence.</p> <p>The implementation is permitted read-only access to the enrollment directory during this phase.</p> | <p>Statistics of the time needed for this operation.</p> |
| | S4 | Search | <p>A template or multiple templates is searched against the enrollment database.</p> <p>The implementation is permitted read-only access to the enrollment directory during this phase.</p> | <p>Statistics of the time needed for this operation.</p> <p>Accuracy metrics - Type I + II error rates.</p> <p>Failure rates.</p> |

586 **3.3. Interfaces**

587 **3.3.1. The VideoEnrollment Interface**

588 The abstract class VideoEnrollment must be implemented by the SDK developer in a class named exactly
 589 SdkVideoEnrollment. The processing that takes place during each phase of the test is done via calls to the methods
 590 declared in the interface as pure virtual, and therefore is to be implemented by the SDK. The test driver will call these
 591 methods, handling all return values.

| | C++ code fragment | Remarks |
|----|---|---|
| 1. | class VideoEnrollment | |
| 2. | { public: | |
| 3. | virtual ReturnCode initialize(const string &configDir, const string &enrollDir, const uint32 t numVideos, uint8_t &numThreads) = 0 ; | Initialize the enrollment session. |
| 4. | virtual ReturnCode generateEnrollmentTemplate(const ONEVIDEO &inputVideo, vector<PERSONREP> &enrollTemplates) = 0; | Generate enrollment template(s) for the persons detected in the input video. This function takes a ONEVIDEO (see 2.3.7) as input and populates a vector of PERSONREP (see 2.3.11) with the number of persons detected from the video sequence. The implementation could call vector::push_back to insert into the vector. |
| 5. | // Destructor | |
| 6. | }; | |

592 **3.3.1.1. Initialization of the video enrollment session**

593 Before any enrollment feature extraction calls are made, the NIST test harness will call the initialization below for video-
594 to-video and still image-to-video.

595 **Table 25 – VideoEnrollment::initialize**

| | | |
|-------------------|--|---|
| Prototype | ReturnCode initialize(const string &configDir, const string &enrollDir, const uint32_t numVideos, uint8_t &numThreads); | |
| | | Input |
| | | Input |
| | | Output |
| Description | This function initializes the SDK under test and sets all needed parameters. This function will be called N=1 times by the NIST application immediately before any $M \geq 1$ calls to generateEnrollmentTemplate. Caution: The implementation should tolerate execution of $P > 1$ processes on the one or more machines each of which may be reading and writing to this same enrollment directory in parallel. File locking or process-specific temporary filenames would be needed to safely write content in the enrollDir. | |
| Input Parameters | configDir | A read-only directory containing any developer-supplied configuration parameters or run-time data files. |
| | enrollDir | The directory will be initially empty, but may have been initialized and populated by separate invocations of the enrollment process. When this function is called, the SDK may populate this folder in any manner it sees fit. Permissions will be read-write-delete. |
| | numVideos | The total number of videos that will be passed to the SDK for enrollment. |
| Output Parameters | numThreads | The maximum number of threads the implementation expects to spawn during a call to VideoEnrollment::generateEnrollmentTemplate(). A master thread that remains idle while the worker threads proceed should not be included in this total. Unthreaded implementations should return T = 1. |
| ReturnCode | Success | Success |
| | MissingConfig | The configuration data is missing, unreadable, or in an unexpected format. |
| | EnrollDirFailed | An operation on the enrollment directory failed (e.g. permission, space). |
| | InitNumData | The SDK cannot support the number of videos. |
| | Vendor | Vendor-defined failure |

596 **3.3.1.2. Video enrollment**

597 A **ONEVIDEO** is converted to enrollment template(s) for each person detected in the **ONEVIDEO** using the function below.

598 **Table 26 – VideoEnrollment::generateEnrollmentTemplate**

| | | |
|------------------|--|----------------------------------|
| Prototypes | ReturnCode generateEnrollmentTemplate(const ONEVIDEO &inputVideo, std::vector< PERSONREP > &enrollTemplates); | |
| | | Input |
| | | Output |
| Description | This function takes a ONEVIDEO , and outputs a vector of PERSONREP objects. If the function executes correctly (i.e. returns a ReturnCode::Success exit status), the NIST calling application will store the template. The NIST application will concatenate the templates and pass the result to the enrollment finalization function. For a video in which no persons appear, a valid output is an empty vector (i.e. size() == 0). If the function gives a non-zero exit status: <ul style="list-style-type: none"> – If the exit status is ReturnCode::FailParse, NIST will debug, otherwise – the test driver will ignore the output template (the template may have any size including zero) – the event will be counted as a failure to enroll. IMPORTANT: NIST's application writes the template to disk. The implementation must not attempt writes to the enrollment directory (nor to other resources). Any data needed during subsequent searches should be included in the template, or created from the templates during the enrollment finalization function. | |
| Input Parameters | inputVideo | An instance of a Table 15 class. |

| | | |
|-------------------|---------------------|---|
| Output Parameters | enrollTemplates | For each person detected in the ONEVIDEO , the function shall identify the person’s estimated eye centers for each video frame where the person’s eye coordinates can be calculated. The eye coordinates shall be captured in the PERSONREP .eyeCoordinates variable, which is a vector of EYEPAIR objects. The frame number from the video of where the eye coordinates were detected shall be captured in the EYEPAIR .frameNum variable for each pair of eye coordinates. In the event the eye centers cannot be calculated (ie. the person becomes out of sight for a few frames in the video), the SDK shall not store an EYEPAIR for those frames. |
| ReturnCode | Success | Success |
| | RefuseInput | Elective refusal to process this kind of ONEVIDEO |
| | FailExtract | Involuntary failure to extract features (e.g. could not find face in the input-image) |
| | FailTempl | Elective refusal to produce a template (e.g. insufficient pixels between the eyes) |
| | FailParse | Cannot parse input data (i.e. assertion that input record is non-conformant) |
| | ImgSizeNotSupported | Input image/frame size too small or large |
| | Vendor | Vendor-defined failure. Failure codes must be documented and communicated to NIST with the submission of the implementation under test. |

599 **3.3.2. The ImageEnrollment Interface**

600 The abstract class ImageEnrollment must be implemented by the SDK developer in a class named exactly
601 SdkImageEnrollment.

| | C++ code fragment | Remarks |
|----|---|---|
| 1. | class ImageEnrollment | |
| 2. | { public: | |
| 3. | virtual ReturnCode initialize(const string &configDir, const string &enrollDir, const uint32_t numPersons, const uint32_t numImages const vector<ImageLabel> &descriptions) = 0 ; | Initialize the enrollment session. |
| 4. | virtual ReturnCode generateEnrollmentTemplate(const MULTIFACE &inputFaces, PERSONREP &outputTemplate) = 0; | This function takes a MULTIFACE (see 2.3.4) as input and outputs a proprietary template represented by a PERSONREP (see 2.3.11). For each input image in the MULTIFACE , the function shall return the estimated eye centers by setting PERSONREP .eyeCoordinates. |
| 5. | // Destructor | |
| 6. | }; | |

602 **3.3.2.1. Initialization of the image enrollment session**

603 Before any enrollment feature extraction calls are made, the NIST test harness will call the initialization below for video-
604 to-still.

605 **Table 27 – ImageEnrollment::initialize**

| | | |
|-------------|--|-------|
| Prototype | ReturnCode initialize(const string &configDir, const string &enrollDir, const uint32_t numPersons, const uint32_t numImages const std::vector<ImageLabel> &descriptions); | |
| | | Input |
| | | Input |
| | | Input |
| | | Input |
| Description | This function initializes the SDK under test and sets all needed parameters. This function will be called N=1 times by the NIST application immediately before any M ≥ 1 calls to generateEnrollmentTemplate. Caution: | |

| | | |
|-------------------|--|---|
| | The implementation should tolerate execution of $P > 1$ processes on the one or more machines each of which may be reading and writing to this same enrollment directory in parallel. File locking or process-specific temporary filenames would be needed to safely write content in the enrollDir. | |
| Input Parameters | configDir | A read-only directory containing any developer-supplied configuration parameters or run-time data files. |
| | enrollDir | The directory will be initially empty, but may have been initialized and populated by separate invocations of the enrollment process. When this function is called, the SDK may populate this folder in any manner it sees fit. Permissions will be read-write-delete. |
| | numPersons | The number of persons who will be enrolled. |
| | numImages | The total number of images that will be enrolled, summed over all identities. |
| | descriptions | A vector of labels one of which will be assigned to each enrollment image. See Table 10 for valid values. NOTE: The identification search images may or may not be labeled. An identification image may carry a label not in this set of labels. The number of items stored in the vector is accessible via the vector::size() function. |
| Output Parameters | none | |
| ReturnCode | Success | Success |
| | MissingConfig | The configuration data is missing, unreadable, or in an unexpected format. |
| | EnrollDirFailed | An operation on the enrollment directory failed (e.g. permission, space). |
| | InitNumData | The SDK cannot support the number of videos. |
| | InitBadDesc | The descriptions are unexpected, or unusable. |
| | Vendor | Vendor-defined failure |

606 **3.3.2.2. Image enrollment**

607 A **MULTIFACE** (see Table 13) is converted to a single enrollment template using the function below.

608 **Table 28 – ImageEnrollment::generateEnrollmentTemplate**

| | | |
|-------------------|---|---|
| Prototypes | ReturnCode generateEnrollmentTemplate(const MULTIFACE &inputFaces, PERSONREP &outputTemplate); | Input Output |
| Description | <p>This function takes a MULTIFACE, and outputs a proprietary template in the form of a PERSONREP object. If the function executes correctly (i.e. returns a ReturnCode::Success exit status), the NIST calling application will store the template. The NIST application will concatenate the templates and pass the result to the enrollment finalization function.</p> <p>If the function gives a non-zero exit status:</p> <ul style="list-style-type: none"> – If the exit status is ReturnCode::FailParse, NIST will debug, otherwise – the test driver will ignore the output template (the template may have any size including zero) – the event will be counted as a failure to enroll. Such an event means that this person can never be identified correctly. <p>IMPORTANT. NIST's application writes the template to disk. The implementation must not attempt writes to the enrollment directory (nor to other resources). Any data needed during subsequent searches should be included in the template, or created from the templates during the enrollment finalization function.</p> | |
| Input Parameters | inputFaces | An instance of a Table 13 structure. |
| Output Parameters | outputTemplate | An instance of a section 2.3.11 class, which stores proprietary template data and eye coordinates. The function shall identify the person's estimated eye centers for each image in the MULTIFACE . The eye coordinates shall be captured in the PERSONREP .eyeCoordinates variable, which is a vector of EYEPAIR objects. In the event the eye centers cannot be calculated, the SDK shall store an EYEPAIR and set EYEPAIR .isSet to false to indicate there was a failure in generating eye coordinates. In other words, for N images in the MULTIFACE . |

FIVE

| | | |
|------------|---------------------|---|
| ReturnCode | Success | Success |
| | RefuseInput | Elective refusal to process this kind of MULTIFACE |
| | FailExtract | Involuntary failure to extract features (e.g. could not find face in the input-image) |
| | FailTempl | Elective refusal to produce a template (e.g. insufficient pixels between the eyes) |
| | FailParse | Cannot parse input data (i.e. assertion that input record is non-conformant) |
| | ImgSizeNotSupported | Input image/frame size too small or large |
| | Vendor | Vendor-defined failure. Failure codes must be documented and communicated to NIST with the submission of the implementation under test. |

609

610

611 **3.3.3. The Finalize Interface**

612 The abstract class Finalize must be implemented by the SDK developer in classes named exactly ImageGalleryFinalize and
 613 VideoGalleryFinalize. The finalize function in this class takes the name of the top-level directory where enrollment
 614 database (EDB) and its manifest have been stored. These are described in section 2.3.7. The enrollment directory
 615 permissions will be read + write.
 616

| | C++ code fragment | Remarks |
|----|---|--|
| 1. | class Finalize | |
| 2. | { public: | |
| 3. | virtual ReturnCode finalize(const string &enrollDir, const string &edbName, const string &edbManifest) = 0; | This function supports post-enrollment developer-optional book-keeping operations and statistical processing. The function will generally be called in a separate process after all the enrollment processes are complete. |
| 4. | // Destructor | |
| 5. | }; | |

617

618 **3.3.3.1. Finalize enrollment**

619 After all templates have been created, the function of Table 29 will be called. This freezes the enrollment data. After this
 620 call the enrollment dataset will be forever read-only. This API does not support interleaved enrollment and search
 621 phases.

622 The function allows the implementation to conduct, for example, statistical processing of the feature data, indexing and
 623 data re-organization. The function may alter the file structure. It may increase or decrease the size of the stored data.
 624 No output is expected from this function, except a return code.

625

Table 29 – Finalize::finalize

| | | |
|-------------------|---|--|
| Prototypes | ReturnCode finalize (const string &enrollDir, const string &edbName, const string &edbManifest); | |
| | | Input |
| | | Input |
| Description | This function takes the name of the top-level directory where enrollment database (EDB) and its manifest have been stored. These are described in section 2.3.7. The enrollment directory permissions will be read + write. | |
| | The function supports post-enrollment developer-optional book-keeping operations and statistical processing. The function will generally be called in a separate process after all the enrollment processes are complete. | |
| | This function should be tolerant of being called two or more times. Second and third invocations should probably do nothing. | |
| Input Parameters | enrollDir | The top-level directory in which enrollment data was placed. This variable allows an implementation to locate any private initialization data it elected to place in the directory. |
| | edbName | The name of a single file containing concatenated templates, i.e. the EDB of section 2.3.7. While the file will have read-write-delete permission, the SDK should only alter the file if it preserves the necessary content, in other files for example. The file may be opened directly. It is not necessary to prepend a directory name. |
| | edbManifest | The name of a single file containing the EDB manifest of section 2.3.7. The file may be opened directly. It is not necessary to prepend a directory name. |
| Output Parameters | None | |
| ReturnCode | Success | Success |
| | FinInputData | Cannot locate the input data - the input files or names seem incorrect. |
| | EnrollDirFailed | An operation on the enrollment directory failed (e.g. permission, space). |
| | FinTemplFormat | One or more template files are in an incorrect format. |

| | | |
|--|--------|---|
| | Vendor | Vendor-defined failure. Failure codes must be documented and communicated to NIST with the submission of the implementation under test. |
|--|--------|---|

626
627

3.3.3.2. Finalize video enrollment - VideoGalleryFinalize

| | C++ code fragment | Remarks |
|----|---|--|
| 1. | <code>class VideoGalleryFinalize : public Finalize</code> | |
| 2. | <code>{</code> <code>public:</code> | |
| 3. | <code> ReturnCode finalize(const string &enrollDir, const string &edbName, const string &edbManifest);</code> | This function supports post-video-enrollment developer-optional book-keeping operations and statistical processing. The function will generally be called in a separate process after all the enrollment processes are complete. |
| 4. | <code> // Constructor/Destructor/Other</code> | |
| 5. | <code>};</code> | |

628
629

3.3.3.3. Finalize still image enrollment - ImageGalleryFinalize

| | C++ code fragment | Remarks |
|----|---|--|
| 1. | <code>class ImageGalleryFinalize : public Finalize</code> | |
| 2. | <code>{</code> <code>public:</code> | |
| 3. | <code> ReturnCode finalize(const string &enrollDir, const string &edbName, const string &edbManifest);</code> | This function supports post-still-image-enrollment developer-optional book-keeping operations and statistical processing. The function will generally be called in a separate process after all the enrollment processes are complete. |
| 4. | <code> // Constructor/Destructor/Other</code> | |
| 5. | <code>};</code> | |

630

3.3.4. The VideoFeatureExtraction Interface

631 The abstract class VideoFeatureExtraction must be implemented by the SDK developer in a class named exactly
632 SdkVideoFeatureExtraction.

| | C++ code fragment | Remarks |
|----|---|---|
| 1. | <code>class VideoFeatureExtraction</code> | |
| 2. | <code>{</code> <code>public:</code> | |
| 3. | <code> virtual ReturnCode initialize(const string &configDir, const string &enrollDir, uint8_t &numThreads) = 0;</code> | Initialize the feature extraction session. |
| 4. | <code> virtual ReturnCode generateIdTemplate(const ONEVIDEO &inputVideo, vector<PERSONREP> &idTemplates) = 0;</code> | Generate identification template(s) for the persons detected in the input video. This function takes a ONEVIDEO (see 2.3.7) as input and populates a vector of PERSONREP (see 2.3.11) with the number of persons detected from the video sequence. The implementation could call vector::push_back to insert into the vector. |
| 5. | <code> // Destructor</code> | |
| 6. | <code>};</code> | |

633 **3.3.4.1. Video feature extraction initialization**

634 Before one or more **ONEVIDEO**s are sent to the identification feature extraction function, the test harness will call the
 635 initialization function below.

636 **Table 30 – VideoFeatureExtraction::initialize**

| | | |
|-------------------|--|--|
| Prototype | ReturnCode initialize(const string &configDir, const string &enrollDir, uint8_t &numThreads); | |
| | | Input |
| | | Output |
| Description | <p>This function initializes the SDK under test and sets all needed parameters. This function will be called once by the NIST application immediately before any $M \geq 1$ calls to generateIdTemplate.</p> <p>The implementation has read-only access to enrollDir (containing prior enrollment data) and to configDir.</p> | |
| Input Parameters | configDir | A read-only directory containing any developer-supplied configuration parameters or run-time data files. |
| | enrollDir | The read-only top-level directory in which enrollment data was placed and then finalized by the implementation. The implementation can parameterize subsequent template production on the basis of the enrolled dataset. |
| Output Parameters | numThreads | <p>The maximum number of threads the implementation expects to spawn during a call to VideoFeatureExtraction::generateIdTemplate(). A master thread that remains idle while the worker threads proceed should not be included in this total.</p> <p>Unthreaded implementations should return $T = 1$.</p> |
| ReturnCode | Success | Success |
| | MissingConfig | The configuration data is missing, unreadable, or in an unexpected format. |
| | EnrollDirFailed | An operation on the enrollment directory failed (e.g. permission). |
| | Vendor | Vendor-defined failure |

637 **3.3.4.2. Video feature extraction**

638 A **ONEVIDEO** is converted to one or more identification templates using the function below. The result may be stored by
 639 NIST, or used immediately. The SDK shall not attempt to store any data.

640 **Table 31 – VideoFeatureExtraction::generateIdTemplate**

| | | |
|-------------------|--|---|
| Prototypes | ReturnCode generateIdTemplate(const ONEVIDEO &inputVideo, std::vector< PERSONREP > &idTemplates); | |
| | | Input |
| | | Output |
| Description | <p>This function takes a ONEVIDEO (see 2.3.7) as input and populates a vector of PERSONREP (see 2.3.11) with the number of persons detected from the video sequence. The implementation could call vector::push_back to insert into the vector.</p> <p>If the function executes correctly, it returns a zero exit status. The NIST calling application may commit the template to permanent storage, or may keep it only in memory (the implementation does not need to know). If the function returns a non-zero exit status, the output template will be not be used in subsequent search operations.</p> <p>The function shall not have access to the enrollment data, nor shall it attempt access.</p> | |
| Input Parameters | InputVideo | An instance of a section 2.3.7 class. Implementations must alter their behavior according to the people detected in the video sequence. |
| Output Parameters | IdTemplates | For each person detected in the video, the function shall create a PERSONREP (see section 2.3.11) object, populate it with a template and eye coordinates for each frame where eyes were detected, and add it to the vector. |
| ReturnCode | Success | Success |
| | RefuseInput | Elective refusal to process this kind of ONEVIDEO |
| | FailExtract | Involuntary failure to extract features (e.g. could not find face in the input-image) |
| | FailTempl | Elective refusal to produce a template (e.g. insufficient pixels between the eyes) |

| | | |
|--|---------------------|---|
| | FailParse | Cannot parse input data (i.e. assertion that input record is non-conformant) |
| | ImgSizeNotSupported | Input image/frame size too small or large |
| | Vendor | Vendor-defined failure. Failure codes must be documented and communicated to NIST with the submission of the implementation under test. |

641 **3.3.5. The ImageFeatureExtraction Interface**

642 The abstract class ImageFeatureExtraction must be implemented by the SDK developer in a class named exactly
 643 SdkImageFeatureExtraction.

| | C++ code fragment | Remarks |
|----|---|---|
| 1. | class ImageFeatureExtraction | |
| 2. | { | |
| | public: | |
| 3. | virtual ReturnCode initialize(const string &configDir, const string &enrollDir) = 0; | Initialize the feature extraction session. |
| 4. | virtual ReturnCode generateIdTemplate(const MULTIFACE &inputFaces, PERSONREP &outputTemplate) = 0; | This function takes a MULTIFACE (see 2.3.4) as input and outputs a proprietary template represented by a PERSONREP (see 2.3.11). For each input image in the MULTIFACE , the function shall return the estimated eye centers by setting PERSONREP .eyeCoordinates. |
| 5. | // Destructor | |
| 6. | }; | |

644 **3.3.5.1. Image feature extraction initialization**

645 Before one or more **MULTIFACE**s are sent to the identification feature extraction function, the test harness will call the
 646 initialization function below.

647 **Table 32 – ImageFeatureExtraction::initialize**

| | | |
|-------------------|---|--|
| Prototype | ReturnCode initialize(const string &configDir, const string &enrollDir); | Input Input |
| Description | This function initializes the SDK under test and sets all needed parameters. This function will be called once by the NIST application immediately before $M \geq 1$ calls to generateIdTemplate. The implementation has read-only access to enrollDir (containing prior enrollment data) and to configDir. | |
| Input Parameters | configDir | A read-only directory containing any developer-supplied configuration parameters or run-time data files. |
| | enrollDir | The read-only top-level directory in which enrollment data was placed and then finalized by the implementation. The implementation can parameterize subsequent template production on the basis of the enrolled dataset. |
| Output Parameters | none | |
| ReturnCode | Success | Success |
| | MissingConfig | The configuration data is missing, unreadable, or in an unexpected format. |
| | EnrollDirFailed | An operation on the enrollment directory failed (e.g. permission). |
| | Vendor | Vendor-defined failure |

648 **3.3.5.2. Image feature extraction**

649 A **MULTIFACE** is converted to one identification template using the function below. The result may be stored by NIST, or
 650 used immediately. The SDK shall not attempt to store any data.

651 **Table 33 – ImageFeatureExtraction::generateIdTemplate**

| | | |
|-------------------|--|---|
| Prototypes | ReturnCode generateIdTemplate(const MULTIFACE &inputFaces, PERSONREP &outputTemplate); | |
| | | Input |
| | | Output |
| Description | <p>This function takes a MULTIFACE (see 2.3.4) as input and populates a PERSONREP (see 2.3.11) with a proprietary template and eye coordinates.</p> <p>If the function executes correctly, it returns a zero exit status. The NIST calling application may commit the template to permanent storage, or may keep it only in memory (the developer implementation does not need to know). If the function returns a non-zero exit status, the output template will be not be used in subsequent search operations.</p> <p>The function shall not have access to the enrollment data, nor shall it attempt access.</p> | |
| Input Parameters | inputFaces | An instance of a Table 13 structure. |
| Output Parameters | outputTemplate | An instance of a section 2.3.11 class, which stores proprietary template data and eye coordinates. The function shall identify the person’s estimated eye centers for each image in the MULTIFACE . The eye coordinates shall be captured in the PERSONREP .eyeCoordinates variable, which is a vector of EYEPAIR objects. In the event the eye centers cannot be calculated, the SDK shall store an EYEPAIR and set EYEPAIR .isSet to false to indicate there was a failure in generating eye coordinates. In other words, for N images in the MULTIFACE . |
| ReturnCode | Success | Success |
| | RefuseInput | Elective refusal to process this kind of MULTIFACE |
| | FailExtract | Involuntary failure to extract features (e.g. could not find face in the input-image) |
| | FailTempl | Elective refusal to produce a template (e.g. insufficient pixels between the eyes) |
| | FailParse | Cannot parse input data (i.e. assertion that input record is non-conformant) |
| | ImgSizeNotSupported | Input image/frame size too small or large |
| | Vendor | Vendor-defined failure. Failure codes must be documented and communicated to NIST with the submission of the implementation under test. |

652 **3.3.6. The Search Interface**

653 The abstract class Search must be implemented by the SDK developer in classes named exactly VideoToVideoSearch,
654 StillToVideoSearch, and VideoToStillSearch.
655

| | C++ code fragment | Remarks |
|----|---|--|
| 1. | <code>class Search</code> | |
| 2. | <code>{</code> <code>public:</code> | |
| 3. | <code>virtual ReturnCode initialize(const string &configDir, const string &enrollDir, uint8_t &numThreads) = 0;</code> | Initialize the search session. |
| 4. | <code>virtual ReturnCode identify(const PERSONREP &idTemplate, const uint32_t candListLength, CANDIDATELIST &candList) = 0;</code> | This function searches a template against the enrollment set, and outputs a vector containing candListLength objects of Candidates (see section 2.3.13). |
| 5. | <code>// Destructor</code> | |
| 6. | <code>};</code> | |

656 **3.3.6.1. Search initialization**

657 The function below will be called once prior to one or more calls of the searching function of **Table 36**. The function might
658 set static internal variables so that the enrollment database is available to the subsequent identification searches.

659 **Table 34 – Search::initialize**

| | |
|-----------|--|
| Prototype | ReturnCode initialize(const string &configDir, |
| | Input |

| | | |
|-------------------|--|--|
| | const string &enrollDir, uint8_t &numThreads); | Input Output |
| Description | This function reads whatever content is present in the enroll_dir, for example a manifest placed there by the Finalize::finalize function. | |
| Input Parameters | configDir | A read-only directory containing any developer-supplied configuration parameters or run-time data files. |
| | enrollDir | The read-only top-level directory in which enrollment data was placed. |
| Output Parameters | numThreads | The maximum number of threads the implementation expects to spawn during a call to Search::identify(). A master thread that remains idle while the worker threads proceed should not be included in this total. Unthreaded implementations should return T = 1. |
| | | |
| ReturnCode | Success | Success |
| | MissingConfig | The configuration data is missing, unreadable, or in an unexpected format. |
| | EnrollDirFailed | An operation on the enrollment directory failed (e.g. permission). |
| | Vendor | Vendor-defined failure |

660
661

3.3.6.2. Video-to-video search

| | C++ code fragment | Remarks |
|----|---|--|
| 1. | class VideoToVideoSearch : public Search | |
| 2. | { public: | |
| 3. | ReturnCode initialize(const string &configDir, const string &enrollDir, uint8_t &numThreads); | Initialize the search session for video-to-video search. |
| 4. | ReturnCode identify(const PERSONREP &idTemplate, const uint32_t candListLength, CANDIDATELIST &candList); | This function searches a template generated from a ONEVIDEO against the enrollment set, and outputs a vector containing candListLength objects of Candidates (see section 2.3.13). |
| 5. | // Constructor/Destructor/Other | |
| 6. | }; | |

662

3.3.6.2.1. Video-to-video identification

663
664
665
666

The function below compares a proprietary identification template against the enrollment data and returns a candidate list.

Table 35 – VideoToVideoSearch::identify

| | | |
|-------------------|---|--|
| Prototype | ReturnCode identify(const PERSONREP &idTemplate, const uint32_t candListLength, CANDIDATELIST &candList); | Searches a template generated from a ONEVIDEO against the enrollment set (video-to-video) Input Input Output |
| Description | This function searches an identification template against the enrollment set, and outputs a vector containing candListLength Candidates (see section 2.3.13). Each candidate shall be populated by the implementation and added to candList. Note that candList will be an empty vector when passed into this function. The candidates shall appear in descending order of similarity score - i.e. most similar entries appear first. | |
| Input Parameters | idTemplate | A template from generateIdTemplate() - If the value returned by that function was non-zero the contents of idTemplate will not be used and this function (i.e. identify) will not be called. |
| | candListLength | The number of candidates the search should return |
| Output Parameters | candList | A vector containing candListLength objects of Candidates. The datatype is defined in section 2.3.13. Each candidate shall be populated by the implementation and added to this vector. The |

| | | |
|------------|------------|---|
| | | candidates shall appear in descending order of similarity score - i.e. most similar entries appear first. |
| ReturnCode | Success | Success |
| | IdBadTempl | The input template was defective. |
| | Vendor | Vendor-defined failure |

667

3.3.6.3. Still-to-video search

| | C++ code fragment | Remarks |
|----|---|---|
| 1. | <code>class StillToVideoSearch : public Search</code> | |
| 2. | <code>{ public:</code> | |
| 3. | <code> ReturnCode initialize(const string &configDir, const string &enrollDir, uint8_t &numThreads);</code> | Initialize the search session for still-to-video search. |
| 4. | <code> ReturnCode identify(const PERSONREP &idTemplate, const uint32_t candListLength, CANDIDATELIST &candList);</code> | This function searches a template generated from a MULTIFACE against the enrollment set, and outputs a vector containing candListLength objects of Candidates. |
| 5. | <code> // Constructor/Destructor/Other</code> | |
| 6. | <code>};</code> | |

668

3.3.6.3.1. Still-to-video identification

The function below compares a proprietary identification template against the enrollment data and returns a candidate list.

670

671

672

Table 36 – StillToVideoSearch::identify

| | | |
|-------------------|---|--|
| Prototype | ReturnCode identify(const PERSONREP &idTemplate, const uint32_t candListLength, CANDIDATELIST &candList); | Searches a template generated from a MULTIFACE against the enrollment set (still-to-video) Input Input Output |
| Description | This function searches an identification template against the enrollment set, and outputs a vector containing candListLength Candidates (see section 2.3.13). Each candidate shall be populated by the implementation and added to candList. Note that candList will be an empty vector when passed into this function. The candidates shall appear in descending order of similarity score - i.e. most similar entries appear first. | |
| Input Parameters | idTemplate | A template from generateIdTemplate() - If the value returned by that function was non-zero the contents of idTemplate will not be used and this function (i.e. identify) will not be called. |
| | candListLength | The number of candidates the search should return |
| Output Parameters | candList | A vector containing candListLength objects of Candidates. The datatype is defined in section 2.3.13. Each candidate shall be populated by the implementation and added to this vector. The candidates shall appear in descending order of similarity score - i.e. most similar entries appear first. |
| ReturnCode | Success | Success |
| | IdBadTempl | The input template was defective. |
| | Vendor | Vendor-defined failure |

673

3.3.6.4. Video-to-still search

674

| | C++ code fragment | Remarks |
|----|---|---------|
| 1. | <code>class VideoToStillSearch : public Search</code> | |
| 2. | <code>{ public:</code> | |

| | | |
|----|---|---|
| 3. | <code>ReturnCode initialize(const string &configDir, const string &enrollDir, uint8_t &numThreads);</code> | Initialize the search session for video-to-still search. |
| 4. | <code>ReturnCode identify(const PERSONREP &idTemplate, const uint32_t candListLength, CANDIDATELIST &candList);</code> | This function searches a template generated from a ONEVIDEO against the enrollment set, and outputs a vector containing <code>candListLength</code> objects of Candidates (see section 2.3.13). |
| 5. | <code>// Constructor/Destructor/Other</code> | |
| 6. | <code>};</code> | |

675

676 **3.3.6.4.1. Video-to-still identification**

677 The function below compares a proprietary identification template against the enrollment data and returns a candidate
678 list.

679

Table 37 – VideoToStillSearch::identify

| | | |
|-------------------|--|---|
| Prototype | <code>ReturnCode identify(const PERSONREP &idTemplate, const uint32_t candListLength, CANDIDATELIST &candList);</code> | Searches a template generated from a ONEVIDEO against the enrollment set (video-to-still) |
| | | Input |
| | | Input |
| | | Output |
| Description | This function searches an identification template against the enrollment set, and outputs a vector containing <code>candListLength</code> Candidates (see section 2.3.13). Each candidate shall be populated by the implementation and added to <code>candList</code> . Note that <code>candList</code> will be an empty vector when passed into this function. The candidates shall appear in descending order of similarity score - i.e. most similar entries appear first. | |
| Input Parameters | <code>idTemplate</code> | A template from <code>generateIdTemplate()</code> - If the value returned by that function was non-zero the contents of <code>idTemplate</code> will not be used and this function (i.e. <code>identify</code>) will not be called. |
| | <code>candListLength</code> | The number of candidates the search should return |
| Output Parameters | <code>candList</code> | A vector containing <code>candListLength</code> objects of Candidates . The datatype is defined in section 2.3.13. Each candidate shall be populated by the implementation and added to this vector. The candidates shall appear in descending order of similarity score - i.e. most similar entries appear first. |
| ReturnCode | Success | Success |
| | <code>IdBadTempl</code> | The input template was defective. |
| | <code>Vendor</code> | Vendor-defined failure |

680 NOTE: Ordinarily the calling application will set the input candidate list length to operationally typical values, say $0 \leq L \leq$
681 200, and $L \ll N$. However, there is interest in the presence of mates much further down the candidate list. We may
682 therefore extend the candidate list length such that L approaches N .

683 **3.3.7. The PoseCorrection Interface**

684 The abstract class **PoseCorrection** must be implemented by the SDK developer in a class named exactly
685 **SdkPoseCorrection**.
686

| | C++ code fragment | Remarks |
|----|---|---|
| 1. | <code>class PoseCorrection</code> | |
| 2. | <code>{ public:</code> | |
| 3. | <code>virtual ReturnCode initialize(const string &configDir, const vector<string> &descriptions, uint32_t &maxOutImages) = 0;</code> | Initialize the pose correction session. |

| | | |
|----|--|---|
| 4. | virtual ReturnCode reconstruct(const MULTIFACE &inputFaces, const uint32_t maxOutImages, MULTIFACE &outputFaces, uint32_t &numOutImages) = 0; | Take a MULTIFACE containing K images of an individual and output $1 \leq L \leq \text{maxOutImages}$ pose-corrected faces in a MULTIFACE structure. |
| 5. | // Destructor | |
| 6. | }; | |

687 **3.3.7.1. Pose correction initialization**

688 Before any template generation or matching calls are made, the NIST test harness will make a call to the initialization of
689 the function in Table 38.

690 **Table 38 – Pose correction initialization**

| | | |
|--------------------------|--|--|
| Prototype | ReturnCode initialize(const string &configDir, const std::vector<ImageLabel> &descriptions uint32_t &maxOutImages); | |
| | | Input |
| | | Input |
| | | Output |
| Description | This function initializes the SDK under test. It will be called by the NIST application before any reconstruction calls. The SDK under test should set all parameters. | |
| Input Parameters | configDir | A read-only directory containing any developer-supplied configuration parameters or run-time data files. The name of this directory is assigned by NIST. It is not hardwired by the provider. The names of the files here are hardwired in the SDK and are unrestricted. |
| | descriptions | A vector of labels one of which will be assigned to each image. See Table 10 for valid values. |
| Output Parameters | maxOutImages | The maximum number of images that the frontal reconstruction algorithms will output – see below. |
| ReturnCode | Success | Success |
| | MissingConfig | Vendor provided configuration files are not readable in the indicated location. |
| | InitBadDesc | The descriptions are unexpected, or unusable. |
| | Vendor | Vendor-defined failure |

691 **3.3.7.2. Pose Correction**

692 The function of Table 39 maps $K \geq 1$ input faces to L frontal faces. When $L = 1$, the algorithm should render a frontal
693 image as close as possible to ISO/IEC 19794-5 Token image geometry [ISO]. When $L > 1$, the implementation should
694 render non-degenerate faces around Token geometry. The non-degenerate aspect is supplier-defined, but should be
695 intended to be of utility to downstream recognition algorithms.

696 **Table 39 – Pose Correction**

| | | |
|-------------------------|---|---|
| Prototypes | int32_t reconstruct(const MULTIFACE &inputFaces, const uint32_t maxOutImages, MULTIFACE &outputFaces, uint32_t &numOutImages); | |
| | | Input |
| | | Input |
| | | Output |
| | | Output |
| Description | This function takes a MULTIFACE containing K images of an individual. It outputs $1 \leq L \leq \text{maxOutImages}$ output faces in a MULTIFACE structure. | |
| Input Parameters | inputFaces | An instance of a Table 12 Class for encapsulating a set of face images from a single person Table 13 Implementations must alter their behavior according to the number of images contained in the structure. |
| | maxOutImages | The number of output faces requested by the calling application. The implementation must support a call with $\text{maxOutImages} == 1$. This will form a baseline result. NIST will additionally report results with larger values $1 < \text{maxOutImages} \leq 9$. The upper bound here would allow the algorithm to render left, left-up, left-down, right, right-up, right-down, frontal, up, down |

FIVE

| | | |
|-------------------|--------------|--|
| | | variants around frontal. The implementation does not need to support values $1 < \text{maxOutImages}$. |
| Output Parameters | outputFaces | <p>A MULTIFACE object with data pre-allocated for maxOutImages entries each of size 640 height by 480 width by 24 bits (RGB). These dimensions afford 120 pixels between the eyes for a Token geometry output. Images smaller than this could be centered with a grey border.</p> <p>This prescription of height and width allows the NIST application to allocate all memory. The implementation should not allocate memory for the output MULTIFACE.</p> <p>Implementers seeking pre-allocated sizes larger than 640x480 should contact NIST.</p> |
| | numOutImages | <p>$0 \leq L \leq \text{maxOutImages}$. The number of faces actually produced. These faces must occupy the first L positions of the output MULTIFACE.faces vector.</p> <p>If 0 faces are rendered, the ReturnCode must be non-zero (i.e. not Success)</p> |
| ReturnCode | Success | Success |
| | RefuseInput | Elective refusal to process this kind of MULTIFACE |
| | FailExtract | Involuntary failure to extract features (e.g. could not find face in the input-image) |
| | FailTempl | Elective refusal to render any output images. |
| | FailParse | Cannot parse input data (i.e. assertion that input record is non-conformant) |
| | Vendor | Vendor-defined failure. Failure codes must be documented and communicated to NIST with the submission of the implementation under test. |

698 **4. References**

| | |
|-----------------|--|
| AN27 | NIST Special Publication 500-271: American National Standard for Information Systems — Data Format for the Interchange of Fingerprint, Facial, & Other Biometric Information – Part 1. (ANSI/NIST ITL 1-2007). Approved April 20, 2007. |
| FRVT 2002 | Face Recognition Vendor Test 2002: Evaluation Report, NIST Interagency Report 6965, P. Jonathon Phillips, Patrick Grother, Ross J. Micheals, Duane M. Blackburn, Elham Tabassi, Mike Bone |
| FRVT 2002b | Face Recognition Vendor Test 2002: Supplemental Report, NIST Interagency Report 7083, Patrick Grother |
| FRVT 2006 | P. Jonathon Phillips, W. Todd Scruggs, Alice J. O'Toole, Patrick J. Flynn, Kevin W. Bowyer, Cathy L. Schott, and Matthew Sharpe. "FRVT 2006 and ICE 2006 Large-Scale Results." NISTIR 7408, March 2007. |
| FRVT 2013 | P. Grother and M. Ngan, Face Recognition Vendor Test (FRVT), Performance of Face Identification Algorithms, NIST Interagency Report 8009, Released May 26, 2014. http://face.nist.gov/frvt |
| IREX III | P. Grother, G.W. Quinn, J. Matey, M. Ngan, W. Salamon, G. Fiumara, C. Watson, Iris Exchange III, Performance of Iris Identification Algorithms, NIST Interagency Report 7836, Released April 9, 2012. http://iris.nist.gov/irex |
| ISO STD05 | ISO/IEC 19794-5:2005 — Information technology — Biometric data interchange formats — Part 5: Face image data. The standard was published in 2005, and can be purchased from ANSI at http://webstore.ansi.org/ Multipart standard of "Biometric data interchange formats". This standard was published in 2005. It was amended twice to include guidance to photographers, and then to include 3D information. Two corrigenda were published. All these changes and new material is currently being incorporated in revision of the standard. Publication is likely in early 2011. The documentary history is as follows. ISO/IEC 19794-5: Information technology — Biometric data interchange formats — Part 5:Face image data. First edition: 2005-06-15. International Standard ISO/IEC 19794-5:2005 Technical Corrigendum 1: Published 2008-07-01 International Standard ISO/IEC 19794-5:2005 Technical Corrigendum 2: Published 2008-07-01 Information technology — Biometric data interchange formats — Part 5: Face image data AMENDMENT 1: Conditions for taking photographs for face image data. Published 2007-12-15 Information technology — Biometric data interchange formats — Part 5: Face image data AMENDMENT 2: Three dimensional image data. JTC 1/SC37/N3303. FCD text of the second edition. Contact pgrother AT nist DOT gov for more information. |
| MBE | P. Grother, G .W. Quinn, and P. J. Phillips, Multiple-Biometric Evaluation (MBE) 2010, Report on the Evaluation of 2D Still Image Face Recognition Algorithms, NIST Interagency Report 7709, Released June 22, 2010. Revised August 23, 2010. http://face.nist.gov/mbe |
| MINEX | P. Grother et al., Performance and Interoperability of the INCITS 378 Template, NIST IR 7296 http://fingerprint.nist.gov/minex04/minex_report.pdf |
| MOC | P. Grother and W. Salamon, MINEX II - An Assessment of ISO/IEC 7816 Card-Based Match-on-Card Capabilities http://fingerprint.nist.gov/minex/minexII/NIST_MOC_ISO_CC_interop_test_plan_1102.pdf |
| PERFSTD INTEROP | ISO/IEC 19795-4 — Biometric Performance Testing and Reporting — Part 4: Interoperability Performance Testing. Posted as document 37N2370. The standard was published in 2007. It can be purchased from ANSI at http://webstore.ansi.org/ . |

699

Annex A

Submission of Implementations to the FIVE

700
701

702 A.1 Submission of implementations to NIST

703 NIST requires that all software, data and configuration files submitted by the participants be signed and encrypted.
704 Signing is done with the participant's private key, and encryption is done with the NIST public key. The detailed
705 commands for signing and encrypting are given here: <http://www.nist.gov/itl/iad/ig/encrypt.cfm>

706 NIST will validate all submitted materials using the participant's public key, and the authenticity of that key will be verified
707 using the key fingerprint. This fingerprint must be submitted to NIST by writing it on the signed participation agreement.

708 By encrypting the submissions, we ensure privacy; by signing the submission, we ensure authenticity (the software
709 actually belongs to the submitter). NIST will reject any submission that is not signed and encrypted. NIST accepts no
710 responsibility for anything that is transmitted to NIST that is not signed and encrypted with the NIST public key.

711 A.2 How to participate

712 Those wishing to participate in FIVE testing must do all of the following, on the schedule listed on Page 2.

713 — IMPORTANT: Follow the instructions for cryptographic protection of your SDK and data here.
714 <http://www.nist.gov/itl/iad/ig/encrypt.cfm>

715 — Send a signed and fully completed copy of the *Application to Participate in the Face In Video Evaluation (FIVE)*. This is
716 available at <http://www.nist.gov/itl/iad/ig/five.cfm>. This must identify, and include signatures from, the Responsible
717 Parties as defined in the application. The properly signed FIVE Application to Participate shall be sent to NIST as a
718 PDF.

719 — Provide an SDK (Software Development Kit) library which complies with the API (Application Programmer Interface)
720 specified in this document.

- 721 • Encrypted data and SDKs below 20MB can be emailed to NIST at five@nist.gov
- 722 • Encrypted data and SDKS above 20MB shall be

723 EITHER

724 ▪ Split into sections AFTER the encryption step. Use the unix "split" commands to make 9MB chunks,
725 and then rename to include the filename extension needed for passage through the NIST firewall.

726 ▪ `you% split -a 3 -d -b 9000000 libFIVE_enron_A_02.tgz.gpg`

727 ▪ `you% ls -l x??? | xargs -iQ mv Q libFIVE_enron_A_02_Q.tgz.gpg`

728 ▪ Email each part in a separate email. Upon receipt NIST will

729 ▪ `nist% cat FIVE2012_enron_A02_*.tgz.gpg > libFIVE_enron_A_02.tgz.gpg`

730 OR

731 ▪ Made available as a file.zip.gpg or file.zip.asc download from a generic http webserver⁸,

732 OR

733 ▪ Mailed as a file.zip.gpg or file.zip.asc on CD / DVD to NIST at this address:

| | |
|--|---|
| FIVE Test Liaison (A203) 100 Bureau Drive A203/Tech225/Stop 8940 NIST Gaithersburg, MD 20899-8940 USA | In cases where a courier needs a phone number, please use NIST shipping and handling on: 301 -- 975 -- 6296. |
|--|---|

⁸ NIST will not register, or establish any kind of membership, on the provided website.

734 **A.3 Implementation validation**

735 Registered Participants will be provided with a small validation dataset and test program available on the website
736 <http://www.nist.gov/itl/iad/ig/five.cfm> shortly after the final evaluation plan is released.

737 The validation test programs shall be compiled by the provider. The output of these programs shall be submitted to NIST.

738 Prior to submission of the SDK and validation data, the Participant must verify that their software executes on the
739 validation images.

740 Software submitted shall implement the FIVE API Specification as detailed in the body of this document.

741 Upon receipt of the SDK and validation output, NIST will attempt to reproduce the same output by executing the SDK on
742 the validation imagery, using a NIST computer. In the event of disagreement in the output, or other difficulties, the
743 Participant will be notified.