

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

# Text Recognition Algorithm Independent Test (TRAIT)

An Evaluation Activity under the DHS Science & Technology  
Child Exploitation Image Analytics Program (CHEXIA)



**Homeland  
Security**

Science and Technology

## Concept, Evaluation Plan and API

Version 0.7, November 17, 2015

Patrick Grother, Mei Ngan, Afzal Godil

Contact via [trait2016@nist.gov](mailto:trait2016@nist.gov)



16

**Provisional Timeline of the TRAIT 2016 Evaluation**

Phase 0 API Development	2015-10-05	Draft evaluation plan
	2015-11-15	Final evaluation plan
Phase 1	2015-12-01	Participation starts: Algorithms may be sent to NIST
	<b>2016-02-08</b>	<b>Last day for submission of algorithms to Phase 1</b>
	2016-03-07	Interim results released to Phase 1 participants
Phase 2	<b>2016-06-12</b>	<b>Last day for submission of algorithms to Phase 2</b>
	2016-07-12	Interim results released to Phase 2 participants
Phase 3	<b>2016-10-19</b>	<b>Last day for submission of algorithms to Phase 3</b>
	2016-Q4	Release of final public report

17

18

19

20

Updates since the last version are highlighted in **cyan**.

21 **Table of Contents**

22 1. TRAIT .....5

23 1.1. Scope.....5

24 1.2. Audience .....5

25 1.3. Market drivers.....6

26 1.4. Test data .....7

27 1.5. Offline testing .....7

28 1.6. Phased testing.....7

29 1.7. Interim reports.....7

30 1.8. Final reports.....7

31 1.9. Application scenarios .....8

32 1.10. Options for participation.....8

33 1.11. Number and schedule of submissions .....8

34 1.12. Core accuracy metrics .....9

35 1.13. Reporting computational efficiency.....9

36 1.14. Hardware specification .....9

37 1.15. Operating system, compilation, and linking environment.....9

38 1.16. Software and Documentation.....10

39 1.17. Runtime behavior.....11

40 1.18. Threaded computations.....11

41 1.19. Time limits.....11

42 2. Data structures supporting the API.....12

43 2.1. Namespace.....12

44 2.2. Overview .....12

45 2.3. Requirement .....12

46 2.4. File formats and data structures.....12

47 3. API Specification.....14

48 3.1. Image-to-location.....14

49 3.2. Image-to-text with provided location information .....16

50 3.3. Image-to-text-and-location.....18

51 Annex A Submission of Implementations to the TRAIT 2016 .....20

52 A.1 Submission of implementations to NIST.....20

53 A.2 How to participate .....20

54 A.3 Implementation validation .....21

56 **List of Figures**

57 Figure 1 – Example of inputs and outputs .....5

58 Figure 2 – Example of input with provided location information and outputs.....6

59

60 **List of Tables**

61 Table 1 – Subtests supported under the TRAIT 2016 activity .....8

62 Table 2 – TRAIT 2016 classes of participation.....8

63 Table 3 – Cumulative total number of algorithms, by class.....8

64 Table 4 – Implementation library filename convention .....10

65 Table 5 – Struct representing a single image.....12

66 Table 7 – Location Types.....12

67 Table 8 – Data structure for location information in an image.....13

68 Table 9 – Enumeration of return codes for API function calls .....13

69 Table 10 – ReturnStatus structure.....13

70 Table 11 – SDK initialization.....15

71 Table 12 – GPU index specification.....15

72 Table 13 – Text detection .....15

73 Table 14 – SDK initialization.....17

74 Table 15 – GPU index specification .....17

75 Table 16 – Text recognition .....17

76 Table 17 – SDK initialization.....19

77 Table 18 – GPU index specification .....19

78 Table 19 – Text processing.....19

79

80

81 **1. TRAIT**

82 **1.1. Scope**

83 This document establishes a concept of operations and C++ API for evaluation of text-in-image detection and recognition  
 84 algorithms submitted to NIST's TRAIT program. See <http://www.nist.gov/itl/iad/ig/trait-2016.cfm> for latest  
 85 documentation.

86 TRAIT proceeds as follows. Algorithm developers send compiled C++ libraries to NIST. NIST executes those algorithms on  
 87 sequestered imagery that has been made available to NIST by, for example, other US Government agencies.

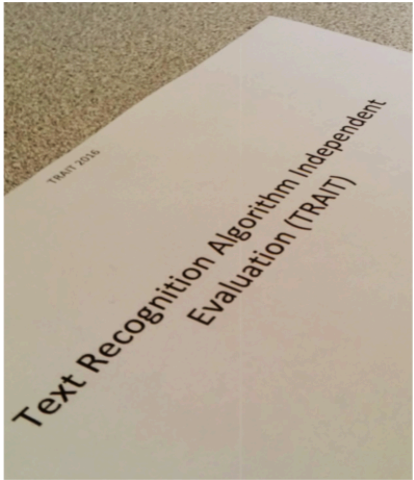
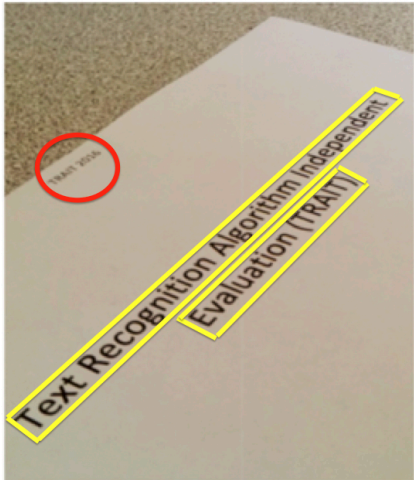
88 **1.2. Audience**

89 This document is aimed at universities, commercial entities and other research organizations possessing a capability to  
 90 detect and recognize unconstrained text in still images. There is no requirement for real-time or streaming-mode  
 91 processing. An example image appears in Figure 1. It is intended only as an example of out-of-plane text, not as some  
 92 representation of widely varying test data.

93

94

**Figure 1 – Example of inputs and outputs**

A simple example of out-of-plane text.	Input image showing geometric markup in yellow, and missed detection in red.	Possible Output text and (dummy, nominal) coordinates
		<p>First string at very top of page is missed</p> <p>Text Recognition Algorithm Independent  <math>(x,y) = (10,10),(10,100),(20,100),(20,10),(10,10)</math></p> <p>Evaluation (TRAIT)  <math>(x,y) = (30,30),(30,60),(40,60),(30,40),(30,30)</math></p>

95

96

97

98

99

100

101

102


103

104

105

106

**Figure 2 – Example of input with provided location information and outputs**

<p>A simple example of out-of-plane text with provided location information marked by the yellow lines.</p> <p>NOTE: When location information is provided to a text recognition algorithm, we provide a simple line instead of a polygon. The reason for this is that our ground truthing process put emphasis on maximizing the number of images i.e. high speed, high volume. This meant that drawing bounding box polygons, which is slow, was deprecated in favor of drawing lines. Implementations that require polygons for recognition will have to estimate that information starting from the provided line.</p>	<p>Possible Output text corresponding to provided locations</p>
	<p>IN CASE OF FIRE DO NOT USE ELEVATOR USE EXIT STAIRS</p>

107

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

111 NIST is particularly interested to evaluate prototypes that have proven useful in prior evaluations organized underneath  
 112 the ICDAR conferences (<http://2015.icdar.org/program/competitions/>) particularly the Robust Reading efforts  
 113 (<http://rrc.cvc.uab.es/>)

114 **1.3. Market drivers**

115 This test is intended to support a plural marketplace of text recognition systems. Our primary driver is to support forensic  
 116 investigations of digital media. Specifically, to allow linking of child exploitation events that occur in a common location,  
 117 or that share other textual clues.

#### 118 **1.4. Test data**

119 NIST will run submitted algorithms on several sequestered datasets available to NIST.

120

121 The primary dataset is an operational child exploitation collection containing illicit pornographic images. The images are  
122 present on digital media seized in criminal investigations. The files include children who range in age from infant through  
123 adolescent. Their faces are the subject of a separate face recognition evaluation and development effort (CHEXIA-FACE  
124 2016). Many of the images contain geometrically unconstrained text. This text is human-legible and sometimes has  
125 investigational value. Such text is visible on certificates, posters, logos, uniforms, sports apparel, computer screens,  
126 business cards, newspapers, books lying on tables, cigarette packets and a long list of more rare objects. **There will also  
127 be instances where watermarks or logos were post-processed into the image.**

128 The text is most commonly in English with French, Spanish, German and Cyrillic present in significant quantity. We do not  
129 intend to test non-Roman alphabets.

130 These images are of interest to NIST's partner law enforcement agencies that seek to employ text recognition in  
131 investigating this area of serious crime. The primary applications are identification of previously known victims and  
132 suspects, as well as detection of new victims and suspects. The presence of text may allow a location to be identified or to  
133 generate leads.

#### 134 **1.5. Offline testing**

135 TRAIT is intended to mimic operational reality. As an offline test intended to assess the core algorithmic capability of text  
136 detection and recognition algorithms, it does not extend to real-time transcription of live image sources. Offline testing is  
137 attractive because it allows uniform, fair, repeatable, and efficient evaluation of the underlying technologies. Testing of  
138 implementations under a fixed API allows for a detailed set of performance related parameters to be measured. The  
139 algorithms will be run only on NIST machines by NIST employees.

#### 140 **1.6. Phased testing**

141 To support development, TRAIT will be conducted in three phases. In each phase, NIST will evaluate implementations on a  
142 first-come-first-served basis and will return results to providers as expeditiously as possible. The final phase will result in  
143 public reports. Providers may submit revised SDKs to NIST only after NIST provides results for the prior SDK and invites  
144 further submission. The frequency with which a provider may submit SDKs to NIST will depend on the times needed for  
145 developer preparation, transmission to NIST, validation, execution and scoring at NIST, and developer review and decision  
146 processes.

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

#### 148 **1.7. Interim reports**

149 The performance of each SDK will be reported in a "score-card". This will be provided to the participant and not publicly.  
150 The feedback is intended to facilitate development. Score cards will: be machine generated (i.e. scripted); be provided to  
151 participants with identification of their implementation; include timing, accuracy and other performance results; include  
152 results from other implementations, but will not identify the other providers; be expanded and modified as revised  
153 implementations are tested, and as analyses are implemented; be produced independently of the status of other  
154 providers' implementations; be regenerated on-the-fly, usually whenever any implementation completes testing, or when  
155 new analysis is added.

156 NIST does not intend to release these test reports publicly. NIST may release such information to the U.S. Government  
157 test sponsors; NIST will request that agencies not release this content.

#### 158 **1.8. Final reports**

159 NIST will publish one or more final public reports. NIST may also publish: additional supplementary reports (typically as  
160 numbered NIST Interagency Reports); in other academic journals; in conferences and workshops (typically PowerPoint).

161 Our intention is that the final test reports will publish results for the best-performing implementation from each  
162 participant. Because "best" is ill defined (accuracy vs. processing time, for example), the published reports may include

163 results for other implementations. The intention is to report results for the most capable implementations (see section  
 164 1.12, on metrics). Other results may be included (e.g. in appendices) to show, for example, examples of progress or  
 165 tradeoffs. **IMPORTANT: Results will be attributed to the providers.**

166 **1.9. Application scenarios**

167 The test will include detection and recognition tasks for still images. As described in Table 1, the test is intended to  
 168 support operations in which an automated text recognition engine yields text that can be indexed and retrieved using  
 169 mainline text retrieval engines.

170 **Table 1 – Subtests supported under the TRAIT 2016 activity**

#		A	B	C
1.	Aspect	Image-to-location	Image-to-text with provided location information	Image-to-text-and-location
2.	Languages	Mostly English. Some French, Spanish and German. While some Cyrillic and Chinese appear also, evaluation will be confined to English roman alphabets only.		
3.	Input	Image(s)	Image(s) and location(s) of text	Image(s)
4.	Output	Given an input image, output detected locations of text. This does not require the algorithm(s) to produce strings of text.	Given an input image and location(s) of text in the image, output strings of text.	Given an input image, output strings of text along with their corresponding locations in the image.

171

172 NOTE 1: The vast majority of images are color. The API supports both color and greyscale images.

173 NOTE 2: For the operational datasets, it is not known what processing was applied to the images before they were  
 174 archived. So, for example, we do not know whether gamma correction was applied. NIST considers that best practice,  
 175 standards and operational activity in the area of image preparation remains weak.

176 **1.10. Options for participation**

177 The following rules apply:

- 178 – A participant must properly follow, complete and submit the TRAIT 2016 Participation Application (see Annex A). This  
 179 must be done once, not before December 1, 2015. It is not necessary to do this for each submitted SDK.
- 180 – Participants may submit class C algorithms only if at least 1 class B algorithm is also submitted.
- 181 – All submissions shall implement exactly one of the functionalities defined in Table 2. A library shall not implement  
 182 two or more classes.

183

**Table 2 – TRAIT 2016 classes of participation**

Function	Image-to-location	Image-to-text with provided location information	Image-to-text-and-location
Class label	A	B	C
Must also submit to class			B
API requirements	3.1	3.2	3.3

184 **1.11. Number and schedule of submissions**

185 The test is conducted in three phases, separated by a few months. The maximum total (i.e. cumulative) number of  
 186 submissions is regulated in Table 3.

187

**Table 3 – Cumulative total number of algorithms, by class**

#	Phase 1	Total over Phases 1 + 2	Total over Phases 1 + 2 + 3
Class A: Image-to-location	2	4	6
Class B: Image-to-text with provided location information	2	4	6
Class C: Image-to-text-and-location	2	4	6



188 The numbers above may be increased as resources allow.

189 NIST cannot conduct surveys over runtime parameters.

### 190 **1.12. Core accuracy metrics**

191 **Recognition:** The evaluation will be performed on the text results provided by each system. We intend to state text  
192 recognition accuracy with at least an edit distance such as the Word Error Rate (WER) [1.12a] between the reference text  
193 and text provided by the system for each line. WER is calculated with the edit distance with equal cost of deletions,  
194 substitutions, and insertions and finally normalize the edit distance by the number of characters in the ground truth  
195 words.

196 [1.12a] J. Fiscus, J. Ajot, N. Radde, and C. Laprun, *Multiple Dimension Levenshtein Edit Distance Calculations for Evaluating*  
197 *Automatic Speech Recognition Systems During Simultaneous Speech*, Proceedings of LREC, 2006.

198 [http://www.itl.nist.gov/iad/mig/publications/storage\\_paper/lrec06\\_v0\\_7.pdf](http://www.itl.nist.gov/iad/mig/publications/storage_paper/lrec06_v0_7.pdf)

199 **Detection:** The text detection task will be evaluated, somewhat similar to prior open evaluations [1.12b]. However, in our  
200 case the ground truth text, is defined by line and curve segments instead of bounding boxes. Hence our methodology will  
201 use a simple matching distance approach between lines and curves as the criteria.

202 [1.12b] C. Wolf and J.-M. Jolion. Object count/Area Graphs for the Evaluation of Object Detection and Segmentation  
203 Algorithms, International Journal on Document Analysis and Recognition, 8(4):280-296, 2006.

204 <http://iris.cnrs.fr/christian.wolf/software/deteval/index.html>

### 205 **1.13. Reporting computational efficiency**

206 NIST will also report timing statistics for all core functions of the submitted SDK implementations. **All timing tests will be**  
207 **executed on unloaded machines running a single process.**

### 208 **1.14. Hardware specification**

209 NIST intends to execute the software on Dual Intel Xeon E5-2695 3.3 GHz CPUs (14 cores each; 56 logical CPUs total) with  
210 Dual NVIDIA Tesla K40 **Graphics Processing Units** (GPUs). NIST will respond to prospective participants' questions on the  
211 hardware by amending this section.

### 212 **1.15. Operating system, compilation, and linking environment**

213 The operating system that the submitted implementations shall run on will be released as a downloadable file accessible  
214 from <http://nigos.nist.gov:8080/evaluations/> which is the 64-bit version of CentOS 7 running Linux kernel 3.10.0.

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

217 NIST will link the provided library file(s) to our C++ language test drivers. Participants are required to provide their library  
218 in a format that is linkable using the C++11 compiler, g++ version 4.8.3.

219 A typical link line might be

220 `g++ -std=c++11 -l. -Wall -m64 -o trait16test trait16test.cpp -L. -ltrait2016_Enron_A_07`

221 The Standard C++ library should be used for development. The prototypes from this document will be written to a file  
222 "trait2016.h" which will be included via

```
#include <trait2016.h>
```

223 The header **and source** files will be made available to implementers at <http://nigos.nist.gov:8080/trait2016>.

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

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

229 Dependencies on external dynamic/shared libraries such as compiler-specific development environment libraries are  
 230 discouraged. If absolutely necessary, external libraries must be provided to NIST upon prior approval by the Test Liaison.  
 231 **Libraries to access the GPU must be provided to NIST as a part of the algorithm submission package.**

232 **1.16. Software and Documentation**

233 **1.16.1. SDK Library and Platform Requirements**

234 Participants shall provide NIST with binary code only (i.e., no source code). Header files (“.h”) are allowed, but these shall  
 235 not contain intellectual property of the company nor any material that is otherwise proprietary. The SDK should be  
 236 submitted in the form of a dynamically linked library file. **A separate library file shall be submitted for each class of**  
 237 **participation (i.e., CLASS\_A, CLASS\_B, CLASS\_C).**

238 The core library shall be named according to Table 4. Additional shared object library files may be submitted that support  
 239 this “core” library file (i.e. the “core” library file may have dependencies implemented in these other libraries).

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

244 **Table 4 – Implementation library filename convention**

Form	libTRAIT2016_provider_class_sequence.ending				
Underscore delimited parts of the filename	libTRAIT2016	provider	class	sequence	ending
Description	First part of the name, required to be this.	Single word name of the main provider EXAMPLE: Enron	Function classes supported in Table 2. EXAMPLE: C	A two digit decimal identifier to start at 00 and increment by 1 every time a library is sent to NIST. EXAMPLE: 07	.so
Example	libTRAIT2016_Enron_C_07.so				

245  
 246 NIST will report the size of the supplied libraries.

247 **1.16.2. Configuration and developer-defined data**

248 The implementation under test may be supplied with configuration files and supporting data files. The total size of the  
 249 SDK, that is all libraries, include files, data files and initialization files shall be less than or equal to 1 073 741 824 bytes =  
 250 1024<sup>3</sup> bytes.

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

252 **1.16.3. Submission folder hierarchy**

253 Participant submissions should contain the following folders at the top level

- 254 • lib/ - contains all participant-supplied software libraries
- 255 • config/ - contains all configuration and developer-defined data
- 256 • doc/ - contains any participant-provided documentation regarding the submission

257 **1.16.4. Installation and Usage**

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

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

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

264 **1.16.5. Documentation**

265 Participants may provide documentation of the SDK and detail any additional functionality or behavior beyond that  
 266 specified here. The documentation might include developer-defined error or warning return codes. The documentation  
 267 shall not include any intellectual property.

268 **1.17. Runtime behavior**

269 **1.17.1. Interactive behavior**

270 The implementation will be tested in non-interactive “batch” mode (i.e., without terminal support). Thus, the submitted  
 271 library shall:

- 272 – Not use any interactive functions such as graphical user interface (GUI) calls or any other calls which require  
 273 terminal interaction, e.g., reads from “standard input”.
- 274 – Run quietly, i.e., it should not write messages to "standard error" and shall not write to “standard output”.
- 275 – If requested by NIST for debugging, include a logging facility in which debugging messages are written to a log file  
 276 whose name includes the provider and library identifiers and the process PID.

277 **1.17.2. Exception Handling**

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

280 **1.17.3. External communication**

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

286 **1.17.4. Stateless behavior**

287 All components in this test shall be stateless, except as noted. This applies to text detection, recognition and  
 288 transcription. Thus, all functions should give identical output, for a given input, independent of the runtime history. NIST  
 289 will institute appropriate tests to detect stateful behavior. If detected, NIST will take appropriate steps, including but not  
 290 limited to, cessation of evaluation of all implementations from the supplier, notification to the provider, and  
 291 documentation of the activity in published reports.

292 **1.18. Threaded computations**

293 All implementations should run without threads, or with exactly one single process. This allows NIST to parallelize the test  
 294 by dividing the workload across multiple cores and multiple machines. **The implementation shall be tolerant of NIST**  
 295 **running N > 2 processes concurrently.** NIST's calling applications are single-threaded.

296 **1.19. Time limits**

297 Given a 12 megapixel input image, the text detection and recognition software should execute in less than 10 seconds.

298

299

300 **2. Data structures supporting the API**

301 **2.1. Namespace**

302 [Redacted]

303 **2.2. Overview**

304 This section describes separate APIs for the core text detection/recognition applications described in section 1.9. All  
 305 SDK's submitted to TRAIT 2016 shall implement the functions required by the rules for participation listed before Table 2.

306 **2.3. Requirement**

307 TRAIT 2016 participants shall submit an SDK, which implements the relevant C++ functions (per class) as specified in Table  
 308 2. C++ was chosen in order to make use of some object-oriented features.

309 **2.4. File formats and data structures**

310 **2.4.1. Overview**

311 In this text detection and recognition test, the input data is a still image.

312 **2.4.2. Data structures for encapsulating a single image**

313 An image is provided to the algorithm using the data structure of Table 5.

314 **Table 5 – Struct representing a single image**

	C++ code fragment	Remarks
1.	struct Image	
2.	{	
3.	uint16_t image_width;	Number of pixels horizontally
4.	uint16_t image_height;	Number of pixels vertically
5.	uint8_t image_depth;	Number of bits per pixel. Legal values are 8 and 24.
6.	uint8_t *data;	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
7.	};	

315

316 **2.4.3. Data structures for reporting detected text**

317 Implementations should report locations of text in each image using the structure of the table below.

318

319 **Table 6 – Structure representing a point in 2D coordinates**

	C++ code fragment	Remarks
1.	struct Coordinate	
2.	{	
3.	uint16_t x;	x-value
4.	uint16_t y;	y-value
5.	};	

320

321 **Table 7 – Location Types**

	Location Type as C++ enumeration	Meaning
	enum class LocationType {	

1.	<code>/* Input only */ Line=1,</code>	Coordinates of line segments drawn through the centroids of the text. There will be N=2 points with coordinates giving the endpoints. This type is used for input to the implementation only, i.e. Class B (Image-to-text with provided location).
2.	<code>/* Output only */ Polygon=2</code>	In reading, clockwise order, the coordinates of a closed polygon drawn around the line of text, where the coordinates of the first and last points in the polygon are the same. This type is used for output from the implementation only, i.e. Class A (Image-to-location) and Class C (Image-to-text-and-location).
3.	<code>};</code>	

322

323

**Table 8 – Data structure for location information in an image**

	C++ code fragment	Remarks
1.	<code>struct Location</code>	
2.	<code>{</code>	
3.	<code>std::vector&lt;Coordinate&gt; points;</code>	Coordinates representing the location of the text
4.	<code>LocationType type;</code>	For Class B (Image-to-text with provided location), this will provide the location information for text in the form of a line through the centroids of the text.  For Class A (Image-to-location) and C (Image-to-text-and-location), the locations returned by the implementation <b>MUST</b> be a Polygon.
5.	<code>};</code>	

324

**2.4.4. Data structure for return value of API function calls**

325

**Table 9 – Enumeration of return codes for API function calls**

	Return code as C++ enumeration	Meaning
	<code>enum class ReturnCode {</code>	
1.	<code>Success=0,</code>	Success
2.	<code>ConfigError=1,</code>	Error reading configuration files
3.	<code>RefuseInput=2,</code>	Elective refusal to process the input
4.	<code>VendorError=3</code>	Vendor-defined error
5.	<code>};</code>	

326

**Table 10 – ReturnStatus structure**

	C++ code fragment	Meaning
	<code>struct ReturnStatus {</code>	
1.	<code>TRAIT2016::ReturnCode code;</code>	Return Code
2.	<code>std::string info;</code>	Optional information string
3.	<code>// constructors</code>	
4.	<code>};</code>	

327 **3. API Specification**

328 **3.1. Image-to-location**

329 **3.1.1. Overview**

330 This section defines an API for algorithms that can perform solely text detection. This does not reflect an operational use-  
 331 case per se, but is included in this evaluation to identify capable algorithms and to support, in-principle, good detection  
 332 algorithms that have poor recognition capability.

333 **3.1.2. API**

334 **3.1.2.1. Interface**

335 The Class A software under test must implement the interface `ImgToLocationInterface` by subclassing this class  
 336 and implementing each method specified therein.

	C++ code fragment	Remarks
1.	<code>class ImgToLocationInterface</code>	
2.	<code>{</code> <code>public:</code>	
3.	<code>virtual ReturnStatus initialize_text_detector( const std::string &amp;configuration_location);</code>	
4.	<code>virtual ReturnStatus detect_text_in_still( const Image &amp;image, std::vector&lt;Location&gt; &amp;textLocations) = 0;</code>	
5.	<code>virtual void set_gpu(uint8_t gpuNum) = 0;</code>	
6.	<code>static ClassAImplPtr getImplementation();</code>	Factory method to return a managed pointer to the <code>ImgToLocationInterface</code> object. This function is implemented by the submitted library and must return a managed pointer to the <code>ImgToLocationInterface</code> object.
7.	<code>};</code>	

337 There is one class (static) method declared in `ImgToLocationInterface`, `getImplementation()` which must  
 338 also be implemented by the SDK. This method returns a shared pointer to the object of the interface type, an instantiation  
 339 of the implementation class. A typical implementation of this method is also shown below as an example.  
 340  
 341

	C++ code fragment	Remarks
	<code>#include "trait2016NullImplClassA.h"</code>	
	<code>using namespace TRAIT2016;</code>	
	<code>NullImplClassA::NullImplClassA() { }</code>	
	<code>NullImplClassA::~NullImplClassA() { }</code>	
	<code>ClassAImplPtr ImgToLocationInterface::getImplementation() { NullImplClassA *p = new NullImplClassA(); ClassAImplPtr ip(p); return (ip); }</code>	
	<code>// Other implemented functions</code>	

342 **3.1.2.2. Initialization**

343 Before any text detection calls are made, the NIST test harness will make a call to the initialization of the function in Table  
 344 11.

345 **Table 11 – SDK initialization**

Prototype	<b>ReturnStatus</b> initialize_text_detector( const std::string &configuration_location);	Input
Description	This function initializes the SDK under test. It will be called by the NIST application before any call to detect_text_in_still() is made.	
Input Parameters	configuration_location	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 in this directory are hardwired in the SDK and are unrestricted.
<b>Return Code</b>	Success	Success
	ConfigError	Vendor provided configuration files are not readable in the indicated location.
	VendorError	Vendor-defined failure

346 **3.1.2.3. GPU Index Specification**

347 For implementations using GPUs, the function of Table 12 specifies a sequential index for which GPU device to execute  
 348 on. This enables the test software to orchestrate load balancing across multiple GPUs.

349 **Table 12 – GPU index specification**

<b>Prototypes</b>	void set_gpu ( uint8_t gpuNum);	Input
<b>Description</b>	This function sets the GPU device number to be used by all subsequent implementation function calls. gpuNum is a zero-based sequence value of which GPU device to use. 0 would mean the first detected GPU, 1 would be the second GPU, etc. If the implementation does not use GPUs, then this function call should simply do nothing.	
<b>Input Parameters</b>	gpuNum	Index number representing which GPU to use.

350 **3.1.2.4. Text detection**

351 The text detection functions of Table 13 accept input imagery and report the location(s) of zero or more lines of text, in  
 352 the form of a closed polygon. NIST may evaluate on images with no text in them.

353 **Table 13 – Text detection**

Prototypes	ReturnStatus detect_text_in_still ( const Image &image, std::vector<Location> &textLocations);	Input Output
Description	This function takes, respectively, a still image and returns the locations of lines of text, if any, in the form of closed polygon(s). The output textLocations vector will initially be empty. It is up to the implementation to populate the vector and ensure textStrings.size() == images.size.	
Input Parameters	Image	A Table 5 structure
Output Parameters	textLocations	A vector of Table 8 structure
<b>Return Code</b>	Success	Success
	RefuseInput	Elective refusal to process the input – e.g., because quality is too poor
	VendorError	Vendor-defined failure. Failure codes must be documented and communicated to NIST with the submission of the implementation under test.

354 **3.2. Image-to-text with provided location information**

355 **3.2.1. Overview**

356 This section defines an API for algorithms that perform recognition given text location in an image. This is not a primary  
 357 operational use-case, but is included for NIST to evaluate the relative difficulties of detection vs. recognition.

358 **3.2.2. API**

359 **3.2.2.1. Interface**

360 The Class B software under test must implement the interface `ImgToTextInterface` by subclassing this class and  
 361 implementing each method specified therein. See

C++ code fragment	Remarks
1. <code>class ImgToTextInterface</code>	
2. <code>{</code>	
3. <code>public:</code>	
3. <code>virtual ReturnStatus initialize_text_recognizer( const std::string &amp;configuration_location);</code>	
4. <code>virtual ReturnStatus recognize_text_in_still( const Image &amp;image, const std::vector&lt;Location&gt; &amp;textLocations, std::vector&lt;std::string&gt; &amp;textStrings) = 0;</code>	
5. <code>virtual void set_gpu(uint8_t gpuNum) = 0;</code>	
6. <code>static ClassBImplPtr getImplementation();</code>	Factory method to return a managed pointer to the <code>ImgToTextInterface</code> object. This function is implemented by the submitted library and must return a managed pointer to the <code>ImgToTextInterface</code> object.
7. <code>};</code>	

362 There is one class (static) method declared in `ImgToTextInterface`, `getImplementation()` which must also be  
 363 implemented by the SDK. This method returns a shared pointer to the object of the interface type, an instantiation of the  
 364 implementation class. A typical implementation of this method is also shown below as an example.  
 365  
 366

C++ code fragment	Remarks
<code>#include "trait2016NullImplClassB.h"</code>	
<code>using namespace TRAIT2016;</code>	
<code>NullImplClassB::NullImplClassB() { }</code>	
<code>NullImplClassB::~NullImplClassB() { }</code>	
<code>ClassBImplPtr</code>	
<code>ImgToTextInterface::getImplementation()</code>	
<code>{</code>	
<code>NullImplClassB *p = new NullImplClassB();</code>	
<code>ClassBImplPtr ip(p);</code>	
<code>return (ip);</code>	
<code>}</code>	
<code>// Other implemented functions</code>	

367



368 **3.2.2.2. Initialization**

369 Before any text recognition calls are made, the NIST test harness will make a call to the initialization of the function in  
 370 Table 14.

371 **Table 14 – SDK initialization**

Prototype	ReturnStatus initialize_text_recognizer( const std::string &configuration_location);	Input
Description	This function initializes the SDK under test. It will be called by the NIST application before any call to recognize_text_in_still().	
Input Parameters	configuration_location	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 in this directory are hardwired in the SDK and are unrestricted.
Return Code	Success	Success
	ConfigError	Vendor provided configuration files are not readable in the indicated location.
	Other	Vendor-defined failure

372 **3.2.2.3. GPU Index Specification**

373 For implementations using GPUs, the function of Table 15 specifies a sequential index for which GPU device to execute  
 374 on. This enables the test software to orchestrate load balancing across multiple GPUs.

375 **Table 15 – GPU index specification**

Prototypes	void set_gpu ( uint8_t gpuNum);	Input
Description	This function sets the GPU device number to be used by all subsequent implementation function calls. gpuNum is a zero-based sequence value of which GPU device to use. 0 would mean the first detected GPU, 1 would be the second GPU, etc. If the implementation does not use GPUs, then this function call should simply do nothing.	
Input Parameters	gpuNum	Index number representing which GPU to use.

376 **3.2.2.4. Text recognition with provided location information**

377 The text recognition functions of Table 16 accept input imagery and locations of text in the image and report zero or more  
 378 lines of recognized text. NIST may provide locations where no text exists, and the implementation should handle that.

379 **Table 16 – Text recognition**

Prototypes	ReturnStatus recognize_text_in_still( const Image &image, const std::vector<Location> &textLocations, std::vector<std::string> &textStrings);	Input Input Output
Description	This function takes a still images and $K \geq 1$ locations of text and returns $K$ (possibly empty) strings of text for the image. The output vector will initially be empty and it is up to the implementation to ensure that <code>textStrings.size() == images.size()</code> . <code>textStrings[i]</code> should correspond to the location information from <code>textLocations[i]</code> . If no text is recognized for a particular provided location, the implementation shall populate <code>textStrings[i]</code> with an empty string (i.e., ""). The provided location information will be in the form of a line, that is, <code>textLocations[i].type=Line</code> .	
Input Parameters	image	A Table 5 structure
	textLocations	A vector of Table 8 structure
Output Parameters	textStrings	A vector of std::string
Return Code	Success	Success
	RefuseInput	Elective refusal to process the input – e.g. because quality is too poor
	VendorError	Vendor-defined failure. Failure codes must be documented and communicated to NIST with the submission of the implementation under test.

380 **3.3. Image-to-text-and-location**

381 **3.3.1. Overview**

382 This section defines an API for algorithms that can perform text recognition in stills. This reflects the primary operational  
 383 use-case.

384 **3.3.2. API**

385 **3.3.2.1. Interface**

386 The Class C software under test must implement the interface `ImgToTextAndLocInterface` by subclassing this class  
 387 and implementing each method specified therein. See

C++ code fragment	Remarks
1. <code>class ImgToTextAndLocInterface</code>	
2. <code>{</code> <code>public:</code>	
3. <code>virtual ReturnStatus initialize_text_processor(</code> <code>const std::string &amp;configuration_location,</code> <code>bool &amp;useGPU);</code>	
4. <code>virtual ReturnStatus process_text_in_still(</code> <code>const Image &amp;image,</code> <code>std::vector&lt;Location&gt; &amp;textLocations,</code> <code>std::vector&lt;std::string&gt; &amp;textStrings) = 0;</code>	
5. <code>virtual void set_gpu(uint8_t gpuNum) = 0;</code>	
6. <code>static ClassCImplPtr getImplementation();</code>	Factory method to return a managed pointer to the <code>ImgToTextAndLocInterface</code> object. This function is implemented by the submitted library and must return a managed pointer to the <code>ImgToTextAndLocInterface</code> object.
7. <code>};</code>	

388 There is one class (static) method declared in `ImgToTextAndLocInterface`, `getImplementation()` which  
 389  
 390  
 391 instantiation of the implementation class. A typical implementation of this method is also shown below as an example.  
 392

C++ code fragment	Remarks
<pre> #include "trait2016NullImplClassC.h"  using namespace TRAIT2016;  NullImplClassC::NullImplClassC() { }  NullImplClassC::~NullImplClassC() { }  ClassCImplPtr ImgToTextAndLocInterface::getImplementation() {     NullImplClassC *p = new NullImplClassC();     ClassCImplPtr ip(p);     return (ip); }  // Other implemented functions                     </pre>	

393

394 **3.3.2.2. Initialization**

395 Before any text recognition/processing calls are made, the NIST test harness will make a call to the initialization of the  
396 function in Table 14.

397 **Table 17 – SDK initialization**

Prototype	ReturnStatus initialize_text_processor( const std::string &configuration_location);	Input
Description	This function initializes the SDK under test. It will be called by the NIST application before any call to process_text_in_still() is made.	
Input Parameters	configuration_location	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 in this directory are hardwired in the SDK and are unrestricted.
Return Code	Success	Success
	ConfigError	Vendor provided configuration files are not readable in the indicated location.
	VendorError	Vendor-defined failure

398 **3.3.2.3. GPU Index Specification**

399 For implementations using GPUs, the function of Table 18 specifies a sequential index for which GPU device to execute  
400 on. This enables the test software to orchestrate load balancing across multiple GPUs.

401 **Table 18 – GPU index specification**

Prototypes	void set_gpu ( uint8_t gpuNum);	Input
Description	This function sets the GPU device number to be used by all subsequent implementation function calls. gpuNum is a zero-based sequence value of which GPU device to use. 0 would mean the first detected GPU, 1 would be the second GPU, etc. If the implementation does not use GPUs, then this function call should simply do nothing.	
Input Parameters	gpuNum	Index number representing which GPU to use.

402 **3.3.2.4. Text processing without location information**

403 The text processing functions of Table 19 accept input imagery and report zero or more lines of text.

404 **Table 19 – Text processing**

Prototypes	ReturnStatus process_text_in_still( const Image &images, std::vector<Location> &textLocations, std::vector<std::string> &textStrings);	Input Output Output
Description	This function takes a still image and returns strings of text and their corresponding locations found in the image. textStrings[i] shall correspond to the location(s) for text in textLocations[i]. The output vectors will initially be empty and it is up to the implementation to ensure that textLocations.size() == textStrings.size(). The implementation shall report the location of lines of text in the form of a closed polygon, that is, textLocations[i].type=Polygon.	
Input Parameters	image	A Table 5 structure
Output Parameters	textLocations	A vector of Table 8 structure
	textStrings	A vector of std::string
Return Code	Success	Success
	RefuseInput	Elective refusal to process the input – e.g. because quality is too poor
	VendorError	Vendor-defined failure. Failure codes must be documented and communicated to NIST with the submission of the implementation under test.

405

406  
407

# Annex A Submission of Implementations to the TRAIT 2016

## 408 A.1 Submission of implementations to NIST

409 NIST requires that all software, data and configuration files submitted by the participants be signed and encrypted.  
 410 Signing is done with the participant's private key, and encryption is done with the NIST public key. The detailed  
 411 commands for signing and encrypting are given here: <http://www.nist.gov/itl/iad/ig/encrypt.cfm>  
 412 NIST will validate all submitted materials using the participant's public key, and the authenticity of that key will be verified  
 413 using the key fingerprint. This fingerprint must be submitted to NIST by writing it on the signed participation agreement.  
 414 By encrypting the submissions, we ensure privacy; by signing the submissions, we ensure authenticity (the software  
 415 actually belongs to the submitter). NIST will reject any submission that is not signed and encrypted. NIST accepts no  
 416 responsibility for anything that is transmitted to NIST that is not signed and encrypted with the NIST public key.

## 417 A.2 How to participate

418 Those wishing to participate in TRAIT 2016 testing must do all of the following, on the schedule listed in this document.

- 419 — IMPORTANT: Follow the instructions for cryptographic protection of your SDK and data here.  
 420 <http://www.nist.gov/itl/iad/ig/encrypt.cfm>
- 421 — Send a signed and fully completed copy of the *Application to Participate in the Text Recognition Algorithm*  
 422 *Independent Test (TRAIT) 2016*. This is available at <http://www.nist.gov/itl/iad/ig/trait-2016.cfm>. This must identify,  
 423 and include signatures from, the Responsible Parties as defined in the application. The properly signed TRAIT 2016  
 424 Application to Participate shall be sent to NIST as a PDF.
- 425 — Provide an SDK (Software Development Kit) library which complies with the API (Application Programmer Interface)  
 426 specified in this document.
  - 427 • Encrypted data and SDKs below 20MB can be emailed to NIST at [trait2016@nist.gov](mailto:trait2016@nist.gov).
  - 428 • Encrypted data and SDKS above 20MB shall be
    - 429 EITHER
      - 430 ▪ Split into sections AFTER the encryption step. Use the unix "split" commands to make 9MB chunks,  
 431 and then rename to include the filename extension need for passage through the NIST firewall.
      - 432 ▪ `you% split -a 3 -d -b 9000000 libTRAIT2016_enron_A_02.tgz.gpg`
      - 433 ▪ `you% ls -l x??? | xargs -iQ mv Q libTRAIT2016_enron_A_02_Q.tgz.gpg`
      - 434 ▪ Email each part in a separate email. Upon receipt NIST will
      - 435 ▪ `nist% cat TRAIT2016_enron_A02_*.tgz.gpg > libTRAIT2016_enron_A_02.tgz.gpg`
    - 436 OR
      - 437 ▪ Made available as a file.zip.gpg or file.zip.asc download from a generic http webserver<sup>1</sup>,
    - 438 OR
      - 439 ▪ Mailed as a file.zip.gpg or file.zip.asc on CD / DVD to NIST at this address:

TRAIT 2016 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.
--	---

<sup>1</sup> NIST will not register, or establish any kind of membership, on the provided website.

**440 A.3 Implementation validation**

441 Registered Participants will be provided with a small validation dataset and test program available on the website.

442 <http://www.nist.gov/itl/iad/ig/trait-2016.cfm> shortly after the final evaluation plan is released.

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

444 Prior to submission of the SDK and validation data, the Participant must verify that their software executes on the  
445 validation images, and produces correct similarity scores and templates.

446 Software submitted shall implement the TRAIT 2016 API Specification as detailed in the body of this document.

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