

Appendix for Collecting and Annotating the Large Continuous Action Dataset

Daniel Paul Barrett · Ran Xu · Haonan Yu · Jeffrey Mark Siskind

Received: date / Accepted: date

A Origin

The evaluations conducted by Mind's Eye included a specific set of tasks, namely Recognition (REC), Description (DES), Gap Filling (GAP), and Anomaly Detection (ANM) with specific evaluation metrics. Such tasks and metrics are expressly *not* part of LCA. The evaluations conducted under Mind's Eye make use of material that is not included in LCA and metrics that are not public. Likewise, LCA contains material that was not available for use during the Mind's Eye evaluations. Thus said evaluations could not be replicated outside of the context of Mind's Eye. Any potential future evaluations conducted with LCA would thus be incomparable to the results obtained under Mind's Eye.

The entire LCA dataset, including the video and the annotations, has been cleared for release by DARPA. The remaining material gathered by DARPA for the Mind's Eye Year 2 evaluation that is not included in LCA may not have been cleared for release. As part of the release process, some video was edited to remove certain portions. Furthermore, the annotation process was performed with the particular versions of the videos included in LCA as provided by DARPA. These may have been transcoded from the original as filmed by the camera. Thus, the time alignment of the annotations can only be guaranteed with the versions of the videos included in LCA. The time alignment may not be correct for any other versions of these videos that may be residual from the DARPA Mind's Eye program.

Daniel Paul Barrett
School of Electrical and Computer Engineering, Purdue University

Ran Xu
Computer Science and Engineering, SUNY Buffalo

Haonan Yu
School of Electrical and Computer Engineering, Purdue University

Jeffrey Mark Siskind
School of Electrical and Computer Engineering, Purdue University
Tel.: +1-765-496-3197
Fax: +1-765-494-2706
E-mail: qobi@purdue.edu

B Collection

The video for this dataset was filmed by DARPA in conjunction with Mitre and several performers from the Mind's Eye program.¹ The bulk of the video was filmed as part of the Mind's Eye Year 2 evaluation. Within the Mind's Eye program, that video was referred to as C-D2a, C-D2b, C-D2c, and the Y2 Evaluation dataset. This video is disjoint from that gathered by Janus Research Group as part of the Mind's Eye Year 1 evaluation. Within the Mind's Eye program, that video was referred to as C-D1a, C-D1b, C-D1, and C-E1. As the LCA dataset contains only a subset of that material, we refrain from using all such terminology in reference to LCA. Also note that the LCA dataset contains some video that was not included in the data used as part of the Mind's Eye evaluations.

The LCA dataset was filmed at three different locations over four periods:

1. The Great Plains Joint Training Center (GPJTC), an army training facility in Kansas, on 22 August 2011. Filming took place in two contexts at GPJTC, a simulated country road and a simulated safe house.
2. Strategic Operations, Inc. (STOPS), a training facility in San Diego, on 14–15 December 2011 and on 6–9 March 2012. Filming during the first period took place in three contexts: two different simulated country roads and one simulated safe house. Filming during the second period took place in five contexts: two different simulated country roads, two different simulated safe houses, and one other.
3. Fort Indiantown Gap (FITG), an army training facility in Pennsylvania, on 19–20 June 2012.

A portion of the video was annotated by Mitre for the Mind's Eye Year 2 evaluation. That annotation is not included in LCA. After the completion of the Mind's Eye Year 2 evaluation, we undertook a systematic annotation effort for a portion of the above video. That annotation forms the basis of LCA. LCA contains all and only the portion of the above video that was annotated as part of this process. This video comprises 190 files as delineated in Table 1. Eight files are MOV format, 46 are MP4 format, and 136 are AVI format. The MOV files all use the MP4V codec and are 640×384 at 60 fps. The MP4 files all use the H264 codec and are 640×360 at 30 fps. The AVI files all use the XVID codec and are either 640×384 at 60 fps or 1440×1080 at 30 fps. This constitutes 2302144 frames and a total of 12 hours, 51 minutes, and 16 seconds of video. For comparison, UCF50 has 1330936 frames and

¹ <http://www.visint.org/>

Table 1 The original names of the files provided by DARPA. Filenames containing GPTC were filmed at GPJTC. Filenames containing STOPS were filmed at STOPS. Filenames consisting solely of a number were filmed at FITG. Numbers of the form YYYYMMDD indicate filming date. CR indicates country road. SH indicates safe house. Indices on CR, SH, and VT indicate variant backgrounds of the given class. CP1, CP2, C1, and C3 indicate camera. Text indicates the staging directions to guide filming. The remaining numbers serve to uniquely identify the video. These videos were renamed to video-XXX for consistency in the release. The release also contains a file, `video-mapping.txt`, which includes the mapping between the original filenames and those in the LCA release.

GPTC_20110822_SH_02_CP1_NOACTIVITY	STOPS_20120307_SH3_06_C1-edited-01	STOPS_20120308_CR1_06a_C1	100
GPTC_20110822_SH_02_CP2_NOACTIVITY	STOPS_20120307_SH3_06_C3-edited-01	STOPS_20120308_CR1_06a_C3	105
GPTC_20110822_SH_07_CP1_EX&RET	STOPS_20120307_SH3_03_C1-edited-01	STOPS_20120306_SH1_04_C1	114
GPTC_20110822_SH_07_CP2_EX&RET	STOPS_20120307_SH3_03_C3-edited-01	STOPS_20120306_SH1_04_C3	118
GPTC_20110822_SH_13_CP1_WARYHO	STOPS_20120307_SH3_02_C1-edited-01	STOPS_20120306_SH1_01_C1	11
GPTC_20110822_SH_13_CP2_WARYHO	STOPS_20120307_SH3_02_C3-edited-01	STOPS_20120306_SH1_01_C3	121
GPTC_20110822_SH_11_CP1_PARAHO	STOPS_20120307_SH3_01_C1-edited-01	STOPS_20120308_CR1_13a_C1	128
GPTC_20110822_SH_11_CP2_PARAHO	STOPS_20120307_SH3_01_C3-edited-01	STOPS_20120308_CR1_13a_C3	133
GPTC_20110822_SH_12_CP1_HOSTILEHO	STOPS_20120308_CR1_07a_C1	STOPS_20120309_VT1_07_C1	154
GPTC_20110822_SH_12_CP2_HOSTILEHO	STOPS_20120308_CR1_07a_C3	STOPS_20120309_VT1_07_C3	158
GPTC_20110822_SH_06_CP1_SUPCACHRET	STOPS_20120309_VT1_23_C1	STOPS_20120309_VT1_25_C1	162
GPTC_20110822_SH_06_CP2_SUPCACHRET	STOPS_20120309_VT1_23_C3	STOPS_20120309_VT1_25_C3	175
GPTC_20110822_SH_09_CP1_GIVE&CONT-BLDG	STOPS_20120308_CR1_NA_C1	STOPS_20120309_VT1_05_C1	188
GPTC_20110822_SH_09_CP2_GIVE&CONT-BLDG	STOPS_20120308_CR1_NA_C3	STOPS_20120309_VT1_05_C3	195
GPTC_20110822_SH_05_CP1_SUPCACHDUMP	STOPS_20120306_SH1_06_C1	STOPS_20120308_CR1_01c_C1	196
GPTC_20110822_SH_05_CP2_SUPCACHDUMP	STOPS_20120306_SH1_06_C3	STOPS_20120308_CR1_01c_C3	200
GPTC_20110822_SH_08_CP1_HO&EX-mixBKG	STOPS_20120306_SH1_NA_C1	STOPS_20120309_VT1_NA_C1	211
GPTC_20110822_SH_08_CP2_HO&EX-mixBKG	STOPS_20120306_SH1_NA_C3	STOPS_20120309_VT1_NA_C3	226
GPTC_20110822_SH_03_CP1_HO&RET	STOPS_20120308_CR1_08a_C1	STOPS_20120308_CR1_03a_C1	22
GPTC_20110822_SH_03_CP2_HO&RET	STOPS_20120308_CR1_08a_C3	STOPS_20120308_CR1_03a_C3	230
GPTC_20110822_SH_04_CP1_HO&RET2	STOPS_20120308_CR1_12a_C1	STOPS_20120306_SH1_02_C1	238
GPTC_20110822_SH_04_CP2_HO&RET2	STOPS_20120308_CR1_12a_C3	STOPS_20120306_SH1_02_C3	243
GPTC_20110822_CR_13_CP1_NOACTIVITY	STOPS_20120308_CR1_01a_C1	STOPS_20120307_SH3_11_C1	252
GPTC_20110822_CR_13_CP2_NOACTIVITY	STOPS_20120308_CR1_01a_C3	STOPS_20120307_SH3_11_C3	265
GPTC_20110822_CR_07_CP1_SALESMAN	STOPS_20120309_VT1_01_C1	STOPS_20120308_CR1_04a_C1	269
GPTC_20110822_CR_07_CP2_SALESMAN	STOPS_20120309_VT1_01_C3	STOPS_20120308_CR1_04a_C3	277
GPTC_20110822_CR_12_CP1_BAGDOWNHO&RET	STOPS_20120309_VT1_26_C1	STOPS_20120307_SH3_10_C1	30
GPTC_20110822_CR_12_CP2_BAGDOWNHO&RET	STOPS_20120309_VT1_26_C3	STOPS_20120307_SH3_10_C3	39
GPTC_20110822_CR_02_CP1_RoutineActivity	STOPS_20120308_CR1_02a_C1	STOPS_20120309_VT1_22_C1	41
GPTC_20110822_CR_02_CP2_RoutineActivity	STOPS_20120308_CR1_02a_C3	STOPS_20120309_VT1_22_C3	46
GPTC_20110822_CR_05_CP1_CONTHUR-NOHO	STOPS_20120308_CR1_05a_C1	STOPS_20120308_CR1_01b_C1	52
GPTC_20110822_CR_05_CP2_CONTHUR-NOHO	STOPS_20120308_CR1_05a_C3	STOPS_20120308_CR1_01b_C3	57
GPTC_20110822_CR_04_CP1_EX	STOPS_20120309_VT1_04_C1	STOPS_20120309_VT1_28_C1	59
GPTC_20110822_CR_04_CP2_EX	STOPS_20120309_VT1_04_C3	STOPS_20120309_VT1_28_C3	62
GPTC_20110822_CR_08_CP1_SALESMAN_INSIST	STOPS_20120306_SH1_07_C1	STOPS_20120309_VT1_20_C1	68
GPTC_20110822_CR_08_CP2_SALESMAN_INSIST	STOPS_20120306_SH1_07_C3	STOPS_20120309_VT1_20_C3	69
GPTC_20110822_CR_11_CP1_DOUBLEAVOIDBADGUY	STOPS_20120309_VT1_21_C1	STOPS_20120309_VT1_03_C1	81
GPTC_20110822_CR_11_CP2_DOUBLEAVOIDBADGUY	STOPS_20120309_VT1_21_C3	STOPS_20120309_VT1_03_C3	86
GPTC_20110822_CR_09_CP1_UPTONOGOOD	STOPS_20120307_SH3_07_C1	STOPS_20120309_VT1_27_C1	88
GPTC_20110822_CR_09_CP2_UPTONOGOOD	STOPS_20120307_SH3_07_C3	STOPS_20120309_VT1_27_C3	97
GPTC_20110822_CR_06_CP1_DISAGREE.NO GIVE.RETURN	STOPS_20120308_CR1_09a_C1	STOPS_20120309_VT1_13_C1	
GPTC_20110822_CR_06_CP2_DISAGREE.NO GIVE.RETURN	STOPS_20120308_CR1_09a_C3	STOPS_20120309_VT1_13_C3	
GPTC_20110822_CR_10_CP1_AVOIDBADGUY	STOPS_20120307_SH3_08_C1	STOPS_20120308_CR1_06b_C1	
GPTC_20110822_CR_10_CP2_AVOIDBADGUY	STOPS_20120307_SH3_08_C3	STOPS_20120308_CR1_06b_C3	
GPTC_20110822_CR_03_CP1_HO	STOPS_20120308_CR1_11a_C1	STOPS_20120309_VT1_02_C1	
GPTC_20110822_CR_03_CP2_HO	STOPS_20120308_CR1_11a_C3	STOPS_20120309_VT1_02_C3	
	STOPS_20120306_SH1_05_C1	STOPS_20120308_CR1_10a_C1	
	STOPS_20120306_SH1_05_C3	STOPS_20120308_CR1_10a_C3	
	STOPS_20120309_VT1_12_C1	STOPS_20120309_VT1_24_C1	
	STOPS_20120309_VT1_12_C3	STOPS_20120309_VT1_24_C3	
	STOPS_20120309_VT1_10_C1	STOPS_20120308_CR1_03b_C1	
	STOPS_20120309_VT1_10_C3	STOPS_20120308_CR1_03b_C3	

13.81 hours, HMDB51 has 632635 frames and 5.85 hours, UCF101 has 27 hours, Hollywood-2 has 20.1 hours, and VIRAT has 8.5 hours.

The Mind's Eye program specified a set of 48 verbs of interest. Of these, the LCA dataset uses only 24 verbs as annotation labels, as delineated in Table 1 in the manuscript.