

Multimodal Focused Interaction Dataset (Readme)

In total there are 19 videos in this dataset. Each folder is named with the captured video file and contains four types of files:

- Video: *.MP4
- Voice Activity Detection (VAD) scores: *.mat
- Sensor data: *.csv
- ELAN annotation: *.eaf
- GT annotations per video frame: GOPRXXXX_GT.csv
- 6 folds for cross-validation: focusedinteraction_6folds_split.txt

1. Video file

- Each video is of 1080p resolution sampled at 25 *fps* captured using a shoulder-mounted GoPro Hero 4. Audio of all videos has been removed due to privacy reason.

2. Voice Activity Detection (VAD) file

- VAD scores are made available for all videos.
- VAD scores are computed using the method proposed by [1]. Their implementation is available at <https://github.com/mvansegbroeck/vad>
- VAD scores are computed every 10 ms, i.e. at 100Hz, given an input audio stream with sampling rate of 8 kHz (the default setting proposed by [1]). We average four consecutive VAD scores, with a step size of four, to get the score at the same rate as that of the video features.

[1] Van Segbroeck, Maarten, Andreas Tsiartas, and Shrikanth Narayanan. "A robust frontend for VAD: exploiting contextual, discriminative and spectral cues of human voice." INTERSPEECH. 2013.

3. ELAN annotation file

- Install ELAN (<https://tla.mpi.nl/tools/tla-tools/elan/>) annotation tool to open *.eaf files.
- Each video is annotated for focused interaction, focused interaction walk, unfocused interaction and no interactions labels using the ELAN annotation tool.
- Each person appearing in the video has a unique ID (e.g. the camera wearer is labelled as Person0) in the ELAN annotation files.
- ELAN file can be read in Matlab using <https://github.com/marc-hanheide/SALEM>

4. Sensor data file

- Sensor_record_XXXXXXXX_XXXXXX_AndroSensor.csv file capture all sensors available on the smartphone that is placed in the front trousers' pocket of the camera wearer. The captured sensors include accelerometer, gyroscope, magnetometer, GPS sampled at 2 Hz.
- Andro Sensor App is used for logging the sensor data.
(<https://play.google.com/store/apps/details?id=com.fivasim.androsensor>)
- Video-Sensor Synchronisation Pulse
At the beginning of each recording session, the AndroSensor app and the camera record buttons were turned on. Since these devices started recording at slightly different times, the data streams from the two devices were not synchronised. A synchronisation pulse was

therefore generated with a clap action performed by the subject while holding the smartphone in the hand and in front of the camera. This pulse can be used for alignment if data from both devices are to be used. The smartphone was then placed in the subject's trouser pocket.

5. GT annotations per video frame

- GOPRXXXX_GT.csv contains ground-truth annotations for each video frame. The ground-truth labels are:
 - 0 – No focused interaction
 - 1 – Focused interaction (non-walk)
 - 2 – Focused interaction (walk)

6. 6 folds for cross-validation

- The file focusedinteraction_6folds_split.txt contains 6-fold split of the focused interaction dataset which has been used in our related publications. Each video name is indexed as follows:
 - 1- GOPR0177.MP4
 - 2- GOPR0184.MP4
 - 3- GOPR0185.MP4
 - 4- GOPR0188.MP4
 - 5- GOPR0193.MP4
 - 6- GOPR0194.MP4
 - 7- GOPR0196.MP4
 - 8- GOPR0198.MP4
 - 9- GOPR0201.MP4
 - 10- GOPR0202.MP4
 - 11- GOPR0203.MP4
 - 12- GOPR0204.MP4
 - 13- GOPR0207.MP4
 - 14- GOPR0208.MP4
 - 15- GOPR0209.MP4
 - 16- GOPR4037.MP4
 - 17- GOPR4038.MP4
 - 18- GOPR4041.MP4
 - 19- GOPR4042.MP4

7. License

The Multimodal Focused Interaction Dataset dataset is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0).

<https://creativecommons.org/licenses/by/4.0/>

8. Contact

For comments, suggestions or feedback, or if you experience any problems with this website or the dataset, please contact Stephen McKenna (s.j.z.mckenna@dundee.ac.uk).