

GEINTRA Overhead ToF People Detection 3 (GOTPD3) Database: Human Activity Detection

Javier Macías-Guarasa and Cristina Losada-Gutiérrez and David Fuentes-Jiménez

October 10, 2019

1 Introduction

This document describes the generalities of the GEINTRA Overhead ToF People Detection 3 database (GOTPD3 from now on)

The GOTPD3 is a multimodal images database containing depth and infrared video-sequences, recorded with a Kinect 2 camera located in overhead position, and monitoring people movements under it, and it was designed to fulfill the following objectives:

- Allow evaluation and fine tuning of the ToF data acquisition system in the GEINTRA research group.
- Allow the evaluation of people detection and human activity detection algorithms based on data generated by ToF cameras (including depth and infrared) in top-view configuration.
- Provide quality data to the research community in people detection and identification tasks.

The people detection task (and the data provided) can also be extended to practical applications such as video-surveillance, access control, people flow analysis, behaviour analysis or event capacity management.

2 Database description

GOTPD3 is composed of 9887 depth and infrared images recorded with a Kinect 2 camera placed in a zenithal position at a height of 3.4 meters, oriented perpendicularly to the floor, and monitoring people movements under it. These images are divided into 34 video-sequences corresponding to 7 people (with different height, hair style and complexion) performing 5 actions: *Walk slowly*, *Run*, *Walk fast*, *Falling down* and *Walk, stop briefly under the camera and continue walking*. Each of the available sequences correspond to only one person performing only one of the actions. Table 1 shows the code assigned to each action as well as the number of depth and infrared frames.

Action	Code	Number of frames
Walk slowly	000	2820
Run	001	1377
Walk fast	002	1764
Falling down	003	1932
Walk normal, stop, continue walking	004	1992
Total		9887

Table 1: Code and number of available frames (depth and infrared) for each action.

Quantitative details on the database content are provided below.

- **Number of frames:** 9887 (depth) + 9887 (infrared)
- **Number of different people:** 7 (5 men and 2 women)
- **Number of actions:** 54
- **Image resolution:** 512×424 pixels

For each image, there are provided the depth map and the infrared image. To give you an idea on what to expect, figure 1 shows some examples of images from the dataset. The images correspond to three different people performing the actions walk slowly, falling down and standing up under the camera. The first row represents the depth maps whereas the second one shows the corresponding infrared images.

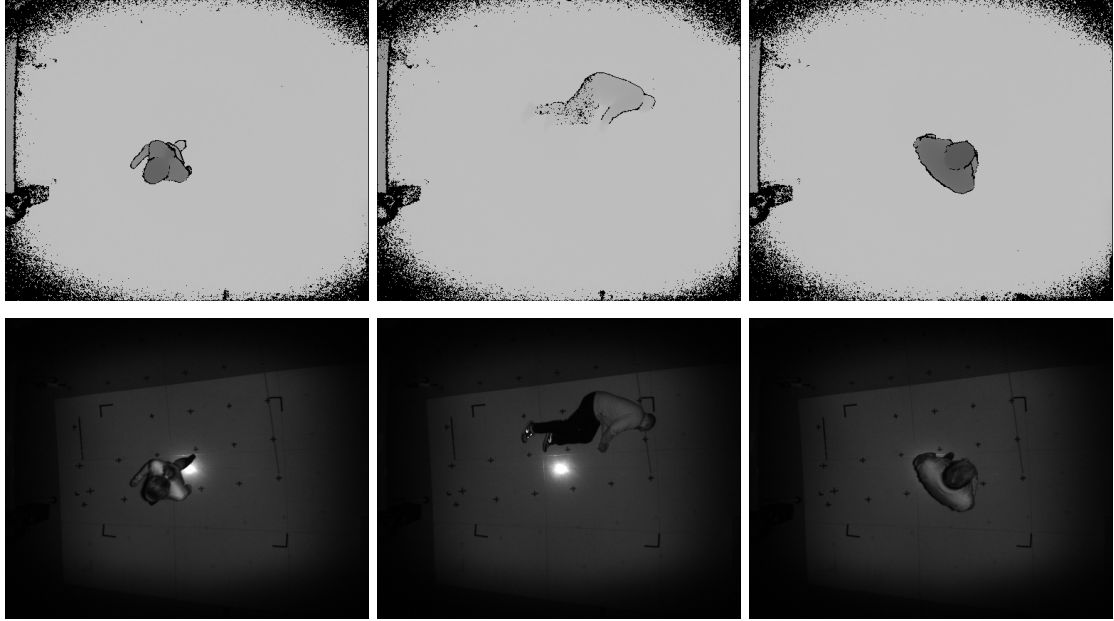


Figure 1: Sample images corresponding to three different people performing diverse actions.

3 File formats and extensions

3.1 Depth data

The depth information (distance to the camera plane) is stored as binary `.z16` file. For each depth map, each pixel represents the depth value in millimeters as a (little endian) unsigned integer of two bytes. Its values range from 0 to 4500.

The infrared intensity information is also stored in plain binary form as a (little endian) signed integer of two bytes, in files with extension `.ir16`.

4 Labeling and file naming conventions

To ease adapting the experimental setup for specific tasks, we have designed a (verbose) naming convention for the file names, that includes information about both, the person who appears in the video and the performed action. Thus, each file is named following this structure:

`actXXX-YYY`, where:

- **XXX:** Person information. `XXX` corresponds to the initialism that stands for the person performing the action.

- **YYY**: Activity information. *YYY* is written in decimal but is meant to refer to a bitmask, as shown in table 1.

All the variables defined above are left zero-padded, so that parsing the filenames is trivial.

4.1 Recorded sequences

Table 2 shows the list of recorded sequences, with the following contents per column:

- **Sequence ID**: The file basename following the convention described in section 4. The same name is used for both, depth and infrared sequences.
- **Action ID**: The code corresponding to the action performed on the sequence as shown in table 1
- **Person ID**: The ID of the persons that appear in the corresponding sequence. This ID is unique along all the recorded sequences, so that it can be used to rigourously select the training/testing partitions, so that no user used in training is used in the evaluation procedures.
- **#Frames**: The number of recorded frames.

5 ToF Camera Specifications

The camera used in our recordings is a Kinect 2 for windows device, with the following main characteristics [1]:

- **Depth sensing**
 - 512 x 424
 - 30 Hz
 - FOV: 70 x 60
 - One mode: 0.5–4.5 meters
- **1080p color camera (not used for the dataset)**
 - 30 Hz (15 Hz in low light)
- **Active infrared (IR) capabilities**
 - 512 x 424
 - 30 Hz
- **Microphone array (4 microphones)**

6 Disclaimer, Licensing, Request and Contributions

This document and the data provided are work in progress and provided as is.

The GEINTRA Overhead ToF People Detection 3 (GOTPD3) Database: Human Activity Detection (and accompanying files and documentation) by Javier Macías-Guarasa, Cristina Losada-Gutiérrez and David Fuentes-Jiménez is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](#).

To request a copy of the dataset, please contact [Javier Macias-Guarasa](mailto:javier.maciasguarasa@uah.es) at javier.maciasguarasa@uah.es.

If you make use of this database and/or its related documentation, you are kindly requested to cite the paper [2].

Also, if you derive additional data, information, publications, etc., using GOTPD3, please [tell us](#) so that we can also publicite your contributions.

Sequence ID	Action ID	Person ID	# frames
act-CLG-000	000	CLG (woman)	417
act-CLG-001	001		184
act-CLG-002	002		247
act-CLG-003	003		288
act-CLG-004	004		320
act-DCP-000	000	DCP (man)	329
act-DCP-001	001		184
act-DCP-002	002		201
act-DCP-003	003		350
act-DCP-004	004		238
act-DFJ-000	000	DFJ (man)	483
act-DFJ-001	001		219
act-DFJ-002	002		289
act-DFJ-003	003		265
act-DFJ-004	004		249
act-FVS-000	000	FVS (man)	382
act-FVS-001	001		195
act-FVS-002	002		238
act-FVS-003	003		292
act-FVS-004	004		293
act-JMG-000	000	JMG (man)	419
act-JMG-001	001		198
act-JMG-002	002		241
act-JMG-004	004		291
act-MMR-000	000		MMR (woman)
act-MMR-001	001	226	
act-MMR-002	002	297	
act-MMR-003	003	418	
act-MMR-004	004	370	
act-PHB-000	000	PHB (man)	348
act-PHB-001	001		171
act-PHB-002	002		251
act-PHB-003	003		319
act-PHB-004	004		233

Table 2: Detailed information about the dataset

References

- [1] Microsoft, “Kinect hardware,” <https://developer.microsoft.com/en-us/windows/kinect/hardware>, accessed june 2016.
- [2] C. A. Luna, C. Losada, D. Fuentes-Jimenez, A. Fernandez-Rincon, M. Mazo, and J. Macias-Guarasa, “Robust people detection using depth information from an overhead time-of-flight camera,” *Expert Systems with Applications*, vol. 71, pp. 240–256, 04/2017 2017.