GEINTRA Overhead ToF People Detection (GOTPD1) Database Description

Javier Macias-Guarasa, Cristina Losada-Gutierrez, David Fuentes-Jimenez,
Raquel Garcia-Jimenez, Carlos A. Luna, Alvaro Fernandez-Rincon, and Manuel Mazo

June 23, 2016

1 Introduction

This document describes the generalities of the GEINTRA Overhead ToF People Detection 1 database (GOTPD1 from now on)[1].

The GOTPD1is a multimodal database (depth and infrared data) of recordings from a Kinect 2 camera located in overhead position, monitoring people movements under it, and it was designed to fulfill the following objectives:

- Allow evaluation and fine tuning of the ToF data adquisition system in the GEINTRA research group.
- Allow the evaluation of people detection algorithms based on data generated by ToF cameras (depth and infrared) placed in overhead position.
- Provide quality data to the research community in people detection 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. In case that only the depth information is used, we can also guarantee that privacy is preserved.

2 General contents

GOTPD1is composed of 48 sequences comprising a broad variety of conditions, with scenarios comprising:

- Single and multiple persons
- Single and multiple non-persons (such as chairs)
- Persons with and without accessories (hats, caps)
- Persons with different complexity, height, hair color, and hair configuration
- Persons actively moving and performing additional actions (such as using their mobile phones, moving their fists up and down, moving their arms, etc.).

The actual video footage is over 28 minutes, with sequence lengths ranging from 4 seconds to 2.63 minutes. Quantative details on the database content are provided in Table 1.

To give you an idea on what to expect, you can have a look at the following videos we prepared from the data (the colormap used is not the best, but enough to understand what is going on):

[1] You can get the latest version of this document at http://www.geintra-uah.org/archivos/GOTPD1/GOTPD1-readme.pdf

Table 1: General quantitative details on `GOTPD1`.

	Numer of frames	Time length	Labelled persons
Total	51418	1713,93 seconds – 28.57 min	21093
Max across sequences	4728	157,6 seconds – 2.62 minutes	4164
Min across sequencs	277	9.23 seconds	21

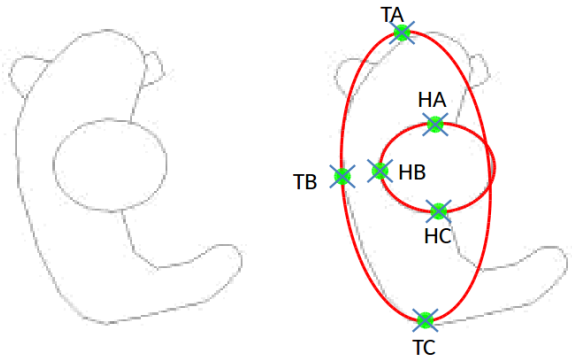


Figure 1: Labeling.

- Persons moving randomly
- Persons moving grouped
- Person moving fists up and down

And you can even get a sample of the provided data in the corresponding `tgz file` (see file format details on section 5).

3 Geometry details

The camera was placed in a zenithal position at a height of 3.4 meters, oriented perpendicularly to the floor.

4 Labeling procedure

For the labeling procedure, we modified the Head Annotation Interface developed by Guillaume Lathoud for the AV16.3 corpus [1], to fulfill our labeling requirements.

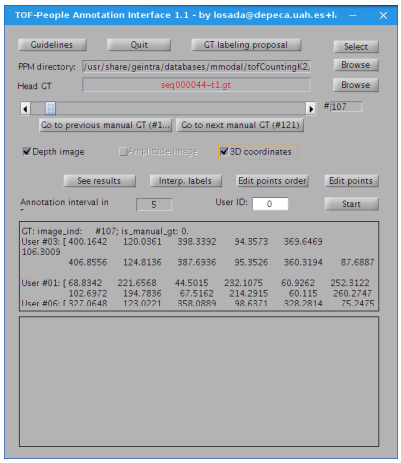
The specific items to label per frame are the ones needed to define the two main physical entities required for our planned detection task, namely, the head contour and the torso contour. To label them, it's expected that using symmetry considerations, we will only need to label tree points per contour (HA, HB, HC for the head, and TA, TB, TC for the torso), as shown in Figure 1.

From this information, it is easy to obtain additional information, such as the head position centroid.

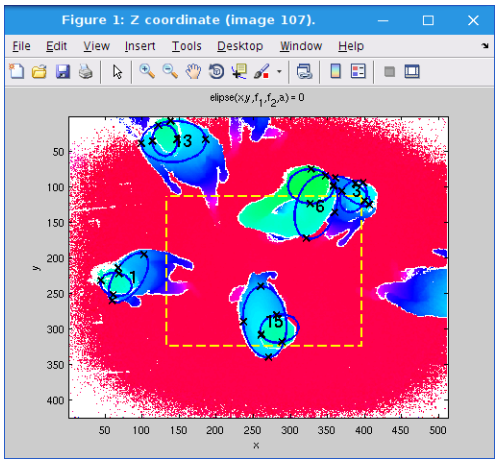
Figure 2 shows some screenshots of the labelling tool.

Labelling is done manually every n frames (usually, $n = 5$ in our recordings, but a limit of 1 second between manual labels is guaranteed), and an automatic interpolation procedure generates the trajectories between manual labels. The ground truth file includes information on which frames have been labelled manually or by interpolation.

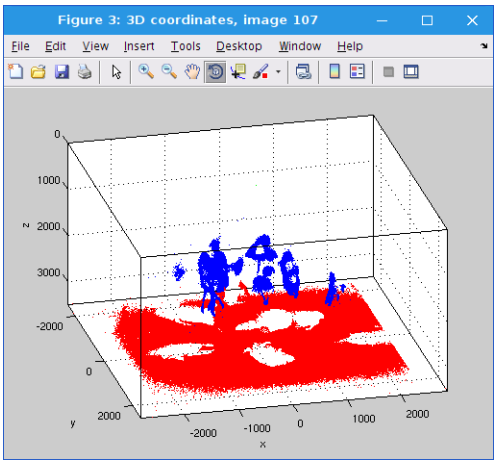
In the areas with high noise or for specially difficult image conditions, the labeling of the torso contour may not be very precise, although the area for the head is usually correct.



(a) Annotation info screen.



(b) Depth + labels window.



(c) 3D point cloud window.

Figure 2: Labeling tool adapted from the Head Annotation Interface developed by Guillaume Lathoud for the AV16.3 corpus [1].

5 File formats

5.1 Depth (height) data

The depth information (distance to the camera plane) in stored in plain binary form, with each pixel distance represented in millimeters as a (little endian) signed integer of two bytes. Its values range from 0 to 4500.

If $Z_{i,j}$ is the depth value for pixel i, j, the image matrix would be:

$$\begin{bmatrix} Z_{0,0} & Z_{1,0} & Z_{2,0} & \cdots & Z_{N_W-2,0} & Z_{N_W-1,0} \\ Z_{0,1} & Z_{1,1} & Z_{2,1} & \cdots & Z_{N_W-2,1} & Z_{N_W-1,0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ Z_{0,N_H-2} & Z_{1,N_H-2} & Z_{2,N_H-2} & \cdots & Z_{N_W-2,N_H-2} & Z_{N_W-1,N_H-2} \\ Z_{0,N_H-1} & Z_{1,N_H-1} & Z_{2,N_H-1} & \cdots & Z_{N_W-2,N_H-1} & Z_{N_W-1,N_H-1} \end{bmatrix} \quad (1)$$

where N_W is the image width, and N_H the image height, both of them in pixels ($N_W = 512$ and $N_H = 424$ for the Kinect 2).

From this depth matrix, the write order in the file is per rows, so that the values in the file would be:

$$[Z_{0,0}\ Z_{1,0}\ \ldots\ Z_{N_W-1,0}\ Z_{0,1}\ Z_{1,1}\ \ldots\ Z_{N_W-1,0}\ \ldots\ Z_{0,N_H-1}\ \ldots\ Z_{N_W-2,N_H-1}\ Z_{N_W-1,N_H-1}]$$

5.2 Infrared data

The infrared intensity information in also stored in plain binary form as a (little endian) signed integer of two bytes.

The write order is exactly the same as with the depth data.

5.3 Position Ground Truth Data

The ground truth information is provided as a text file, with each line using the following format:

FFFFFF X [UID_i $\text{HA}x_i$ $\text{HA}y_i$ $\text{HB}x_i$ $\text{HB}y_i$ $\text{HC}x_i$ $\text{HC}y_i$ $\text{TA}x_i$ $\text{TA}y_i$ $\text{TB}x_i$ $\text{TB}y_i$ $\text{TC}x_i$ $\text{TC}y_i$]$*$

where:

- FFFFF is the frame number, starting at 000000. Six digits are used and they are zero padded.
- X indicates whether the ground truth information has been generated manually (X = 1), or automatically (X = 0).
- [UID_i $\text{HA}x_i$ $\text{HA}y_i$ $\text{HB}x_i$ $\text{HB}y_i$ $\text{HC}x_i$ $\text{HC}y_i$ $\text{TA}x_i$ $\text{TA}y_i$ $\text{TB}x_i$ $\text{TB}y_i$ $\text{TC}x_i$ $\text{TC}y_i$] is the information for person i in the scene (the asterisk $*$ indicates that there maybe from 0 up to an arbitraty number of persons in the scene, with $i = 0, 1, ...$):
  - UID_i is the person ID number
  - $\text{HA}x_i$ $\text{HA}y_i$ are the image plane coordinates (x, y) for point HA in Figure 1
  - $\text{HB}x_i$ $\text{HB}y_i$ are the image plane coordinates (x, y) for point HB in Figure 1
  - $\text{HC}x_i$ $\text{HC}y_i$ are the image plane coordinates (x, y) for point HC in Figure 1
  - $\text{TA}x_i$ $\text{TA}y_i$ are the image plane coordinates (x, y) for point TA in Figure 1
  - $\text{TB}x_i$ $\text{TB}y_i$ are the image plane coordinates (x, y) for point TB in Figure 1
  - $\text{TC}x_i$ $\text{TC}y_i$ are the image plane coordinates (x, y) for point TC in Figure 1

If you want to get just the centroid of the head, the calculation is very simple. If $\text{C}_i = (\text{C}x_i\ \text{C}y_i)$ are the image plane coordinates of the head labelled centroid, you can apply the following trivial equations (see Figure 1 for geometrical references):

$\text{HC}x_i$ $\text{HC}y_i$

$$\text{C}x_i = \text{HC}x_i + \frac{\text{HA}x_i - \text{HC}x_i}{2}; \text{C}y_i = \text{HC}y_i + \frac{\text{HA}y_i - \text{HC}y_i}{2}; \quad (2)$$

6 File naming conventions

6.1 File names

To ease adapting the experimental setup for specific tasks, we have designed a (verbose) naming convention for the file names.

Each file is named following this structure: `seq-P`XX`-M`YY`-A`$UUUU$`-G`XX`-C`WW`-S`VVV, where:

- `P`XX: Number of persons in the scene. XX is the maximum number of people than can be seen simultaneously in the scene. Note that there may be multiple users recorded in a given sequence, but at most XX will be seen at the same time.
- `M`YY: Movement information. YY is written in decimal but is meant to refer to a bitmask, with the following convention:
  - `00` N/A

  - `01` static
  - `02` mostly regular around scene
  - `04` mostly random
  - `08` reduced movements (almost static, probably turning on)

- `A`*UUUU*: Activity information. *UUUU* is written in decimal but is meant to refer to a bitmask, with the following convention:
  - `0000` N/A
  - `0001` Normal walking
  - `0002` Looking to smartphone (texting) or looking to floor
  - `0004` Talking to phone on ear (phone call)
  - `0008` Facing fists up and down
  - `0016` Standing still moving up and down
  - `0032` Moving chairs
  - `0064` User pushing chair
  - `0128` User making squats to modify his/her height
  - `0256 Standing nearby image border (turning on)`

- `G`*XX*: Grouping information. *XX* is written in decimal but is meant to refer to a bitmask, with the following convention:
  - `00` N/A
  - `01` mostly not forced
  - `02` mostly forced to be close

- `C`*WW*: Accesories information. *WW* is written in decimal but is meant to refer to a bitmask, with the following convention:
  - `00` No
  - `01` Some users have hats
  - `02` All user have hats
  - `04` Some users have caps
  - `08` All user have caps

- `S`*VVVV*: Sequence number information. *VVVV* is the sequence number and is unique across the database.

All the variables defined above (*XX*, *YY*, *UUUU*, *XX*, *WW*, *VVV*) are left zero-padded, so that parsing the filenames is trivial.

6.2 Filename extensions

The distributed filenames have an extension that identifies their type, as follows:

- `z16`: Depth information file.
- `ir16`: Infrared information file.
- `gt`: Ground truth information.
- `info`: Contains detail on the number of frames, file length (in secods), frames labelled, and persons labelled.

6.3 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 6.1.
- Person IDs: The IDs 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.
- Comment: A general comment on what to expect on the sequence.
- #Frames: The number or recorded frames.
- Lengh (s): The time length of the recorded sequence (in seconds)
- Labelled persons[2]: The number of persons labelled in the ground truth files (note that this can be higher than the number of frames, as there may be multiple users labelled per frame.

[2] -1 indicates that the ground truth file is not yet available for distribution

Table 2: Detailed DB information

Sequence ID	Person Ids	Comment	# Frames	Length (s)	Labelled persons
seq-P01-M02-A0001-G00-C00-S0001	000000 000014	Single person moving regularly along the space	3685	122.83	861
seq-P01-M02-A0001-G00-C00-S0002	000013 000003	Single person moving regularly along the space	3924	130.80	801
seq-P01-M02-A0001-G00-C00-S0003	000006 000006	Single person moving regularly along the space	3566	118.86	594
seq-P01-M02-A0001-G00-C00-S0004	000001 000012	Single person moving regularly along the space	3732	124.40	771
seq-P01-M02-A0001-G00-C00-S0005	000000	Single person moving regularly along the space	1700	56.66	1700
seq-P01-M02-A0001-G00-C00-S0006	000017	Single person moving regularly along the space	418	13.93	410
seq-P01-M02-A0001-G00-C00-S0007	000017	Single person moving regularly along the space	413	13.76	410
seq-P01-M02-A0001-G00-C00-S0008	000017	Single person moving regularly along the space	502	16.73	500
seq-P01-M02-A0001-G00-C00-S0009	000017	Single person moving regularly along the space	390	13.00	390
seq-P01-M02-A0001-G00-C00-S0010	000017	Single person moving regularly along the space	580	19.33	580
seq-P01-M02-A0001-G00-C00-S0011	000007	Single person moving regularly along the space	334	11.13	334
seq-P01-M04-A0001-G00-C00-S0012	000020	Single person moving randomly along the space	417	13.90	330
seq-P01-M04-A0002-G00-C00-S0013	000020	Person looking down to phone screen	516	17.20	456
seq-P01-M04-A0001-G00-C00-S0014	000034	Single person moving randomly along the space	473	15.76	380
seq-P01-M04-A0001-G00-C00-S0015	000028	Single person moving randomly along the space	624	20.79	536
seq-P01-M04-A0002-G00-C00-S0016	000028	Person looking down to phone screen	514	17.13	510
seq-P01-M04-A0001-G00-C00-S0017	000035	Single person moving randomly along the space	615	20.50	540
seq-P01-M04-A0002-G00-C00-S0018	000035	Person looking down to phone screen	333	11.10	333
seq-P01-M02-A0001-G00-C00-S0019	000038	Single person moving regularly along the space	783	26.10	680
seq-P01-M02-A0001-G00-C00-S0020	000007	Single person moving regularly along the space	332	11.06	332
seq-P01-M02-A0001-G00-C00-S0021	000007	Single person moving regularly along the space	380	12.66	380
seq-P01-M02-A0001-G00-C00-S0022	000007	Single person moving regularly along the space	277	9.23	277
seq-P02-M02-A0001-G02-C00-S0023	000000 000014 000013 000003	Two persons moving regularly along the space, keeping close to each other	2047	68.23	532
seq-P02-M02-A0001-G02-C00-S0024	000015 000006 000001 000012	Two persons moving regularly along the space, keeping close to each other	1526	50.86	689
seq-P02-M02-A0001-G02-C04-S0025	000000 000014 000013 000003	Two persons moving regularly along the space, keeping close to each other, one of them wearing a cap	1766	58.86	593
seq-P02-M02-A0001-G02-C02-S0026	000015 000006 000001 000012	Two persons moving regularly along the space, keeping close to each other, one of them wearing a hat	1799	59.96	602
seq-P07-M02-A0001-G02-C00-S0027	000014 000013 000003 000015 000006 000001 000012	Persons moving vertically along the image, forming a line (parallel to the movement), two of them cannot be seen as they are walking on the image floor border	421	14.03	415

... *Continued on next page* ...

Table 2 Detailed DB information (*Continued from previous page*)

Sequence ID	Person Ids	Comment	# Frames	Length (s)	Labelled persons
seq-P08-M04-A0001-G01-C00-S0028	000000 000014 000013 000003 000015 000006 000001 000012	8 persons moving randomly	1911	63.70	4164
seq-P08-M02-A0001-G02-C00-S0029	000000 000014 000013 000003 000015 000006 000001 000012	8 persons moving regularly very close each other	891	29.70	1098
seq-P05-M04-A0001-G03-C00-S0030	000002 000016 000000 000017 000018	5 persons moving randomly, some of them very close each other	508	16.93	1815
seq-P05-M04-A0001-G01-C00-S0031	000016 000000 000019 000022	5 persons moving randomly, some of them very close each other at times	536	17.86	1898
seq-P05-M04-A0001-G01-C05-S0032	000016 000000 000019 000022 000018	5 persons moving randomly, some of them very close each other at times, two of them wearing a hat, and one of them wearing a cap	582	19.40	2349
seq-P04-M04-A0001-G01-C00-S0033	000002 000016 000000 000017	4 persons moving quite regularly keeping close to each other	440	14.66	1325
seq-P05-M04-A0001-G01-C05-S0034	000002 000016 000000 000017 000018	5 persons moving randomly, some of them very close each other at times, two of them wearing a hat, and one of them wearing a cap	545	18.16	2045
seq-P01-M02-A0032-G00-C00-S0035	000000	Chairs moving vertically along the image middle vertical. Person is eventually seen to recover chair	891	29.70	21
seq-P00-M02-A0032-G00-C00-S0036	-	Chairs moving vertically along the image left vertical	868	28.93	0
seq-P00-M02-A0032-G00-C00-S0037	-	Chairs spinning and moving vertically along the image middle vertical	921	30.70	0
seq-P01-M02-A0008-G00-C00-S0038	000017	Person with arms in square angle, fists up, moving them up&down, wears hood not covering head (in back)	485	480	480
seq-P01-M02-A0004-G00-C00-S0039	000017	Person speaking on the phone, wears hood not covering head (in back)	446	14.86	435
seq-P01-M02-A0064-G00-C00-S0040	000017	Person pushing a chair	501	16.70	495
seq-P01-M02-A0001-G00-C00-S0041	000009	Single person moving regularly along the space	714	23.79	-1
seq-P01-M02-A0001-G00-C00-S0042	000021	Single person moving regularly along the space	534	17.79	-1
seq-P01-M02-A0001-G00-C00-S0043	000023	Single person moving regularly along the space	873	29.10	-1
seq-P01-M02-A0001-G00-C00-S0044	000024	Single person moving regularly along the space	749	24.96	-1
seq-P01-M04-A0001-G00-C00-S0045	000026	Single person moving randomly along the space	354	11.80	-1

. . . *Continued on next page* . . .

Table 2 Detailed DB information (*Continued from previous page*)

Sequence ID	Person Ids	Comment	# Frames	Length (s)	Labelled persons
seq-P01-M04-A0001-G00-C00-S0046	000036	Single person moving randomly along the space	433	14.43	-1
seq-P01-M02-A0001-G00-C02-S0047	000000 000014 000013 000003 000015 00000	Single person with hat	4728	157.60	-1
seq-P01-M02-A0001-G00-C02-S0048	000001 000012	Single person with hat	1441	48.03	-1

7 ToF Camera Specifications

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

- Depth sensing
  - 512 x 424
  - 30 Hz
  - FOV: 70 x 60
  - One mode: 0.5–4.5 meters
- 1080p color camera
  - 30 Hz (15 Hz in low light)
- Active infrared (IR) capabilities
  - 512 x 424
  - 30 Hz
- Microphone array (4 microphones)

8 Disclaimer, Licensing, Request and Contributions

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

To request a copy of the dataset, please contact Javier Macias-Guarasa at macias@depeca.uah.es.

If you derive additional data, information, publications, etc., using `GOTPD1`, please tell us so that we can also publicite your contributions.

References

[1] G. Lathoud, J.-M. Odobez, and D. Gatica-Perez, "Av16.3: An audio-visual corpus for speaker localization and tracking," in *Procceedings of the MLMI*. Springer-Verlag, 2004, pp. 182–195.

[2] Microsoft, "Kinect hardware," https://developer.microsoft.com/en-us/windows/kinect/hardware, accessed june 2016.