GEINTRA Synthetic Depth People Detection (GESDPD) Database Description

David Fuentes-Jiménez, Roberto Martín-López, Cristina Losada-Gutiérrez, Javier Macías-Guarasa and Carlos A. Luna

July 26, 2019

1 Introduction

This document describes the generalities of the GEINTRA Synthetic Depth People Detection database (GESDPD from now on) [1].

The GESDPD is a depth images database containing 22000 frames , that simulates to have been taken with a sensor in an elevated front position, in an indoor environment. It was designed to fulfill the following objectives:

- Allow the train and evaluation of people detection algorithms based on depth , or RGB-D data, without the need of manually labeling.
- Provide quality synthetic data to the research community in people detection tasks.

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

2 General contents

GESDPD is composed of 22000 depth synthetic images, that simulates to have been taken with a sensor in an elevated front position, in a rectangular, indoor working environment. These have been generated using the simulation software Blender [1].

The synthetic images show a room with different persons walking in different directions. The camera perspective is not stationary, it moves around the room along the database, which avoids a constant background. Some examples of the different views are shown in figure 1

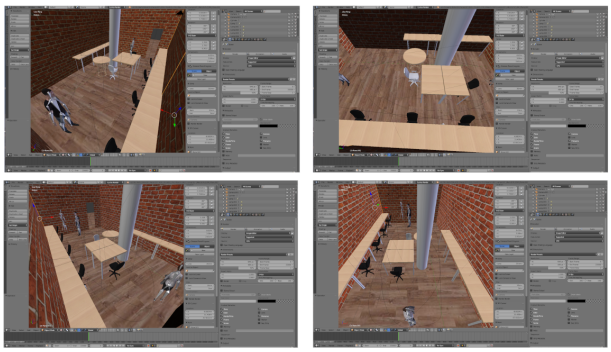

Figure 1: Example of different views of the simulated room.

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

Quantitative details on the database content are provided below.

- **Number of frames:** 22000
- **Number of different people:** 4 (3 men and 1 woman)
- **Number of labeled people:** 20800
- **Image resolution:** 320×240 pixels

For each image, the are provided the depth map and the ground truth including the position of each person in the scene.

To give you an idea on what to expect, figure 2 shows some examples of images from the dataset. In this figure, depth values are represented i millimeters, using a colormap.

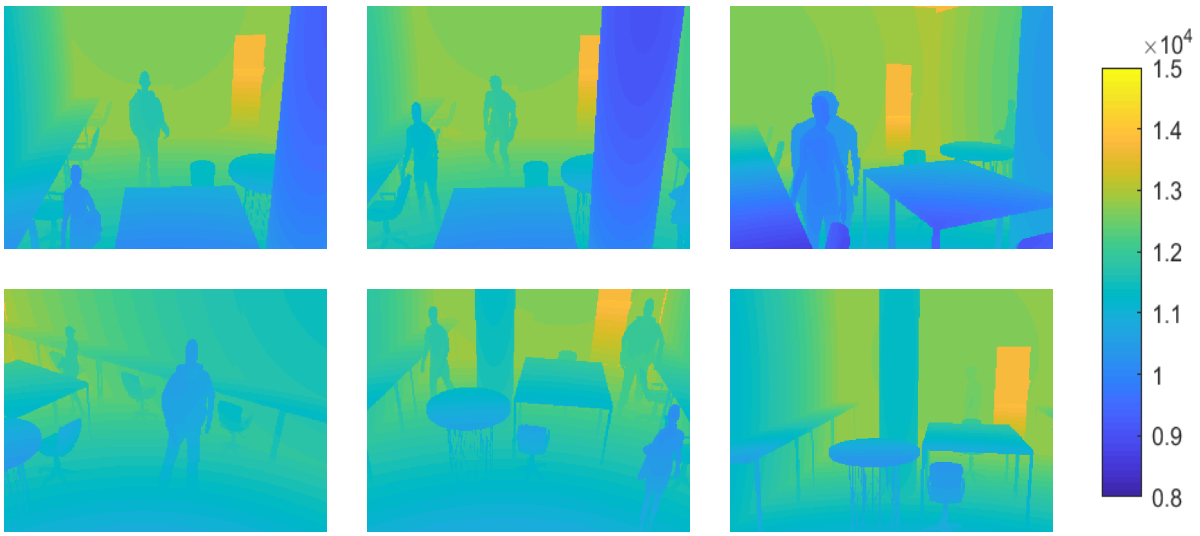

Figure 2: Example of depth images belonging to GESDPD dataset showing depth values using a color map.

3 Geometry details

As it has been said before, the dataset simulates to have been taken with a sensor in an elevated front position, in a rectangular indoor working environment. Specifically, the camera it placed at a height of 3 meters, and it rotates along the sequence.

Regarding the room (whose distribution is shown in figure 3), its dimensions are $8.56 \times 5.02m$, and it has a height of $3.84m$.

4 File formats

4.1 Depth data

The depth information (distance to the camera plane) in stored as a .png image, in which each pixel represent the depth value in millimeters as a (little endian) unsigned 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:

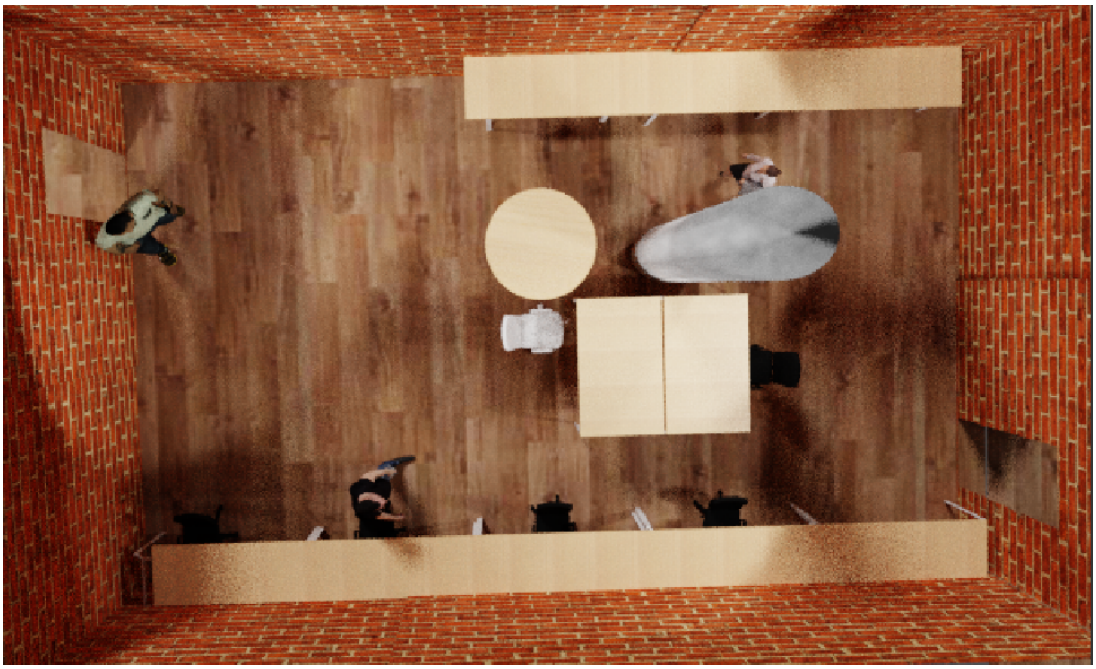

Figure 3: Zenithal view of the simulated room

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

where N_W is the image width, and N_H the image height, both of them in pixels ($N_W = 320$ and $N_H = 240$ for our images).

Sicne depth data is stored in PNG16 format, it should be scaled, using equation 2:

$$d(x,y) = \frac{D_{max} P(x,y)}{2^{16} - 1} \quad (2)$$

where d is the depth value in millimeters and D_{max} is the maximum distance in the dataset images (that is set to $D_{max} = 15000mm$).

4.2 Position Ground Truth Data

The ground truth information is also provided as a .png file, with the same dimensions that the generated images (320 × 240 pixels). The ground truth files have in their names the same number than the corresponding depth files.

For labeling people positions, there have been placed Gaussian functions over the centroid of the head of each person in the scene, so that the centroid corresponds to the 2D position of the center of the head and has a normalized value of one. The standard deviation has a value of 15 pixels for all the Gaussians, regardless of the size of each head and the distance from the head to the camera. This value has been calculated based on an estimated value of the average diameter of a person head, taking into account anthropometric considerations.

It is worth to highlight that, when two heads are very closely or overlapping with each other, instead of adding both Gaussian functions, the maximum value of them prevail. That modification provides a set of Gaussians that are always separated, so that the CNN can learn to generate that separation between Gaussians in its output. Figure 4 shows an example of two Gaussian functions.

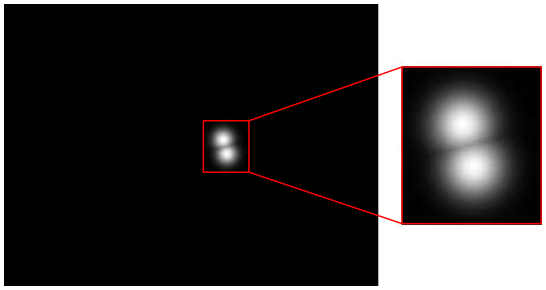

Figure 4: Labeled Gaussians Detail.

5 Disclaimer, Licensing, Request and Contributions

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

The GEINTRA Synthetic Depth People Detection (GESDPD) Database (and accompanying files and documentation) by David Fuentes-Jiménez, Roberto Martín-López, Cristina Losada-Gutiérrez, Javier Macías-Guarasa and Carlos A. Luna is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

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

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

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

References

[1] T. Roosendaal, R. Hess, and B. Foundation, *The Essential Blender: Guide to 3D Creation with the Open Source Suite Blender*, ser. No Starch Press Series. No Starch Press, 2007. [Online]. Available: https://books.google.es/books?id=Y0N5NAAACAAJ

[2] R. M. López, D. F. Jiménez, C. L. Gutiérrez, and C. L. Vázquez, "Detección de personas en imágenes de profundidad mediante redes neuronales convolucionales," in *Libro de actas del XXVI Seminario Anual de Automática, Electrónica Industrial e Instrumentación*, 2019, pp. 114–119.