GEINTRA Multiple Actions Dataset in Cruises (`GSMADC`)

Carlos Cob-Parro, Cristina Losada-Gutiérrez, Marta Marrón-Romera,
Rodrigo A. Vidal Pinto, Ignacio Bravo, Alfredo Gardel

April 3, 2024

1 Introduction

This document summarizes the main characteristics of the dataset created by the research group GEIN-TRA[1], at the University of Alcalá, named GEINTRA Multiple Actions Dataset in Cruises (`GSMADC` from now on).

The `GSMADC` dataset has been developed within the framework of the *PALAEMON: A holistic passenger ship evacuation and rescue ecosystem* European project (H2020-PALAEMON-814962). Furthermore, we would like to thanks Astilleros de Santander S.A.U. (ASTANDER)[2], for allowing the recording of the sequences in an ship.

The `GSMADC` is a multimodal images database containing depth and infrared video-sequences, recorded with an Intel Real Sense D435 camera. The dataset contains a total of 74 videos with an average length of 40 seconds each. Those videos are equally divided into RGB and Infrared (IR) as the used camera provides both RGB and IR recordings, which have been taken in different scenarios included in the cruise ship.

The used camera (Intel Realsense D435) contains a low-power vision processor for real-time depth sensing. An IR projector and a RGB module are also integrated, which allows the camera to output the same video recorded in the two different formats (IR and RGB). Each of these modules has its own characteristics, what are presented in table 1.

Module	Resolution (pixels)	Frame rate	Field of View
IR	1920x1080	90 fps	$87^o \times 58^o$
RGB	1280x720	30 fps	$69^o \times 42^o$

Table 1: Characteristics of the Intel Realsense D435 camera used for the acquisiton of the `GSMADC` dataset.

2 Recording scenarios

The scenarios in which the recordings of the `GSMADC` dataset have taken place are shown along with a brief description of them below. In each figure, there is shown a sample image of the corresponding scenario. It is worth noting that all the sample images shown in this section includes both RGB and IR images from each scenario.

1. **Indoor:** consists of a long corridor with mirrors at its sides that has three main exits (sections of the scenario in from which persons can enter and leave the scene): the beginning and the end of the corridor and the stairs located at the beginning of it. Figure 1 shows the scenario. It is worth noting that all the sample images shown in this section includes both RGB and IR images from each scenario.

---

[1] https://geintra-uah.org/index.php?q=en
[2] https://www.astander.es/
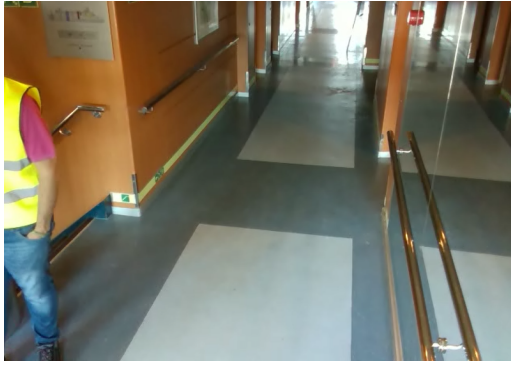
| RGB image | IR image |

Figure 1: Common area scenario shown from RGB (left) and IR (right) camera perspective

2. **Cabins:** consists of a long corridor that has two main exits, the beginning and the end of the corridor. It is important to highlight that the middle section of the corridor has different lighting that the rest of it, as it can be seen in Figure 2, that shows the scenario.
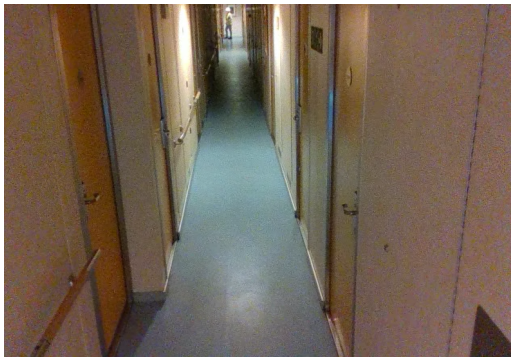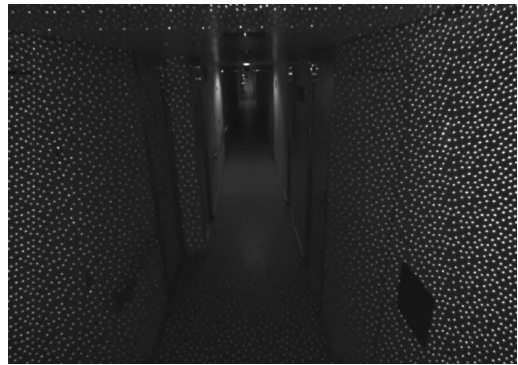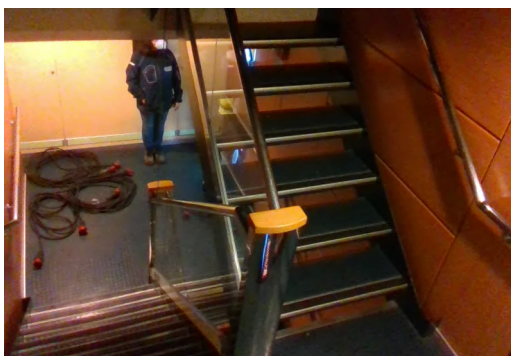

| RGB image | IR image |

Figure 2: Cabins corridor scenario shown from RGB (left) and IR (right) camera perspective

3. **Exit:** consists in a set of two stairs. The viewpoint of the camera is located at the middle of both and it has two main exits: a door located in the lower floor and the end of the stairs located in the upper floor. Figure 3 shows the scenario.
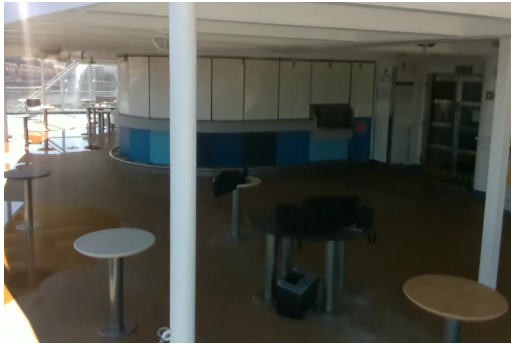

| RGB image | IR image |

Figure 3: Exit stairs scenario shown from RGB (left) and IR (right) camera perspective

4. **Exterior:** consists in an open area located in the surface of the cruise ship. It contains some tables and columns and it has three main exits, the door, the left side of the scenario and the right side. Figure 4 shows a picture of this scenario.

RGB image   IR image

Figure 4: Exterior scenario shown from RGB (left) and IR (right) camera perspective

5. **Lifeboat** consists of a section of the cruise ship that contains the lifeboats. This scenario has two exits: behind the camera and at the end of the hallway. Figure 5 shows the scenario.

RGB image   IR image

Figure 5: Lifeboat scenario shown from RGB (left) and IR (right) camera perspective

6. **Stairs** consists of a staircase, the camera point of view allows to see the lower floor, this scenario has two exits, the lower floor and behind the camera. Figure 6 shows the scenario.

RGB image   IR image

Figure 6: Stair corridors scenario shown from RGB (left) and IR (right) camera perspective

3 Dataset description

The dataset is composed by 32 videos with the participation of 17 people (users): three of them women and the rest men. These videos include people performing different actions, and in some of them, there are several people at the same time to have a realistic scenario.

The dataset is balanced taking into account that it contains the same number of videos recorded in RGB and in IR, (as the camera provides two outputs of the same video) and the camera view point of both is the same, with the exception than the IR camera has a larger FOV than the RGB one. This causes that the video obtained from the IR camera contains more information as it is capable of providing a wider view from x and y axis.

The actions performed along the dataset can be included in two groups:

1. Individual actions performed by each of the individuals independently from each other that include: walking, running, sitting, falling down and going through stairs.

2. Group actions performed by all the individuals or the majority of them in the video, including: stampede and blockade

Most of the videos included in the dataset contain three to four persons performing different actions in the scenario at the same time. This group of four people includes two men and two women with different heights. This selection of persons try to cover the diversity of the people that can be found inside a cruise ship in a real-world scenario.

It is important to highlight that the dataset is not balanced as different scenarios contain a different number of videos, different number of persons performing actions and different number of frames per scenario. Table 2 includes a summary of the distribution of videos of the dataset by scenario with the number of persons and the number of frames.

Scenario	Number of videos	Average number of persons	Number of frames
Indoor	10	4	8628x2
Cabins	5	4	6874x2
Exit	5	6	2739x2
Exterior	6	4	7413x2
Lifeboat	5	4	2336x2
Stairs	6	4	5854x2

Table 2: Video, persons and frames distribution of the HARS dataset

As it has been explained, the recorded frames, correspond to different people performing different actions, sometimes, simultaneously. Besides, table 3 includes the number of frames corresponding for an action per scenario included in the dataset.

Scenario	Walk	Run	Fall Down	Stationary	Stairs	Blockade	Stampede
Indoor	7604	4641	0	5735	0	0	1006
Cabins	5915	7076	1500	3676	0	1091	2349
Exit	1941	100	0	443	2338	0	0
Exterior	4237	3029	295	4120	462	332	0
Lifeboat	760	652	0	2458	462	0	0
Stairs	4079	0	1010	368	7440	986	0
Total	24536x2	15498x2	2805x2	16797x2	10240x2	2409x2	3355x2

Table 3: Video, persons and frames distribution of the HARS dataset

4 File formats and extensions

The dataset is composed of 6 folders named: common-area (Indoor), corridor-cabins (Cabins), exit-stairs (Exit), lifeboat, Outdoor, Stairs-corridors. In each folder there are videos with the same name of the folder but with a enumeration and one letter. The letter can be ?I? of infrared or ?c? of colour. The format of the videos are .mp4. The groundtruth has the same name of the video but the files with this information are allocated in the groundtruth folder. In addition this groundtruth files have the extension .gt.

For every video clip there is a file named videoXXX.gt that is the main annotation file, the Ground Truth. The term Ground Truth (GT) refers to the process of collecting specific data. That is, when creating a dataset and working with it, some files resulting from the annotation of the dataset where the information is collected are generated that wants to be studied, as for example the position of the person and the action developed by it. In this dataset two specific features are studied: firstly the position of each person in the video at each moment to know the tracking of all the people, and secondly the action that each person is doing in each frame. The GT file includes a line for each labelled frame containing the following data:

NIm	M/A	UserID0	X01	Y01	X02	Y02	Act0	Blockade	Stampede

being:

- nIm: the number of the frame in the video

- M/A: variable that indicates if the annotation is manual (1), automatic by a neural network called YOLO (-1), or automatic by the linear interpolation with an annotation tool (0)

And for each person labeled in the frame, there are six values:

- UserID: the ID of the user: all the users have an ID that starts to count in 0 an it increments by the way in which they appear in the different videos. Also, the intruders have an ID that starts with the number 100 in each video and incrementing the counter in the video.

- Xxx Yyy: four values that are the coordinates of the upper left corner and lower right corner of the bounding box where each person is.

- Actx: the number associated with the action that is performing each person

Finally, *Blockade* and *Stampede* are boolean values. In the case of blockade, "0" means no blockade and "1" the other wise. With Stampede field is the same.

5 Disclaimer, Licensing, Request and Contributions

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

The GEINTRA Multiple Actions Dataset in Cruises (`GSMADC`) Database (and accompanying files and documentation) by Carlos Cob-Parro, Cristina Losada-Gutiérrez, Marta Marrón-Romera, Rodrigo Antonio Vidal Pinto, Ignacio Bravo, Alfredo Gardel is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License ⊚①⊜.

To request a copy of the dataset, please contact Marta Marrón Romera at marta.marron@uah.es.