GBA Video Dataset for Human Action Recognition

Javier Macías-Guarasa, Cristina Losada-Gutierrez, David Fuentes-Jiménez, Carlos A. Luna, Manuel Mazo, Carlos Martínez-García, Marta Marrón-Romera, Carlos Cob-Parro, Valeria Boggian.

The following document is a resume of different characteristics from the dataset created by the research group GEINTRA, at the University of Alcalá. The dataset is composed of images sequences recorded indoor with people performing different actions.

GBA (GEINTRA Behaviour Analysis) dataset is a set of videos recorded at the Polytechnic School of the University of Alcalá, acquired using a stationary Full HD 2.7K camera, the GoPro HERO4, an action camera, compact, lightweight and resistant with a resolution of 1920x1080 pixels at 50 frames per second (fps) framerate. This camera has an ultra-wide-angle field of view and reduced distortion. The store format is .MP4. The camera has been located in the south building of the Polytechnic School (Figure 1), with northwest orientation at a height of approximately 1.8m, to increase the area of view reduce possible occlusions.

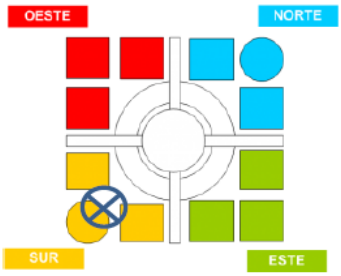

Figure 1 Top schematic view from the UAH's Polytechnic School

The chosen scenario includes a corridor with columns, stairs, an elevator, vending machines and a bench, as shown in Figure 2 and Figure 3. It is a realistic scenario including natural lighting that changes during the day.

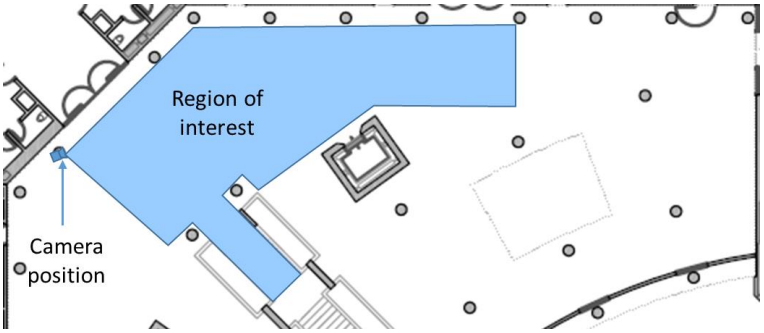

Figure 2 Region of interest. Top view

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. It makes GBA a challenging dataset for both people detection and tracking and human action recognition in the wild.

The aim of this dataset is to select a collection of daily behaviours, such as walk or sit down, in order to have sequences of images representing real situations which can occurs in people day-to-day. Besides, users perform these actions spontaneously, considering in every moment that not all people walk with same manners, and there are possibilities to have intruders at the scene in order to have a more realistic observation of the situation.

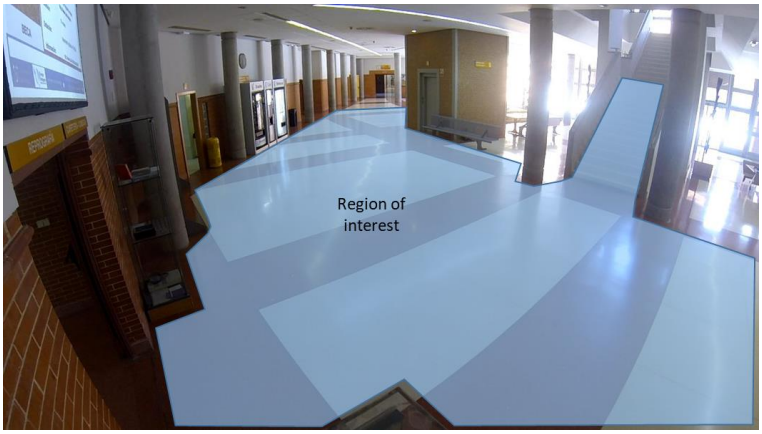

Figure 3 Region of interest. Camera view

The recorded videos include six actions: walking, running, sitting down, falling and going up or down stairs, being the number of sequences not balanced between actions. Table 1 shows a summary the number of people performing an action and the number of available videos.

Table 1 People and videos for each the actions

Action	People	Videos
Walk	17	32
Run	13	13
Sit down	16	17
Fall	14	15
Stairs	16	19

Each video is split into pieces, including a person performing one action that can be used for training.

Table 2 shows the actions included in each recorded videos and the number of segments for each action, as well as the total number and percentage of segments for each action.

Table 2. Summary of the included actions and the number of segments for each action in the recorded videos.

		Number of sequences				
Video	**Performed actions**	**Walk**	**Run**	**Sit down**	**Fall**	**Stairs**
0000	walk, fall	7	0	0	6	0
0001	walk, sit down, stairs, run	59	19	7	0	11
0002	walk, fall	8	0	0	4	0
0003	walk, sit down, stairs, run	52	15	7	0	6
0004	walk, fall	8	0	0	4	0
0005	walk, sit down, stairs, run	68	16	11	0	14
0006	walk, fall	10	0	0	4	0
0007	walk, sit down, stairs, run	50	16	6	0	7
0008	walk, fall	6	0	0	5	0
0009	walk, sit down, stairs, run	48	11	7	0	9
0011	walk, fall, stairs	7	0	5	0	5
0012	walk, sit down, stairs	79	0	11	0	3
0013	walk, sit down, stairs	45	0	6	0	4
0014	walk, fall	7	0	0	4	0
0015	walk, sit down, stairs, run	59	10	21	0	20
0016	walk, fall, stairs	10	0	0	3	19
0017	walk, sit down, stairs, run	38	13	7	0	3
0018	walk, fall	4	0	0	4	0
0019	walk, sit down, stairs, run	66	14	8	0	19
0020	walk, fall	10	0	0	6	0
0021	walk, sit down, stairs, run	67	14	7	0	14
0022	walk, fall, stairs	8	0	0	6	5
0023	walk, sit down, stairs	51	0	6	0	4
0024	walk, sit down, stairs, run	61	11	10	0	6
0025	walk, fall	6	0	0	3	0
0026	walk, sit down, stairs, run	47	14	6	0	7
0027	walk, fall	6	0	0	0	5
0028	walk, sit down, stairs, run	53	0	6	0	4
0029	walk, fall	6	0	0	5	0
0030	walk, fall	8	0	0	5	0
0031	walk, sit down, stairs, run	55	31	13	0	9
Total		**1009**	**184**	**144**	**59**	**174**
Percentages		**64.27%**	**11.72%**	**9.17%**	**3.76%**	**11.08%**

GBA Video Dataset for Human Action Recognition

The dataset is composed of 32 folders named "XXXX", being XXXX the number of the video. In each file, there are four components:

- Video recorded in format .MP4 called *GOPR0XXX.MP4* being XXX the number of the video.
- The ground truth associated, called *videoXXX.gt*. The ground truth is the document where specific data are collected to be studied, such as the position of the people.
- A file with the all the frames in .jpg format that composed the video. This file is called *frames_GROPR0XXX*. All the frames are called with the numeration corresponded with 4 digits. This means that the first frame is going to be the archive 0001.jpg and the frame 205 is going to be the archive 0205.jpg.
- A file called *xml_GROPR0XXX* with the ground truth in XML format. This file is composed of as many archives as labelled frames the video has. The name of each file is the name of the frame i.e. 0240.xml
- A file called *pieces_GOPR0XXX-fix* with all the pieces that composed the video. Each piece is what it is called sequence, a one second video in .avi format with a person doing one action. Each video is named with the ID of the person with 4 characters, the action being done and a counter, for example, 00170301.avi:

This means that is the video number 1 of the person with ID 17 doing the action 3. It has to be said / that the counter starts in 0.

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	$UserID_0$	X_{01}	Y_{01}	X_{02}	Y_{02}	Act_0	$UserID_1$	X_{11}	Y_{11}	X_{12}	Y_{12}	Act_1	...

- 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)

For each person labelled 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.

- X_{xx} Y_{yy}: four values that are the coordinates of the upper left corner and lower right corner of the bounding box where each person is.
- Act_x: the number associated with the action that is performing each person

This information is also generated in xml format, in files named xml_GROPR0XXX. In this file there is an archive for each frame labelled called like the associated frame. This means that if the frame is called 0001.jpg, the xml archive from the ground truth of this image is called 0001.xml as well.

In those archives the path where the frame folder is located, as well as the frame that correspond with the ground truth is stored. Also is stored the size of the image (1920x1080x3), the name of the person and the coordinates of the bounding box.

If you want to get a copy of the datasets, please contact marta.marron@uah.es

If you make use of this datasets and/or its related documentation, you are kindly requested to cite the paper:

- M. Baptista-Ríos, C. Martínez-García, C. Losada-Gutiérrez and M. Marrón-Romera, "Human activity monitoring for falling detection. A realistic framework," 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 2016, pp. 1-7, doi: 10.1109/IPIN.2016.7743617.