

Oct 15, 2020

## Panasonic and Stanford Vision and Learning Lab Release "Home Action Genome", the World's Largest<sup>1</sup> Multimodal Dataset for Living Space AI Development

Panasonic Corporation and Stanford Vision and Learning Lab (SVL) in the US have compiled the world's largest<sup>1</sup> multimodal datasets<sup>2</sup> for living space AI development, called Home Action Genome, and make it available to researchers. In addition, the parties host an International Challenge on Compositional and Multimodal Perception (CAMP), a competition for action recognition algorithm development using this dataset.

Home Action Genome is an image and measurement dataset compiled from multiple sensor data, including data from cameras and heat sensors, in residential scenes where daily life actions are mimicked. The data includes annotation<sup>3</sup> that characterizes human action content for each scene.

Information concerning data acquisition and CAMP participation methods is available at the CAMP website.  
<https://camp-workshop.stanford.edu/>

Most living space datasets released to date have been small and composed largely of audio and image data. The new dataset combines Panasonic's data measurement technology with annotation expertise from SVL to achieve the world's largest multimodal datasets in living space.

AI researchers will be able to apply this dataset to machine learning, and utilize it for research into AI to support people in the home.

To realize individualized Lifestyle Updates that make living better day by day, Panasonic will accelerate AI development for home by promoting collaborative use of the dataset.

### Notes

1. As of October 15, 2020. As a multimodal dataset for living space (Panasonic survey)
2. Data generated from simultaneous readings from multiple sensor types
3. Human-supplied semantic information relating to the data

### <Home Action Genome dataset details>

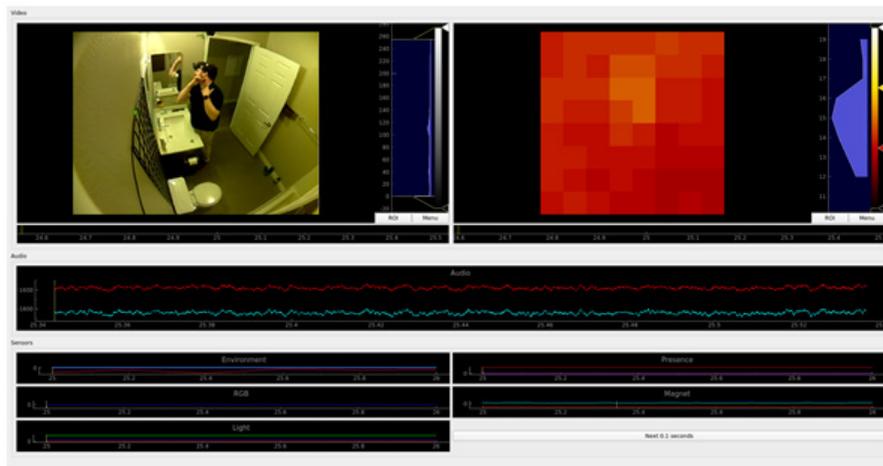
#### • Scale of data

The dataset is composed of 3,500 scenes with multiple location and human subject data in 70 action categories. Each scene consists of sequences with duration of approximately two to five minutes.

• Sensor data

Data type	Detail
Video	Camera image data
IR	Graphic representation of IR sensor grid-based heat data for humans and objects
Audio	Microphone audio data
RGB Light	Intensity for red, green, and blue visible light
Light	Interior light intensity data
Acceleration	Angular speed and acceleration data from gyro sensors and accelerometers
Presence	Human presence data from IR sensors
Magnet	Geomagnetic sensor data

• Graphic representation of sensor-measured data (example)



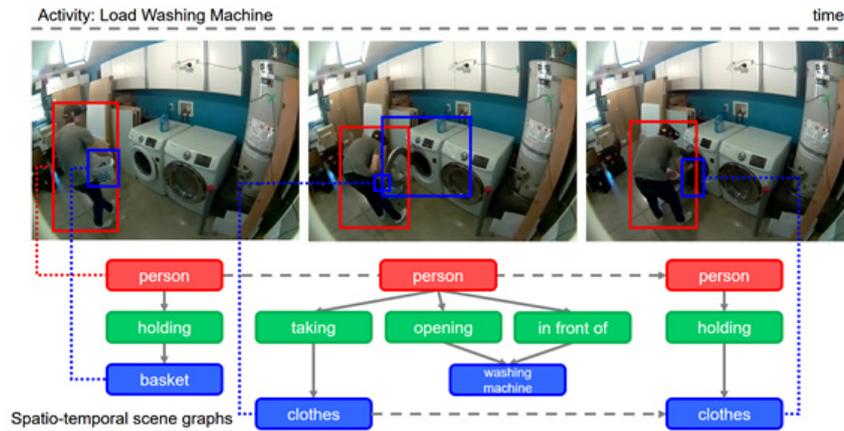
Example of "Shaving" scene image data.  
 Graphs represent each sensor's data in a time series

## ·Annotation information

The multimodal dataset includes the following information.

- Video data: Human and object locations
- Scene data: In-scene human action categories

Please visit the CAMP website for further information.



Diagrammatic of annotation example

*\*The content in the following news releases is accurate at the time of publication but may be subject to change without notice. Please note therefore that these documents may not always contain the most up-to-date information.*