Thermset: A thermal-videos dataset of elder people bedrooms

Pablo Pusiol
pdp0109@famaf.unc.edu.ar

Federico Polacov
fdp0108@famaf.unc.edu.ar

Universidad Nacional de Córdoba - FaMAF
October 15, 2016

1 Introduction

Our interest is to automatically detect clinically relevant daily activities of seniors living independently or in nursing homes. To do that we will use thermal information and develop innovative computer vision technology. Modern computer vision is based on deep learning techniques and has proven to be very effective at automatizing visual recognition tasks such as detecting objects in images, tracking and detecting activities in videos, etc. A caveat of these techniques is the need of big amounts of labeled data for training their models. In practice, to achieve models that could reliably work in new environments (e.g. homes, people, etc.) would require of the training dataset to contain activity examples covering the full space of the manifold defined by all possible configurations of the target activities and environments.

While video cameras are traditionally used for daily activity monitoring, they need additional algorithms to overcome their inherent vulnerability to low light conditions, too much light, shadows, complex scenes. Far infrared sensors (thermal sensors) do not have any these issues because they create a crisp image based temperature data from a scene, so almost all the problems traditional RGB cameras face in processing and classifying images are avoided [1].

In this document, we introduce a new dataset: Thermset, an incremental dataset of thermal information recorded for different time intervals of (so far) 13actualizar senior citizens at their bedrooms with automatic annotations as described in subsection 2.4. Each of the instances of the dataset (videos) corresponds to a series of fram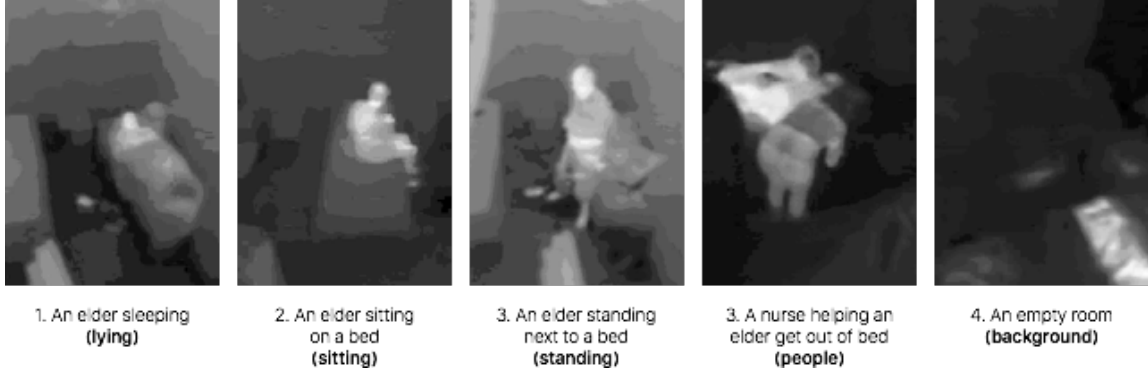es where each frame corresponds to the matrix of absolute temperature values captured by a far-infrared sensor. The sensor captures for intervals of as little as 2 seconds or as long as 15 seconds. Until the date of writing this document, Thermset has about 767actualizar hours of recordings, where typical examples of data reflect an elder sleeping, wandering, getting ready to bed, receiving assistance from a caregiver. The dataset also contains thermal data of the different bedrooms with no people as shown in Figure 1 Thermset is available through `http://forelderly.weebly.com/data`.

In section 2 we provide technical details of how Thermset is structured, characteristics of data, annotations and how data is being gathered. The rest of the document is structured as follows: In section 3 we provide qualitative information of the seniors recorded and environments and finally in section 4 we provide an example of how impactful Thermset can be to improve seniors wellbeing.

2 Technical details

2.1 Structure

Thermset is structured in groups. A group is a sequence of chunks with the same duration. A chunk is represented with a sequence of thermal values matrixes expressed in Celsius degrees (from now on we will call each thermal matrix a frame). Each group has chunks of a unique person recorded for a certain time interval with a certain configuration of our recording tools as we will describe in this section. Is important to notice that there is a time gap between each chunk of a group. Its duration depends on the recording configuration used and is explained in paragraph 2.3.1.

1. An elder sleeping (lying)   2. An elder sitting on a bed (sitting)   3. An elder standing next to a bed (standing)   3. A nurse helping an elder get out of bed (people)   4. An empty room (background)

Figure 1: *Examples of thermal data extracted from Thermset. (Images where stretched for easy reading)*

Group names follow the structure **Name_YYY**, where Name refers to an elder's alias and **YYY** to the **YYY**-th group recorded of that elder. For instance, the group **Marge_001** is the 1st group with recordings of an elder whose alias is Marge.

2.2 Diversity of data

Below we provide graphics to help understand better the distribution of the data collected. The majority of recordings were collected in nursing homes, but there are two groups recorded at two private addresses.

grafico de porcentajes hombre/mujer

grafico de edades

grafico de largos de videos

2.3 Recording tools

To record and create annotations we use Apple iOS devices and FLIR One sensors. The app used was Thermix[5] , available for free in the App Store.

We relied on 3 different devices to run Thermix: an iPod Touch 5th generation, an iPhone 5 and an iPhone 5s. The far-infrared sensors used were Dongle and SLED versions of the FLIR One. We configure Thermix in the iPhone 5 with an SLED sensor and in the iPhone 5s and the iPod Touch with a dongle sensor[4][3]. Because Thermix excludes all frames during a sensor recalibration, we estimate the average fps (frame per second) for each configuration achieving 5fps, 3fps and 2fps respectively. CONTROLAR SI SE ENTIENDE.

2.3.1 Recording Configurations

Groups in Thermset were recorded in different ways. Each of these are what we call *Recording Configuration or RC*. There is a total of 7 Recording Configurations, labeled *RC1* to *RC8*. Information of the configuration and device/sensor used for each group is available at http://forelderly.weebly.com/data. Below we describe 3 aspects of each configuration, and at Table 1 we describe each Recording Configuration by this 3 aspects.

Precision of thermal data For $R = raw$ *thermal data returned from the sensor* we construct a frame F, where $F(i, j)$ corresponds to the element i, j of the frame adjusted by precision.

- **Low precision:**

$$F(i,j) = \begin{cases} 2 * R(i,j), & \text{if } 1 \leq R(i,j) \leq 120 \\ 0, & \text{Otherwise} \end{cases}$$

  Where $R(i, j)$ is the temperature expressed in Celsius degrees with no decimal places. *RC1, RC3* correspond to this precision.

- **High precision:**

$$F(i,j) = \begin{cases} round(2 * R(i,j)), & \text{if } 1 \leq R(i,j) \leq 120 \\ 0, & \text{Otherwise} \end{cases}$$

  Where $R(i, j)$ is the temperature expressed in Celsius degrees with one decimal place. The *round* function returns the integral value nearest to its argument, rounding half-way cases away from zero.

Video lenght Each group records videos for a certain number of time that ranges between 2 and 15 seconds. The amount of frames per video is related to the fps the device is capable of and the time each device was recording for:

$$number\_of\_frames = fps(device) * video\_lenght(RC) \quad (1)$$

Dead intervals Thermix records and uploads small chunks of video to the cloud. There are two possible configurations: either Thermix saves chunks to the device disk continuously (continuous mode) or waits until a chunk is sent to start recording the next one (one-by-one mode).

- **Continuous mode:**
  Video chunks in groups recorded with continuous mode are separated by a small time window (between 2 and 5 seconds in average), that correspond to the time Thermix takes to compose the chunk[1]. This means, Thermix did not wait for a video chunk to be sent before start recording the next one.

- **One-by-one mode:**
  Videos in groups recorded with one-by-one mode are separated by the time Thermix takes to compose the video chunk as in continuos mode plus the time the device takes to upload it to the cloud.

← revisar no me gusta

2.4 Annotations

Thermset provides automatic annotations generated by Thermix for each video. The annotation corresponds to the output of the classification of the first frame of the video using the method described at section 4. Classes are background, standing, sitting, lying and people. Annotations for each group are listed on file (named **GROUP_NAME**_annotations.txt) stored in each group folder. Annotations follow the structure:

---

[1]The composition of the chunk consists in generating a video from thermal data and annotating the first thermal frame using the method described in section 4

Figure 2: *Frame with annotation:*
2016-09-06_03.53.33   *1:0.998713*
4:0.000638   2:0.000534   5:0.000074
3:0.000040

$video\_url \{<class\_id:score>\}^5$,

where *class_id* corresponds to classes as detailed in Table 2, and *score* correspond to the output of the classification each *class_id*. An example of an annotation is shown in Figure 2.

3 Qualitative description of the dataset

Recording environments Videos where recorded at several bedrooms in 2 different nursing home facilities and at 2 private homes. All videos where recorded during winter.

About the people Until the date of writing this document, Thermset has data of the following elders:

- **Marge:** 90+ years old, Female
  Diseases or disabilities: Hyperthyroidism. Had fractures due to falls
  Situation: Living at a healthcare facility. Shares the bedroom with another person. Takes naps daily. Sleeps with a diaper. Had a hip fracture after a fall. Moves around with walker and shes afraid of walking alone. Requires constant assistance for moving, entering, leaving bed, etc.

RC	Video lenght	Precision of thermal data	Sampling dead intervals
1	10 seconds	Low precision	One-by-one
2	15 seconds	Low precision	Continuous
3	15 seconds	Low precision	Continuous
4	10 seconds	High precision	One-by-one
5	2 seconds	High precision	One-by-one
6	2 seconds	High precision	One-by-one
7	5 seconds	High precision	One-by-one
8	5 seconds	High precision	One-by-one

**Table 1:** *Details of recording configurations comprehended by Themset*

id	class
1	lying
2	sitting
3	standing
4	group
5	background

Table 2: *Classes and their corresponding ids.*

- **Peter:** 80+ years old, Male
  Diseases or disabilities: Had fractures due to falls
  Situation: Lives at his place, alone. He does not require assistance, but due to a knee fracture has difficulties to move. Uses a walking stick. Takes naps and reads the newspaper while in bed.

- **Anne:** 60+ years old, Female
  Diseases or disabilities: Alzheimers
  Situation: Living at a healthcare facility. Shares the bedroom with another person. Wanders during the day.

- **Henry:** 55+ years old, Male
  Diseases or disabilities: Vascular dementia
  Situation: Living at a healthcare facility. Shares the bedroom with Victor. Wanders frequently during the day.

- **Victor:** 80+ years old, Male
  Diseases or disabilities: Visual impairment
  Situation: Living at a healthcare facility. Shares the bedroom with Henry. Takes frequent naps during daytime. Due to his visual impairment requires frequent assistance but he uses a snub to urinate next to his bed without assistance. Urinates frequently.

- **Fiona:** 75+ years old, Female
  Diseases or disabilities: Overweight, Hyperthyroidism
  Situation: Lives at her place, alone. She does not require assistance. Takes naps and watches TV while in bed.

- **Irene:** 75+ years old, Female
  Diseases or disabilities: Advanced Alzheimer's disease
  Situation: Living at a healthcare facility. Shares the bedroom with Anne. Due to her incapacity of moving by herself, requires a wheelchair. Sleeps in a special bed to prevent falls.

- **Charles:** 70+ years old, Male
  Diseases or disabilities: Schizophrenia, Alcoholism
  Situation: Living at a healthcare facility. Shares the bedroom with two other male elders. Wanders during the day but has no major mobility issues.

- **Rick:** 75+ years old, Male
  Diseases or disabilities: Visual impairment, Vascular dementia, Limited mobility.
  Situation: Living at a healthcare facility. Shares bedroom with Henry and Victor. Requires assistance to move but uses a snub to urinate next to his bed without assistance

- **Julien:** 90+ years old, Male
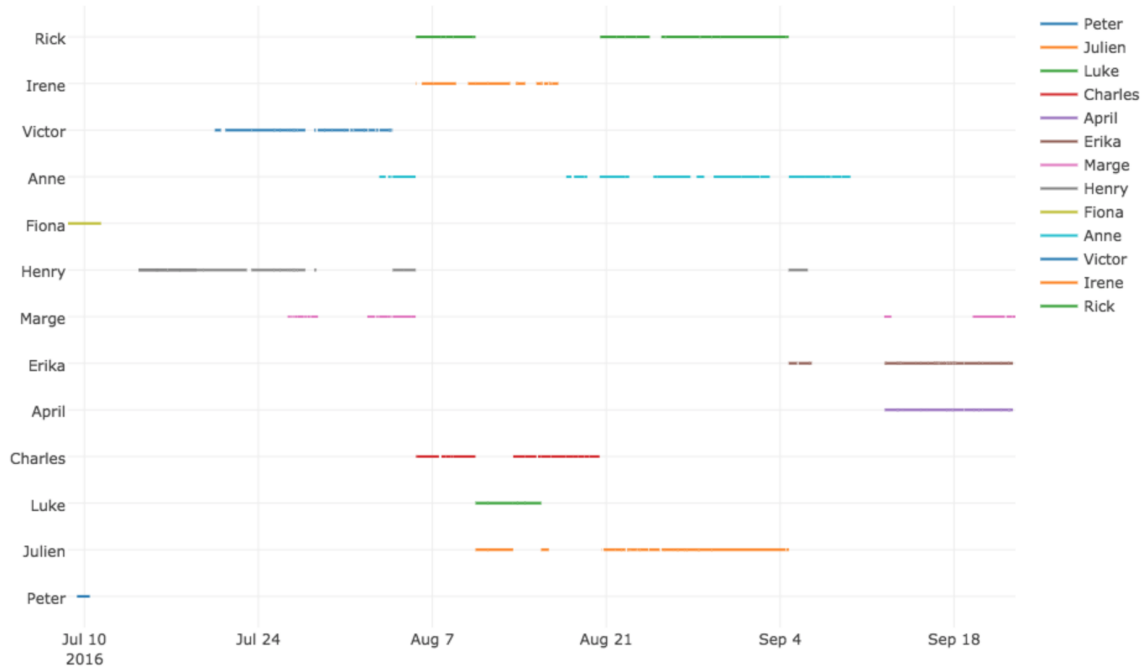
Figure 3: *Distribution in time of recordings per elder. Blank spaces represent a group change or intervals where the system was not recording for more than 5 minutes*

Diseases or disabilities: Alzheimer's disease, mild arthritis

Situation: Living at a healthcare facility. Shares the bedroom with Charles and Luke. Does not require assistance to move. Does hydrotherapy to combat arthritis.

- **Luke:** 80+ years old, Male
  Diseases or disabilities: Alzheimer's disease
  Situation: Living at a healthcare facility. Shares the bedroom with Charles and Julien. Does not require assistance to move.

- **Erika:** 70+ years old, Female
  Diseases or disabilities: Senile dementia
  Situation: Living at a healthcare facility. Shares the bedroom with 2 other women. Requires full-time assistance.

- **April:** 70+ years old, Female
  Diseases or disabilities: Senile dementia
  Situation: Living at a healthcare facility.

Shares the bedroom with 2 other women. Due muscular spasms product of her advanced dementia requires full-time assistance.

4 Application: Pose estimation

In the context of daily activities detection, collecting a training set large enough to accurately detect clinically relevant activities is hard. It might require of hundreds of data feeds for several days, or years if we consider the environmental temperature fluctuations and the different appearances (e.g. clothes) that a person will have during the stations. In addition, it would require of several human annotators working 24/7 watching and extracting the target activities appearing in the streams. Likely, the number of labelers would be proportional to the number of data streams that need to be analyzed. In addition to the complexity of collecting a big labeled training dataset, the size

of each labeled example could present a problem as well. Optimizing an utility function over long-term visual streams would require a massive amount of computing power and memory.

To overcome the limitations described above, instead of thinking in rigid models to detect all activities at once it is more convenient to design dynamic systems capable of learning incrementally new activities when new labeled data arrives. Daily living activities are hierarchically composed of sub-activities occurring in a variable period of time. In general, the deeper we move in the taxonomy of activities, the less training-data is needed for building reliable models to detect them. This observation is related to the lower dimensionality of their manifold in comparison with more complex activities. We are interested in detecting complex activities in incremental bottom-up steps. First, by detecting anchoring activities lying at the lower end of the human activities taxonomy. These anchor activities are persistent human features appearing as sub-activities of more complex ones. Enabling a robust detection of these anchors will enable our second stage of learning, which will use the anchors for detecting and analyzing complex long-term human activity. For the rest of this section we will focus in the first stage of our algorithm: defining and modeling the detection of anchoring activities.

Pose Anchors One of our anchor activities is the detection of pose. Human pose is a highly descriptive and persistent feature, capable of characterizing complex activities. For senior adults living independently, changes of pose in the particular contexts could be describing a high risk situation. For example, a person changes from "standing" to "lying in the floor" could indicate the occurrence of a fall. In nursing-homes, detecting and alerting the "coming out of bed" could enable promptly caregiver's assistance.

The detection of pose anchors is connected to the monitoring of clinically relevant activities. Here we borrow some of the target daily living activities described in [6], where several of the activities presented there can be inferred directly from the human pose. For example:

1. **Falls:** corresponds to transitions from standing to lying or from sitting to lying -as long as lying is not happening in the bed area

2. **Front-door-Loitering:** corresponds to detecting a person standing in selected locations

3. **Day-Night-Reversals and night wanders:** corresponds to detecting a person standing at unusual night hours

4. **Sleep:** corresponds to detecting a person lying in bed, even when the person is covered by a blanket

5. **Immobility (bed or chair):** corresponds to long periods of time where the person is either sitting or lying (in bed)

6. **Restlessness:** corresponds to detecting and tracking pose changes for long periods of time where the person is perpetually agitated or in motion.

4.1   Model

We built our model as a Fully Convolutional Network derived Matt Zeiler's model [7], which takes an input of 225x255x3 thermal image is convolved through 5 layers. The last layer is fully connected, taking features from the top convolutional layer as input in vector. The final layer is a 5-way softmax function, 5 being the number of classes: *lying*, *sitting*, *standing*, *group* and *background*. All filters and feature maps are square in shape.

4.2   Dataset

The dataset used to train this model was labelled manually and contains 30825 frames with 6165 frames by class, where 12726 were selected from Thermset videos and 18099 where collected from a controlled environment.

We randomly split 80%/20% of data into training and testing datasets. Since our training units are videos, we are careful of not including any frame of a testing video in the training set (and vice versa)
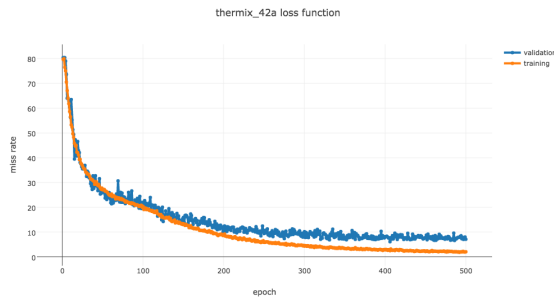
Figure 4: *Cross entropy loss for logits*

4.3 Evaluation

The dataset was split into training and validation datasets, 80%/20% respectively. We are careful of not including any frame of a testing video in the validation set (and vice versa). Training was done for 500 epochs with a Dropout rate of 0.5. Results were obtained running the algorithm in the validaation dataset with a generalization error of 7.09%. Our loss function is cross entropy for logits. (See Figure 4).

Thermix provides a publicly available account where this model is being used live for presence detection for 3 people living in a nursing home, this helps nursing home staff to monitor activity in those scenes. Access can be requested via `http://forelderly.weebly.com/data`.

5 Discussion and further work

5.1 Extending Thermset

Annotations Thermset provides automatic annotations created using the procedure explained at section 4. To train the model, we labelled 30825 frames. Creating different ground truths, such as localization of people would allow a better understanding of the scene.

Size Thermix provides a plug-and-play interface to easily gather thermal data. So far we collected 767actualizar hours of recordings of 13actualizar seniors. At the moment of writing

this document, Thermix is set up in three different locations and we have plans of augmenting the dataset with more seniors in different locations.

Exploiting Thermset Thermset provides hours of recordings of seniors being aided, sitting, wandering, people that felt to the ground, nurses, etc. Such data is highly valuable for training new algorithms aimed to help the elderly live better and benchmarking existing algorithms in real life situations.

References

[1] Hong Chengl, Zicheng Liu, Yang Zhaol, Guo Yel *Real world activity summary for senior home monitoring.* 2011.

[2] FLIR One SDK, `http://developer.flir.com/sdk-documentation/`, accessed August 2016.

[3] FLIR One SLED Tech Specs, `http://www.flir.com/flirone/press/FLIRONE_Fast_Facts_Tech_Specs.pdf`, accessed August 2016.

[4] FLIR One Dongle Tech Specs, `http://www.flirmedia.com/flir-instruments/industrial/datasheets/flir-one-datasheet.html`, accessed August 2016.

[5] Thermix for FLIR One, `https://appstore.com/thermixforflirone/`, accessed August 2016.

[6] Stanford PAC Seniorcare website, `http://vision.stanford.edu/pac/seniorcare/`, accessed August 2016.

[7] Matthew D. Zeiler and Rob Fergus *Visualizing and Understanding Convolutional Networks*, ECCV 2014