IAS0360 Final Project Proposal: Topic 2 Thermal Sensor Based Human Detection

Felix Lukas Kühner, Omar El Nahhas, Tobias Weiss
Department of Computer Systems
Tallinn University of Technology
Tallinn, Estonia

I. Introduction

This project tackles the concept of recognizing and classifying heat patterns in order to detect humans. The heat data is collected from a thermal sensor attached to the ceiling of a room. It measures any object which emits heat. For example, the object could be an oven, human being or any other heat-emitting object. In order to detect humans, a model shall be created to distinguish heat patterns. The hypothesis of this research is that humans can be distinguished from other heat emitting objects using a thermal sensor. As such, the project is an object detection task.

II. Data

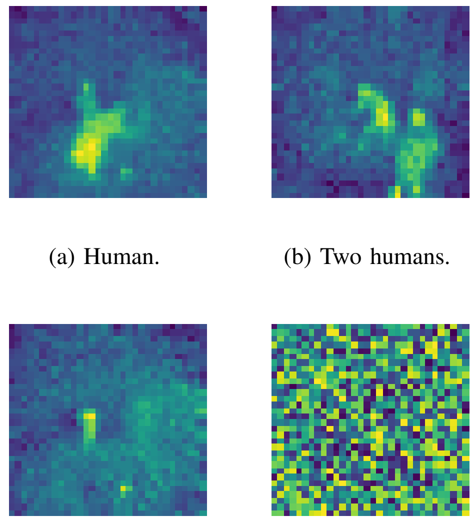
Only raw, unlabeled data is available for this project. It consists of 32×32 infrared images which should be the input for the model (see Figure 1). The images are indoor images with humans present. The dataset contains 10,000 of these images which were taken with a sampling rate of 10 samples per second. But since two different sensors were used to record the same scene, only 5,000 unique data points are in the dataset. In addition to the image array, each data point includes information about the thermal sensor, with which the data was recorded, the current room temperature, the strength of the signal and the validity of the data point.

Since the dataset is not labelled, there is uncertainty regarding the complexity of the data as well as the problem which needs to be tackled during the project. For example, it is not clear whether the data of the two sensors should be fused or should be seen as separate entities. Moreover, the number of classes of objects that are sufficiently well detectable remains uncertain as of the beginning of the project.

III. Preliminary Assumptions

Certain preliminary assumptions can be made about the scenes that were recorded in order to narrow down the scope of the problem:

- The human body temperature will not exceed 40.0 °C. Moreover, most of the body is covered with clothing, causing the measured temperatures to be even lower;



(a) Human. (b) Two humans.

(c) Non-human heat. (d) Random pixels.

Fig. 1: Different examples of thermal images.

- The human body heats up the environment and air around it, so the area of hot pixels is larger than the actual body size;
- There is a measurable cool-down time for objects which were warmed up by humans. The cool-down time depends on the material of the object as well as the starting temperature;
- Overlaps between humans and other heat emitting objects may appear.

IV. Description of the Methodology

In order to prove this research's hypothesis, a methodology is defined to conduct experiments with the presented data. Initially, the raw data will be converted into 32×32 images. Unsupervised learning and classic statistical methods will be applied in order to get an initial estimate of the complexity of the data. When the data is too complex to be solved in an unsupervised manner, the images will then be labelled according to a to-be-determined ground truth. Having labelled the images, supervised learning methods will be applied in order to create a model which is able to distinguish humans from heat emitting objects, therefore proving the hypothesis to be correct. The model validation metrics to be used are precision, recall, F1-score and accuracy for supervised learning, and entropy and inter-intra cluster ratio for unsupervised learning.

On an operational level, we apply the *Cross Industry Standard Process for Data Mining* (CRISP-DM) [1] framework. This helps us to structure our steps and implies the report structure. CRISP-DM comprises following steps: Business understanding, Data understanding, Data preparation, Modeling, Evaluation and Deployment.

A. Data augmentation

As part of the data preparation step, we augment previously labelled samples. To enrich the data set, we investigate not only transformation but investigate further methods, as described in [2]. This approach will lead to an increased data set size and may lead to better predictions [3].

B. (State-of-the-Art) ML models

As first model, we evaluate boosted trees (XGBoost) [4]. For this approach we construct features as object size, object average temperature, object temperature variation, object movement, etc.

The second model we evaluate is Convolutional Neural Networks (CNN) [5]. The CNN approach shall classify into three categories: (1) No heat-emitting objects, (2) non-human heat-emitting object(s) and (3) human(s). We might use certain CNN layout, e.g. SqueezeNet [6].

C. Parameter tuning

In order to optimize our results, we search for promising configuration sets for our models. As discussed during the lecture, a random search approach [7] is state-of-the-art. Previous tuning attempts with Tensorboard were not satisfying. In this project we utilize the Tune framework [8], which promises advanced parameter optimization opportunities and distributed search possibilities.

V. Expected outcomes and constraints

The aimed-at accuracy of the proposed model is more than 85% on the given validation set, considering the following constraints:

- Model must be convertible to/executable in C language;
- Model must fit into 1-core ARM MCU (STM32F4, nRF52840 or nRF9160);
- Model should run with at least 1 FPS on selected MCU.

VI. Task assignment

Each team member gets assigned one model to train:

- Felix - Supervised learning: Gradient Boosting
- Omar - Unsupervised learning: Expectation-Maximisation, Density-Based Clustering, K-Means, or other to-be-determined clustering methods
- Tobias - Supervised learning: Convolutional Neural Network

Afterwards, we will then compare the results and select the best approach.

References

[1] R. Wirth and J. Hipp, "Crisp-dm: Towards a standard process model for data mining," in *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, vol. 1. Springer-Verlag London, UK, 2000.

[2] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019.

[3] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," 2017.

[4] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[5] Y. LeCun, Y. Bengio *et al.*, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.

[6] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <0.5mb model size," *arXiv:1602.07360*, 2016.

[7] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization." *Journal of machine learning research*, vol. 13, no. 2, 2012.

[8] R. Liaw, E. Liang, R. Nishihara, P. Moritz, J. E. Gonzalez, and I. Stoica, "Tune: A research platform for distributed model selection and training," *arXiv preprint arXiv:1807.05118*, 2018.