Applying Faster R-CNN in Extremely Low-Resolution Thermal Images for People Detection

Diego M. Jiménez-Bravo
Expert Systems and Applications Lab, Faculty of Science,
University of Salamanca
Plaza de los Caídos s/n, 37002 Salamanca, Spain
dmjimenez@usal.es

Pierre Masala Mutombo
IDLab research group,
University of Antwerp - imec,
Sint-Pietersvliet 7, 2020 Antwerp, Belgium
PierreMutombo.Masala@uantwerpen.be

Bart Braem
IDLab research group,
University of Antwerp - imec,
Sint-Pietersvliet 7, 2020 Antwerp, Belgium
Bart.Braem@uantwerpen.be

Johann M. Marquez-Barja
IDLab research group,
University of Antwerp - imec,
Sint-Pietersvliet 7, 2020 Antwerp, Belgium
Johann.Marquez-Barja@uantwerpen.be

Abstract—In today's cities, it is increasingly normal to see different systems based on Artificial Intelligence (AI) that help citizens and government institutions in their daily lives. This is possible thanks to the Internet of Things (IoT). In this paper we present a solution using low-resolution thermal sensors in combination of deep learning to detect people in the images generated by those sensors. To verify whether the deep learning techniques are appropriate for this type of images of such low resolution, we have implement a Faster Region-Convolutional Neural Network. The results obtained are hopeful and undoubtedly encourage to continue improving this research line. With a perception of 72.85% and given the complexity of the problem presented we consider the results obtained to be highly satisfactory and it encourages us to continue improving the work presented in this article.

Index Terms—Convolutional Neural Network, Faster Region-Convolutional Neural Network, Grid Eye, Internet of Things, Low-Resolution Images, People Detection, Thermal Images

I. INTRODUCTION

In recent decades the world population is increasing more than ever, a 20% in the last decade [1]. This population usually share the same space in streets, squares, parks, buildings, factories, etc. Therefore, it is important to know how these people act when they are in a shared space. This information is important from the perspective of different research fields like urbanism, psychology, sociology, safety, engineering, computer vision among others.

Engineering and computer vision fields are the ones that provide automated solutions to detect, track and/or count humans in big sharing spaces. There are several technologies proposed for these tasks. Mainly sensors that provide to us the data that a computer system needs to determine the position of a human, or to determinate and count if there are people in the sensing area of the sensor.

However, many of these solutions [2]–[4] do not provide a total solution (detect, count, and track tasks). Furthermore, the technologies that allow us to detect, count and track different people provide a privacy-invasive solution. Therefore, the work we present in this paper, aims to verify if a low-resolution thermal sensor can provide a privacy-preserving detecting, counting, and tracking solution for human detection in Internet of Things (IoT) devices. To do this, we have used a thermal sensor, the Panasonic Grid Eye [5], and Convolutional Neural Networks (CNN) that will detect objects (in this study people) within the thermal images.

II. BACKGROUND

Counting the number of people in a room or an open space has been a difficult task that several researchers have tried to solve. Currently, several technologies offer a solution to this problem; Density [2] and non-commercial solutions like as an example the use of optical sensors [3], PIR sensors [4], WiFi tracking [6], presence sensor [7], ultrasonic sensors [8] and thermal sensors [9]. However, many of these technologies provide an inaccurate and non-privacy solution violating people's privacy rights.

Several studies make use of these kinds of sensors to determine the number of people mainly in an indoor environment. An example of this is the research developed by Trofimova et al. in Italy [10]. They used the Grid Eye sensor [5] which provides an 8 x 8 matrix with the temperatures of the room.

A similar approach is developed by Beltran et al. [11] the authors of the Grid Eye sensor and also by Jeong et al. [12]. However, the Grid Eye sensor is not the only

European Social Fund, Junta de Castilla y León and European H2020 framework programme.

*978 – 1 – 7281 – 7343 – 6/20/$31.00 ©2020 IEEE

sensor used for human detection in low-resolution thermal images. Tyndall et al. [13] used the MAL90620 4 x 16 thermal sensor.

From another point of view, there are numerous studies that deal with the problem of detection of people, although in images of a greater resolution and in which in general the forms of the people are much more noticeable. A first example is the study carried out by Chen et al. [14] but there are many others [15], [16].

Therefore, in this article, we propose a new approach to obtain and to locate the people inside a room by using a low-resolution thermal sensor. In this study, we used the Grid Eye sensor to obtain the data that is processed to eliminate the background, and to obtain the new heat sources that are present in the field of view of the sensor. These features vectors are used to generate images that are classified by a Faster R-CNN (Region-Convolutional Neural Network) [17] model previously trained. The output of the model provides labeled images in which we can obtain the number of people who are in the image.

III. PROPOSAL

In this study, three Grid Eye sensors are used to increase the field of view. They are mounted on the Octa-Connect Platform develop by the imec [18]. To detect groups of people a Faster R-CNN [17] is used. This kind of network will detect possible person and/or groups of people in the images. It is essential to use a convocational neural network to solve the problem that arises since the aim is to identify the areas in which people are located. Therefore, other architectures of this type of neural network have been evaluated. However, since the Faster R-CNN is one of the most powerful architectures [19] in this first approach to this problem, it has been decided to use this architecture.

The temperatures obtained by the Grid Eye sensor are used to generate a set of images for training and testing the architecture. During this process it is important to have a variety of data; therefore it is crucial to obtain data with the different classes that the system is going to deal with and in different positions as well. Our proposed system works with three different classes, one person group, two people group, and three people group. Once the system has the temperature values per every frame captured we normalized the data using the minimun and maximun values of the dataset.

After the normalization, we can create images according to the new values for every pixel in every frame. However, these images cannot be used for training the model because they can contain heat sources that are not related to people. Hence, before training the model the system has to remove the background of these images. To solve this problem, the system collects data when there are no people in the field of view of the sensor so, it can obtain statistics for every pixel. With these statistics, the system can compare if the values of the pixels are extreme. If the extreme value is over the normal values, this pixel is identified as a part of a person or a group of people. After that, interpolation is applied to every image to have more sharpened edges that allow the model to be trained in a better way.

It is at that moment when the system can start the model training. A subset of these images will be the ones that the system is going to use for training a Faster R-CNN model. The rest of the images are going to be used for testing the generated model. The training subset has to be labeled manually so the NN learns how to interpret the heat sources printed in the images. Once the training process is finished, a model will be generated. This model will be tested with the rest of the images to determine the efficiency of the proposed system.

Figure 1 includes an illustrated summary of the proposal's process.

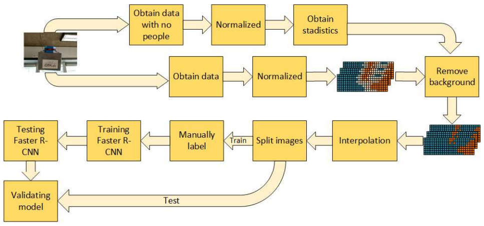


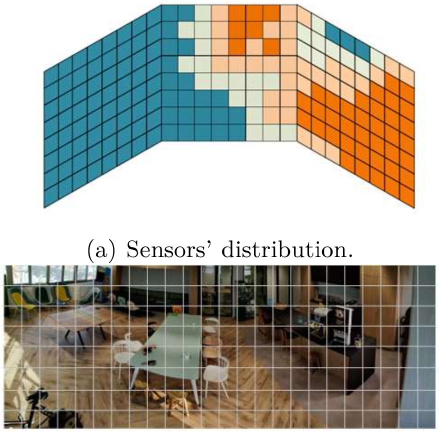
Fig. 1: Proposal's process.

IV. CASE STUDY

The thermal sensors are integrated on the Octa-Connect platform, a fast-prototyping platform for IoT applications. Once the sensors are mounted on the Octa-Connect platform it is deployed in a cafe room inside an office of the Antwerp University and imec research facilities. The experiments had been performed under controlled circumstances.

For this study, three Grid Eye sensors have been used. Some related works explained in Section II use only one Grid Eye sensor located on the ceiling of the rooms [10], [11]. Placing the sensor in this position improves the results of the classifiers but the system seems to be very limited. Therefore, in this experiment, the sensors are more diagonally concerning the floor of the room. With this position and with the use of three sensors the system can increase its field of view and have a more panoramic view of the room. In Figure 2 we can see the distribution of the sensors and an approximation of its view.

To obtain the statistics that allow the system to remove the background of the images, it is necessary to collect data with no people in the room. For this study, data has been collected in different hours of the day to contemplate changes in the temperature inside the room during the different hours of the day. Also, to train and test the Faster R-CNN model data has been recollected. This data has been collect taking into account, three different classes.



(a) Sensors' distribution.

(b) Sensors' view.

Fig. 2: Sensor's distribution and view.

The data obtained for the training and testing processes include every one of these classes in the most balanced possible way. During the collecting people enter and exit the field of view of the sensors.

Hence, once the images are generated according to the process explained in Section III we configure the Faster R-CNN model. In this study, the architecture used is developed by Bardool [20]. This Github repository includes a Faster R-CNN developed in Keras and Python. The implementation of the code allows implementing the NN without applying relevant changes. However, it is extremely important to configure the data in the correct format to proceed with the training properly.

V. RESULTS

First, we analyze the performance of the filter applied to the thermal images to remove their background. If the images are analyzed one by one by manually verified, the results of the background removal seem to work perfectly. It is true that in certain images in which people or groups of people are far from the sensor (±4 meters), the technique used is not able to determine the existence of them. But even for a person who reviews the images, it is really difficult to determine whether or not there are people at those points in the images. So it could be said that the technique used works correctly. In Figure 3 you can see a set of images before and after applying the mentioned filter.

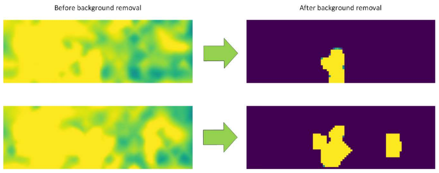


Fig. 3: Background removal.

After the pre-processing the images, it is turn to analyze the model generated by the Faster R-CNN. It is worth to mention that the results have not been compared with any other model since the distribution and orientation of the sensors are not the same as in other studies carried out with the same sensor. To evaluate this model, the results obtained from the training stage and the test stage will be distinguished. Table I presents the parameters used for the training. Figure 4 and Table II show the results of the training.

TABLE I: Training parameters

N° of classes	3
Classes	person, group2, group3
Instances class person	413
Instances class group2	347
Instances class group3	205
Number of instances	851
Number of training instances	695
Number of validation instances	156
Configured epochs	2000
Elapsed epochs	208

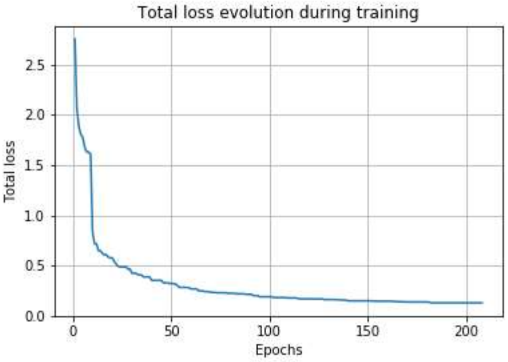


Fig. 4: Total loss evolution during training process.

TABLE II: Model losses

Loss RPN classifier	0.0180868
Loss RPN regression	0.0031488
Loss Detector classifier	0.1008769
Loss Detector regression	0.0087452
Total loss	0.1308577

After finishing the training process, the generated model will be able to detect in the test images the different heat zones, Figure 5. If we analyze the output of the model we observe that the model has a good accuracy given the complexity of the problem. Its accuracy value is 72.85%. By looking at the classified images manually, it can be seen that errors occur when heat areas have been detected away from the sensors (±4 meters). In these cases, the heat sources are practically identical to each other. This behaviour is something normal in object detection problems.

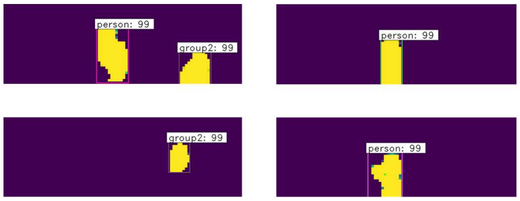


Fig. 5: Examples of outputs of the system.

Finally, we have decided to include a comparison table between our model and other proposed by other authors who also use the Grid-Eye sensor for counting people. However, before looking at Table III it is necessary to specify again that these studies are not comparable since the disposition of the sensors is not the same and therefore, they are studies that have not been carried out under the same circumstances.

TABLE III: Models comparison

Study	RMSE
[11]	0.346
Our study	0.524

VI. CONCLUSIONS

This article shows a deep learning model that detects human beings in low-resolution thermal images extracted with three Grid Eye sensors on a diagonal position.

In this way, and taking into account the conclusions drawn from the experiments carried out several future lines of research has been detected. The replacement of Grid Eye sensors by other types of sensors. If it is possible to obtain a model based on CNN of high efficiency after changing the type of sensor that captures the data, another interesting aspect to work as a future line of research is the creation of a new model capable of working in IoT devices of low capacities. From another point of view, it is also interesting to test the current model with a much larger set of data.

ACKNOWLEDGMENT

The research of Diego M. Jiménez-Bravo has been co-financed by the European Social Fund and Junta de Castilla y León (Operational Programme 2014-2020 for Castilla y León, EDU/574/2018 BOCYL). This work was supported by the European H2020 framework programme project Fed4Fire+ (732638).

References

[1] Sandra L. Colby and Jennifer M. Ortman, "Projections of the Size and Composition of the U.S. Population: 2014 to 2060," U.S. Department of Commerce, Tech. Rep., 2015.

[2] "People Counting Sensors &amp; Software for Businesses." [Online]. Available: https://www.density.io/

[3] L. Spinello and K. O. Arras, "People detection in RGB-D data." Institute of Electrical and Electronics Engineers (IEEE), 12 2011, pp. 3838–3843.

[4] K. N. Ha, K. C. Lee, and S. Lee, "Development of PIR sensor based indoor location detection system for smart home," in 2006 SICE-ICASE International Joint Conference, 2006, pp. 2162–2167.

[5] V. L. Erickson, A. Beltran, D. A. Winkler, N. P. Esfahani, J. R. Lusby, and A. E. Cerpa, "ThermoSense: thermal array sensor networks in building management," in Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems - SenSys '13. New York, New York, USA: ACM Press, 2013, pp. 1–2. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2517351.2517437

[6] C. Wu, Z. Yang, Z. Zhou, X. Liu, Y. Liu, and J. Cao, "Non-invasive detection of moving and stationary human with WiFi," IEEE Journal on Selected Areas in Communications, vol. 33, no. 11, pp. 2329–2342, 11 2015.

[7] "US9224284B2 - Detecting presence using a presence sensor network - Google Patents." [Online]. Available: https://patents.google.com/patent/US9224284B2/en

[8] "US4779240A - Ultrasonic sensor system - Google Patents." [Online]. Available: https://patents.google.com/patent/US4779240A/en

[9] M. Kuki, H. Nakajima, N. Tsuchiya, and Y. Hata, "Multi-human locating in real environment by thermal sensor," in Proceedings - 2013 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2013, 2013, pp. 4623–4628.

[10] A. A. Trofimova, A. Masciadri, F. Veronese, and F. Salice, "Indoor Human Detection Based on Thermal Array Sensor Data and Adaptive Background Estimation," Journal of Computer and Communications, vol. 05, no. 04, pp. 16–28, 2017. [Online]. Available: http://www.scirp.org/journal/doi.aspx?DOI=10.4236/jcc.2017.54002

[11] A. Beltran, V. L. Erickson, and A. E. Cerpa, "ThermoSense: Occupancy Thermal Based Sensing for HVAC Control," in Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings - BuildSys'13. New York, New York, USA: ACM Press, 2013, pp. 1–8. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2528282.2528301

[12] Y. Jeong, K. Yoon, and K. Joung, "Probabilistic method to determine human subjects for low-resolution thermal imaging sensor," in 2014 IEEE Sensors Applications Symposium (SAS). IEEE, 2 2014, pp. 97–102. [Online]. Available: http://ieeexplore.ieee.org/document/6798925/

[13] A. Tyndall, R. Cardell-Oliver, and A. Keating, "Occupancy Estimation Using a Low-Pixel Count Thermal Imager," IEEE Sensors Journal, vol. 16, no. 10, pp. 3784–3791, 5 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7409929/

[14] T. Y. Chen, C. H. Chen, D. J. Wang, and Y. L. Kuo, "A people counting system based on face-detection," in Proceedings - 4th International Conference on Genetic and Evolutionary Computing, ICGEC 2010, 2010, pp. 699–702.

[15] C. Gao, P. Li, Y. Zhang, J. Liu, and L. Wang, "People counting based on head detection combining Adaboost and CNN in crowded surveillance environment," Neurocomputing, vol. 208, pp. 108–116, 10 2016.

[16] H. Ma, H. Lu, and M. Zhang, "A real-time effective system for tracking passing people using a single camera," in Proceedings of the World Congress on Intelligent Control and Automation (WCICA), 2008, pp. 6173–6177.

[17] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," 6 2015. [Online]. Available: http://arxiv.org/abs/1506.01497

[18] "OCTA-Connect · GitHub." [Online]. Available: https://github.com/octa-connect

[19] "Deep Learning for Object Detection: A Comprehensive Review." [Online]. Available: https://towardsdatascience.com/deep-learning-for-object-detection-a-comprehensive-review-73930816d8d9

[20] Kevin Bardool, "kbardool/keras-frcnn - Github." [Online]. Available: https://github.com/kbardool/keras-frcnn