Object localization in handheld thermal images for fireground understanding

Florian Vandecasteele, Bart Merci, Azarakhsh Jalalvand, and Steven Verstockt

In: PROCEEDINGS OF SPIE, 10214 (Thermosense: Thermal Infrared Applications 39), 2017.

Object Localization in Handheld Thermal Images for Fireground Understanding

Florian Vandecasteele*[a], Bart Merci[b], Azarakhsh Jalalvand[a], and Steven Verstockt[a]

[a]Ghent University, imec, ELIS Department - IDlab
[b]Ghent University, Department of Flow, Heat and Combustion Mechanics
Sint-Pietersnieuwstraat 41, 9000 Ghent, Belgium

ABSTRACT

Despite the broad application of the handheld thermal imaging cameras in firefighting, its usage is mostly limited to subjective interpretation by the person carrying the device. As remedies to overcome this limitation, object localization and classification mechanisms could assist the fireground understanding and help with the automated localization, characterization and spatio-temporal (spreading) analysis of the fire. An automated understanding of thermal images can enrich the conventional knowledge-based firefighting techniques by providing the information from the data and sensing-driven approaches. In this work, transfer learning is applied on multi-labeling convolutional neural network architectures for object localization and recognition in monocular visual, infrared and multispectral dynamic images. Furthermore, the possibility of analyzing fire scene images is studied and their current limitations are discussed. Finally, the understanding of the room configuration (i.e., objects location) for indoor localization in reduced visibility environments and the linking with Building Information Models (BIM) are investigated.

Keywords: Thermal fire analysis; object detection and recognition; convolutional neural networks

1. INTRODUCTION

Currently, the images retrieved by handheld thermal imagers during a fire incidence are used in a passive way without generating any automatic feedback. This work is, hence, a first step towards more active/automated fire sensing and investigates the possibilities of transfer learning for object detection and classification in the visual, thermal and multispectral domain. The object localization module presented in this paper is a key building block of a larger architecture that helps in the understanding of real fires.

Automatic image recognition is a hot topic research in the field of visual data analysis, but there is limited study on the application of such techniques on infrared images. Infrared object recognition can be vital and beneficial in circumstances with reduced visibility or environments with varying illumination. Uptil now, however, the majority of research on infrared image analysis only focuses on the detection of humans,[1,2] discarding the other types of objects. An exception is the work of Gundogdu et al.[3] that proposes a binary decision framework in combination with deep convolutional features to classify thermal images as ship/boat, tank, plane or helicopter. As will be explained later in this paper, this approach is slightly similar to our proposed methodology.

Thermal images are useful modalities for investigating low-visibility or dark environments, e.g., during a fire when there is limited visibility due to smoke. In buildings, for example, that contain many different objects made of different materials (e.g., wood, plastic, steel), a thermal imager can be used to analyze the amount of heat radiated by each material. For many objects, these radiation profiles are unique, making it possible to recognize objects independent of the illuminance. For this reason, the focus of this paper is the recognition of building interior objects by means of handheld thermal images, i.e., a research topic that has not been addressed so far.

Further author information: send correspondence to Florian vandecasteele: E-mail: florian.vandecasteele@ugent.be

The remainder of this paper is organized as follows. Section 2 presents the global work flow of the fireground understanding mechanisms. Subsequently, Section 3 describes the transfer learning process for object localization and recognition along with results of the indoor object detection. In Section 4, we discuss the utilization of handheld thermal imagers and the integration of the object detection within a BIM package. Finally, Section 5 concludes the paper and points out directions for future work.

2. FIREGROUND UNDERSTANDING WITH THERMAL IMAGES

The localization of objects in real fire thermal images is helpful in different ways (e.g., to understand which element is burning or to easily find the way back). However, it is important to mention that only items which irradiate infrared light can be detected. If the indoor setting is made of equal materials and has no external heating, the mechanism will fail, i.e., a limitation of the proposed methodology.

Our framework for automated fireground understanding is shown in Figure 1. The first component of the framework is the detection and reporting of the element that is burning (i.e., fhe focus of this paper). The output of this building block can be further used as input for modeling and forecasting of fires with Computational Fluid Dynamic (CFD) packages.[4] The second component is used to understand the room configuration (i.e., object appearance and location) and helps to localize the interiors in reduced visibility environments. Thirdly, object classification is used to facilitate the scene recognition task, which classifies the images into different scene classes such as kitchen, cellar or living room. Finally, the object detection from monocular thermal images will facilitate the interpretation of the fire to guide autonomous firefighting robots. An example of the final textual output of our proposed framework is "The fridge is burning in the left corner of the cellar and the fire is located in the bottom".

In Section 3.1, we will thoroughly discuss our indoor dataset, but it is important to remark that the selected semantic objects are especially useful for the determination of the heat release rate (HRR) in an indoor/ room setting. Kim et al.[5] created a verification set in which the maximum heat release rate, the thermal penetration length and the characterization time are indicated for typical room objects such as bed, closet, couch, chair, table. In our dataset, similar object classes are determined to make the interchangeability of the data possible in the future. Furthermore, with the heat release rate it is possible to estimate the potential of an occurring flashover.[6]

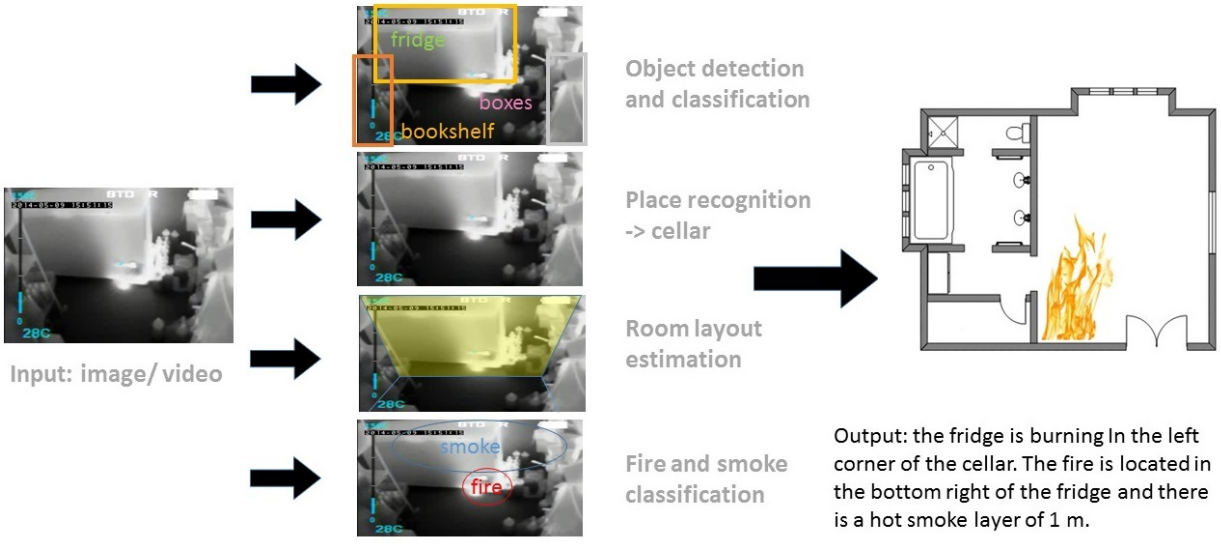

Figure 1. Thermal image analysis pipeline for fireground understanding.

Within the next section we will thoroughly describe the object localization block for fireground understanding. This module for automated reporting of the fireground is only a part of the proposed framework. Future work will discuss the other building blocks (i.e., scene classification, the room layout estimation block and the smoke-fire segmentation block).

3. ACTIVE TRANSFER LEARNING FOR OBJECT LOCALIZATION

Object localization consists of two major stepts: object detection and object recognition. The aim of *Object detection* is to find out where an object is located in the image, whereas *object recognition* identifies which semantic class (e.g., couch, chair, table) an object belongs to. Even though it is a very hot topic in the field of computer vision, there are still some major difficulties to fully automate the recognition pipeline. This is mostly due to the large variety of object appearances due to changes in illumination, viewpoint and non-rigid deformations. A common approach to tackle this issue in the visual domain is to employ multi-label convolutional neural network (CNN) architectures.[7,8] In this work, we apply fine-tuning techniques to obtain comparable results in the thermal and multispectral domain. Girshick et al.[9] stated that supervised pre-training on a large auxiliary dataset such as ImageNet [*], followed by domain-specific fine-tuning on a small dataset is an effective paradigm for learning high-capacity CNNs when the available data is scarce. Lu et al.[10] on the other hand showed that transfer learning largely enhances the localization accuracy in the dark environments.

3.1 Thermal and visual aligned dataset creation

In order to evaluate the proposed methodology, a dataset is constructed with thermal, visual and Multi Spectral Dynamic Imaging (MSX) images. We believe that it is the first indoor scene dataset that combines different spectra of indoor objects. Vidas et al.[11] introduced a related dataset with aligned RGB-D and thermal images from building interiors, but no bounding box annotations were provided. The images of our dataset are captured with a Flir One [†] thermal camera. The main advantage of this device is that thermal and visual images are aligned. This enables us to easily transfer the annotations in the visual domain to the thermal domain without any further calibration, outlining or rectifying compared to the previous visual-thermal datasets, in which a check-board with high contrast in thermal wavelengths was used.[12] Furthermore, FLIR offers a MSX-technique that combines the visual and thermal image in one multispectral image. This results in a sharper look image compared to a standard thermal image. It is also worth to mention that the samples of the current dataset has been collected in a standard indoor, non-heated circumstances. As a future work, we will investigate the impact of heating sources during a fire for the object localization module. In this regard, several real fire footages will be collected and annotated. Furthermore, novel multispectral devices[13] will also be evaluated in the fire investigation context.

The visual domain samples of our dataset were annotated with a bounding box and semantic class label with the well-known LabelMe toolkit.[14] The annotations are then transported to the thermal and MSX-domain. Table 3.1 lists an overview of the amount of samples per class. Finally, it is important to remark that the thermal images are taken with a low-cost handheld device and there could be some artifacts in the images due to clutter or motion instability.

Table 1. Enclosure object dataset

Object name	Number of samples	Object name	Number of samples
Closet	247	Window	57
TV or Screen	29	Person	8
Lamp	227	Table	141
Bed	30	Chair	173
Couch	141	Potted plant	56

[*] www.image-net.org
[†] http://www.flir.eu/flirone/ios-android/

3.2 Multi-label convolutional network

The neural network that we use is based on the Fast-CNN (Recurrent Conventional Neural Network) architecture introduced in.[8] This work initiated the automated region proposal technique which combines the object bounds detection and the objectiveness score on a position. This is contradictory to conventional methods that utilize selective search[15] or low-level features such as superpixels.[16] First, we adapted the final Fast-CNN layer to the ten classes in our dataset. Then, the weights of the VGG16 (Visual Geometry Group)[17] pre-trained on the COCO-dataset[18] are used to initiate the architecture. In the next step, the learning rate was lowered and early stopping was used to avoid overfitting on the new indoor object dataset. Beside the small learning rate, it is also possible to use an aggressive decaying learning rate[19] every few epochs, but this did not lead to any major changes/improvements. Finally, the joint approximate training was used to jointly train the Region Proposal Network (RPN) module[8] and the Fast-RCNN network. The main advantage of this approach is that the network automatically learns the underlying representation of the data without manual intervention.

3.3 Evaluation and discussion of the object detection module

The fine-tuned network has been evaluated with cross validation on 80% of the dataset while the rest of the dataset was considered as the test set. The mean Average Precision (mAP), with values between 0 and 1 (higher is better), is commonly used for the evaluation of object localization. While the average corresponds to the area under the precision-recall curve for a class, the mean of these average individual-class-precision gives the mean Average Precision. In addition, a qualitative evaluation of the object detection is performed on the training and test sets. Some examples are shown in Figure 2. From these examples we observe that there are more objects detected in the visual images. Also, the certainty of the predicted objects in the visual spectrum is higher compared to the thermal images (as could be epected). Nevertheless, the fact that many objects are correctly detected in the thermal image supports the potential and value of our proposed methodology.

In order to improve the recognition performance, some traditional methods as histograms stretching and gamma correction are used to make the images more clear and to increase the contrast. In the case of the thermal images, the mean average precision increased by 22% relatively compared to the non-stretched images. Table 3.3 depicts the mean average precision for the object detection module on the testset. Please note that the threshold value is placed on 75% overlap with the labeled dataset. In this table, we also compare the output of the pre-trained weights of the COCO dataset. According to the results, the fine-tuned system significantly suppresses the traditional COCO system. This confirms the statement of Girshick et al.[9] The results for the thermal images are lower than the visual images, but this was already indicated by the subjective evaluation. Furthermore, there are some objects that radiate the same amount of heat in standard environments which makes the object localization task more difficult.

Table 2. Mean average precision on the test dataset

Object name	mAP COCO visual	mAP fine-tuned visual	mAP thermal	mAP MSX
Bed	0.171	0.418	0.143	0.477
Chair	0.102	0.348	0.027	0.365
Closet	0.006	0.237	0.024	0.196
Couch	0.330	0.381	0.108	0.413
Lamp	Not in the dataset	0.171	0.059	0.132
Person	0.035	0.500	0.500	0.500
Potted plant	0.000	0.089	0.001	0.083
Table	0.000	0.281	0.023	0.319
Tv or screen	0.250	0.129	0.018	0.533
Window	Not in the dataset	0.141	0.056	0.135
Average	0.112	0.269	0.096	0.315

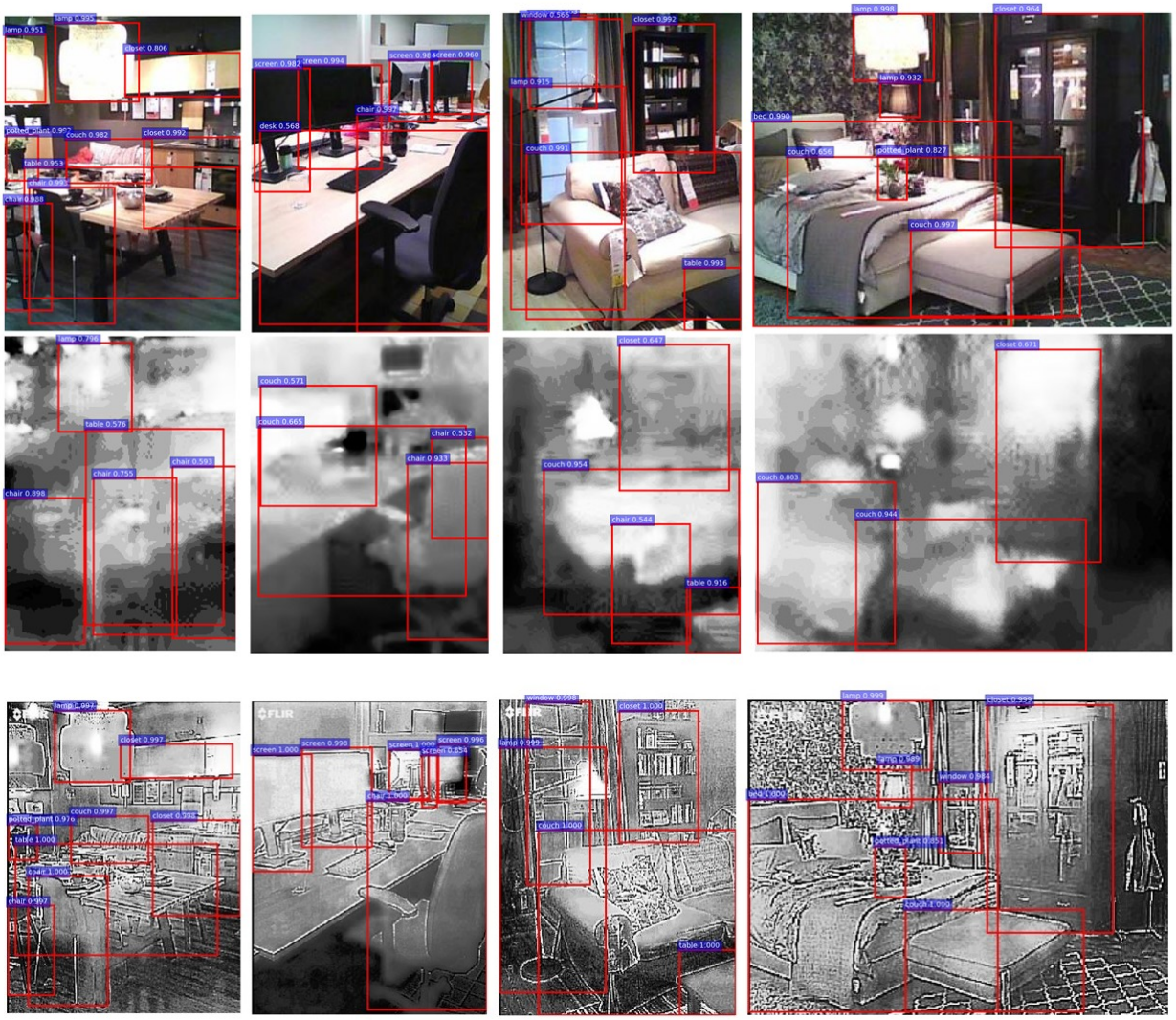

Figure 2. Subjective evaluation of the object detection framework in the Visual, Thermal and MSX domain

The precision on the multispectral imageset (MSX) is higher compared to the thermal images for some classes, Furthermore, some noticeable improvement is measured with the average mAP compared to the fine-tuned visual dataset. However, it will not be possible to use this traditional technique of combining visual and thermal images in reduced visibility environments during a fire (i.e., visual objects are not visible). Merging different spectra could possibly solve this issue and increase the situational awareness and the understanding of the fireground during a fire and will be part of future work. Finally, given that the current dataset is rather small, we plan to enlarge the dataset with additional samples in the near future, which can potentially improves the performance. More data samples will most likely increase the final precision of the system, but our initial results show the possibilities of our proposed approach.

4. PRACTICAL INTEGRATION OF OBJECT DETECTION

A Building information models (BIM) is a rich source of information for object verification, i.e., BIMs have many valuable information for fireground understanding, evacuation or fire safety design. Uptil now, however, BIM are not actively used in firefighting. In the proposed framework, the detected objects could be linked to a building information model to update the building model or to retrieve more detailed information of the object. This is further discussed in Section 4.1. Furthermore, it is important to mention that the performance of the object localization framework in real fire incidences is not only dependent on the selected thermal imaging device, but also on the correct handling of the device. For this reason, operational guidelines are discussed in Section 4.2.

4.1 Linking object detection and Building information models

Uptil now, limited research focuses on the possibilities of linking object detection and building information data. Kropp et al.[20] match visual objects in a construction site with a building model through Histogram Of Gradients (HOG) features and a Support Vector Machine (SVM) classifier. Choi et al.[21] on the other hand investigated the semantic understanding of 3D-furniture objects from images. Combining both research tracks within the Industry Foundation Class (IFC) or the IndoorGML[22] standard for BIM could facilitate the representation, storage or exchange of spatio-temporal information of the indoor space. The generation or matching of the indoor scene data is possible with the position and the semantic labels of the detected objects (i.e., thermal or visually annotated). Future work will focus on the matching process between our detected objects and the furniture elements (e.g., chair, couch, table) in the building model. In the IFC model the furniture elements are described within the IfcFurnishingElement class. They can independently be simple objects with relatively simple geometries or complex objects composed of different geometries. Relating both types of geometrical data will be the major challenge of future research. Finally, the integration of the recognition pipeline in a handheld device will be another challenge to fully optimize the fireground understanding.

4.2 Object detection for fire incident management

The image quality of handheld thermal cameras is increasing while the price is decreasing. As a result, more and more fire brigades actively use these devices. Albeit, there are currently no practice guides to use these imaging devices. Amon et al.[23] investigated the problems to use a thermal imaging device in a firefighting context, but limited focus was put on the correct handling of the device. To fully integrate our proposed object detection mechanism for fireground understanding, more research on this topic is needed. Meanwhile, some practical advices for proper use are given based on our current expertise:

- point the camera foreward to get a complete overview of the room; usually during a fire incident the camera is only pointed upwards,
- avoid the presence of persons (i.e., other firefighters) in the field-of-view to ensure a maximal detection and recognition of objects,
- avoid the pointing on reflective surfaces, such as mirrors or metal doors, to avoid false positives for the object detection,
- move the camera slowly to prevent the image from freezing and to avoid/limit the artifacts creation,
- while moving into a building, occasionally look behind you to make sure the path you have taken is clear, and keep track of any smoke or fire changing conditions,
- it is important to know that a thermal camera can not look through furniture or under debris. Flames or furniture objects could not been detected if they are not in the field-of-view.

Besides the research on the application of thermal imaging devices in a firefighting context there is ongoing research to make the handling of a thermal device easier. Scott Sight built an in-mask thermal sensor [‡] and

[‡] http://www.scottsafetynation.com/scott-sight-nfpa-1981-certified-now-shipping/

VIZIR [§] offers a hands-free operation mask. There are some differences between both systems. For instance, the Scott Sight has no external camera and a thermal viewer in the corner, while the VIZIR has its augmented reality images in the center. Further research will be necessary to achieve full fireground understanding from these devices. Currently, there is no interaction between different crews and the information seen on the screen is not automatically interpreted. To handle this issue, some recent papers try to make the thermal imaging more interactive. Paugam et al.,[24] for example, use the hand held thermal devices to derive Fire Radiative Power (FRP) and flame front Rate Of Spread (ROS) in case of a wildland fire. Kim et al.[25] use a probabilistic classification to identify fire and smoke regions from a single thermal camera. In combination with the proposed object detection tool (Section 3) this will contribute to the automated fireground understanding.

5. CONCLUSIONS AND FUTURE WORK

In this work, we proposed a framework for automated fireground understanding. Furthermore, we investigated the application of multi-labeling convolutional neural networks on object detection in visual, infrared and multi spectral dynamic images. Subsequently, some advices for practical use of a thermal imaging device for firefighting were given. The work presented in this paper is in its early phase and further evaluation and research will be necessary. Nevertheless, it is important to automate the fireground understanding to guide firefighting robots or to assist firefighters.

Future work will evaluate the proposed methodology on a larger dataset and in more challenging circumstances, e.g., during real fire incidents and experiments. Furthermore, the fireground understanding framework will be further extended with automated scene recognition and room layout estimation. Moreover, recent released models and framework for object detection[26] will be evaluated in this context. Finally, the object detector will be used to automatically generate or update building model information.

6. ACKNOWLEDGMENTS

The research activities as described in this paper are funded by imec and Ghent University through GOA project BOF16/GOA/004.

REFERENCES

[1] Wu, Z., Fuller, N., Theriault, D., and Betke, M., "A thermal infrared video benchmark for visual analysis," in [*The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*], (2014).

[2] Berg, A., Ahlberg, J., and Felsberg, M., "A thermal object tracking benchmark," in [*Advanced Video and Signal Based Surveillance (AVSS), 2015 12th IEEE International Conference on*], 1–6, IEEE (2015).

[3] Gundogdu, E., Koç, A., and Alatan, A. A., "Object classification in infrared images using deep representations," in [*Image Processing (ICIP), 2016 IEEE International Conference on*], 1066–1070, IEEE (2016).

[4] Beji, T., Verstockt, S., Van de Walle, R., and Merci, B., "On the use of real-time video to forecast fire growth in enclosures," *Fire Technology* **50**(4), 1021–1040 (2014).

[5] Kim, H.-J. and Lilley, D. G., "Heat release rates of burning items in fires," *Journal of propulsion and power* **18**, 866–870 (2002).

[6] Babrauskas, V., "Estimating room flashover potential," *Fire Technology* **16**, 94–103 (1980).

[7] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A., "You only look once: Unified, real-time object detection," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 779–788 (2016).

[8] Ren, S., He, K., Girshick, R., and Sun, J., "Faster r-cnn: Towards real-time object detection with region proposal networks," in [*Advances in neural information processing systems*], 91–99 (2015).

[9] Girshick, R., Donahue, J., Darrell, T., and Malik, J., "Rich feature hierarchies for accurate object detection and semantic segmentation," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 580–587 (2014).

[§] http://www.darix.ch/

[10] Lu, G., Yan, Y., Ren, L., Saponaro, P., Sebe, N., and Kambhamettu, C., "Where am i in the dark: Exploring active transfer learning on the use of indoor localization based on thermal imaging," *Neurocomputing* **173**, 83–92 (2016).

[11] Vidas, S., Moghadam, P., and Bosse, M., "3d thermal mapping of building interiors using an rgb-d and thermal camera," in [*Robotics and Automation (ICRA), 2013 IEEE International Conference on*], 2311–2318, IEEE (2013).

[12] Engström, P., Larsson, H., and Rydell, J., "Geometric calibration of thermal cameras," in [*SPIE Security+ Defence*], 88970C–88970C, International Society for Optics and Photonics (2013).

[13] Lambrechts, A., Gonzalez, P., Geelen, B., Soussan, P., Tack, K., and Jayapala, M., "A cmos-compatible, integrated approach to hyper-and multispectral imaging," in [*Electron Devices Meeting (IEDM), 2014 IEEE International*], 10–5, IEEE (2014).

[14] Russell, B. C., Torralba, A., Murphy, K. P., and Freeman, W. T., "Labelme: a database and web-based tool for image annotation," *International journal of computer vision* **77**, 157–173 (2008).

[15] Uijlings, J. R., Van De Sande, K. E., Gevers, T., and Smeulders, A. W., "Selective search for object recognition," *International journal of computer vision* **104**, 154–171 (2013).

[16] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Süsstrunk, S., "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE transactions on pattern analysis and machine intelligence* **34**, 2274–2282 (2012).

[17] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al., "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision* **115**, 211–252 (2015).

[18] Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L., "Microsoft coco: Common objects in context," in [*European Conference on Computer Vision*], 740–755, Springer (2014).

[19] Yosinski, J., Clune, J., Bengio, Y., and Lipson, H., "How transferable are features in deep neural networks?," in [*Advances in neural information processing systems*], 3320–3328 (2014).

[20] Kropp, C., König, M., and Koch, C., "Object recognition in bim registered videos for indoor progress monitoring," in [*EG-ICE International Workshop on Intelligent Computing in Engineering*], (2013).

[21] Choi, W., Chao, Y.-W., Pantofaru, C., and Savarese, S., "Indoor scene understanding with geometric and semantic contexts," *International Journal of Computer Vision* **112**(2), 204–220 (2015).

[22] Kim, J.-S., Yoo, S.-J., and Li, K.-J., "Integrating indoorgml and citygml for indoor space," in [*International Symposium on Web and Wireless Geographical Information Systems*], 184–196, Springer (2014).

[23] Amon, F. K., Bryner, N. P., and Hamins, A., [*Thermal imaging research needs for first responders: workshop proceedings*], US Department of Commerce, Technology Administration, National Institute of Standards and Technology (2005).

[24] Paugam, R., Wooster, M. J., and Roberts, G., "Use of handheld thermal imager data for airborne mapping of fire radiative power and energy and flame front rate of spread," *IEEE Transactions on Geoscience and Remote Sensing* **51**, 3385–3399 (2013).

[25] Kim, J.-H. and Lattimer, B. Y., "Real-time probabilistic classification of fire and smoke using thermal imagery for intelligent firefighting robot," *Fire Safety Journal* **72**, 40–49 (2015).

[26] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A. C., "Ssd: Single shot multibox detector," in [*European Conference on Computer Vision*], 21–37, Springer (2016).