Object Detection from Thermal Infrared and Visible Light Cameras in Search and Rescue Scenes

Adrián Bañuls, Anthony Mandow, Ricardo Vázquez-Martín, Jesús Morales and Alfonso García-Cerezo[1]

***Abstract*— Visual object recognition is a fundamental challenge for reliable search and rescue (SAR) robots, where vision can be limited by lighting and other harsh environmental conditions in disaster sites. The goal of this paper is to explore the use of thermal and visible light images for automatic object detection in SAR scenes. With this purpose, we have used a new dataset consisting of pairs of thermal infrared (TIR) and visible (RGB) video sequences captured from an all-terrain vehicle moving through several realistic SAR exercises participated by actual first response organisations. Two instances of the open source YOLOv3 convolutional neural network (CNN) architecture are trained from annotated sets of RGB and TIR images, respectively. In particular, frames are labelled with four representative classes in SAR scenes comprising both persons (*civilian* and *first-responder*) and vehicles (*Civilian-car* and *response-vehicle*). Furthermore, we perform a comparative evaluation of these networks that can provide insight for future RGB/TIR fusion.**

I. INTRODUCTION

A thermal infrared (TIR) camera was employed in the first reported life save by a robot in 2013 [1]. At this point, infrared imagery is a decisive imaging modality not only for search and rescue (SAR) [2][3][4], but also for other robotic applications such as surveillance [5], military [6] and autonomous driving [7][8]. In comparison with visible light cameras (RGB), TIR cameras can be more robust against smoke, fog and lighting conditions [9]. Nevertheless, thermal radiation produces images lacking contrast and texture information [10], so the combination with other modalities can be advantageous for effective object identification [2].

Specially, synergy between thermal and visible images can be helpful to distinguish between rescuers and civilians, to identify survivors, or to recognise different kinds of vehicles. Besides, the RGB/TIR combination can produce an intuitive modality output for human rescuers [11] and can also benefit from recent deep learning tools for automatic object detection and scene understanding. Recently, state-of-the-art convolutional neural networks (CNN) models for object detection, such as single shot multi-box detector (SSD) [12] and YOLO [13] have achieved impressive real time performance with visible light images [7][14] in growing application domains [15][16].

A few works have extended the use of YOLO to thermal imaging, mainly with a focus on nighttime person detection. Thus, [17] addressed the problem of detecting distant persons and vehicles with small pixel sizes for surveillance and border control. In [5], a model trained on a TIR dataset clearly outperformed the original RGB-trained model for person detection under different weather conditions. Furthermore, the real-time qualities of YOLO were exploited in [18] for nighttime pedestrian detection from a moving TIR camera by applying a prior saliency stage. Other works have used CNNs to boast performance and accuracy when visible images are combined with other sensing modalities [19][20]. Thus, YOLO networks were used in [21] for semantic mapping from RGB images with depth information by incorporating a three-dimensional (3D) segmentation algorithm, and in [8] for combining frame- and event-driven images for pedestrian detection.

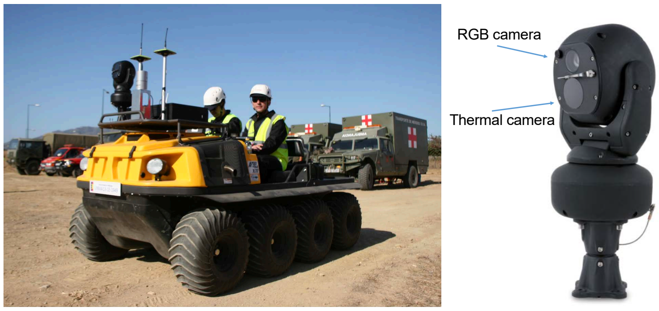

Fig. 1. All-terrain vehicle used to obtain the datasets (left), and the Oculus TI dual camera [22] (right).

Another indication of the growing interest on TIR image processing is the recent publication of different datasets, such as a far infrared (FIR) dataset for on-road pedestrian detection [23], a combination of visual and thermal data for person tracking in urban environments [24], and a multispectral dataset for day and nighttime driving [25]. Moreover, a specific dataset for SAR robotics has been constructed with multimodal (RGB, small field-of-view thermal, and depth) measurements of several indoor search scenarios as well as semi-synthetic images of victims [26].

In this work, our goal is to contribute to filling the gap in combined use of TIR and visible light images in the disaster robotics field. In particular, we explore automatic object detection in SAR scenes with TIR images and their complementarity with visible images. The major novel contributions of the paper are the following:

- We use a specific SAR dataset consisting of pairs of thermal and visible video sequences captured from an all-terrain vehicle (see Fig. 1) moving through several realistic SAR exercises performed by actual first re-

This work has been done in the framework of the TRUST-ROB project, funded by the Spanish Government (RTI2018-093421-B-I00).

[1] Universidad de Málaga, Robotics and Mechatronics Group, Andalucía Tech, 29071 Málaga, Spain. amandow@uma.es, ajgarcia@uma.es