Thermal Imaging Dataset for Person Detection

M. Krišto, M. Ivašić-Kos
Department of Informatics, University of Rijeka, Rijeka, Croatia
matekrishto@gmail.com, marinai@uniri.hr

Abstract – In this paper will be presented an original thermal dataset designed for training machine learning models for person detection. The dataset contains 7412 thermal images of humans captured in various scenarios while walking, running, or sneaking. The recordings are captured in the LWIR segment of the electromagnetic (EM) in various weather condition- clear, fog and rain at different distances from the camera, different body positions (upright, hunched) and movement speeds (regular walking, running). In addition to the standard lens of the camera, we used a telephoto lens for recording, and we compared the image quality at different weather conditions and at different distances in both cases to set parameters that provide the level of detail that is enough to detect the person.

Keywords: Person detection, thermal imaging, surveillance, access control, dataset, biometric

I. Introduction

The security surveillance system is expected to alarm any unauthorized movement (sneaking, running, etc.) in the area of the protected facilities and the state border by day and night and all weather conditions. Our project aims to define a machine learning model for the detection and recognition of people on the thermal images that could be implemented in the surveillance system to assist the protection of people, facilities, and areas.

To learn a model, necessary is to choose an adequate learning dataset. A learning dataset to be of high quality and reliability must contain a large number of images where objects are displayed in different ways, on different lighting conditions, recording angles, different distances from the camera, etc. When a dataset satisfies these conditions, it is more likely that a learned model will generalize better for various situations that may occur in real life, or in our case, in the surveillance area.

A data set can be created by using and combining images from existing data sets, capturing new images according to the defined scenarios that best fit the needs and research goals, and combining images from existing data sets and own recordings.

In our case, we have created a dataset by recording our images, which is perhaps the hardest and most demanding approach. On the other hand, in such a way, we had the opportunity to create a dataset that best suits the needs of our research and the goal of detecting people in harsh weather conditions. We recorded people in clear, rainy and foggy weather at different distances from the camera and in different body positions, as well as different motion speeds, maximizing the simulated conditions for detecting people in the surveillance and monitoring areas.

The choice of equipment to be used for recording also depends on the purpose of the research. Due to night conditions and recording in harsh weather conditions, an infrared thermo-vision camera will be used.

In recent years, thermal cameras have been increasingly present in surveillance systems. Concerning the spread of the areas of application of thermal-vision systems, such as the use of autonomous vehicles, automatic screening at airports, video surveillance usage, recognition of face expression for high-end security applications, monitoring body temperature for medical diagnostics, etc. [1], their price was drastically reduced, their characteristics improved, and availability increased. We choose the long-wave segment of the electromagnetic spectrum of thermal cameras (LWIR segment, wavelength 7 - 14 μm) because of its physical properties such as the ability to record objects under night conditions, in bad weather conditions, with high shooting ranges [2].

Part of the images is recorded by a telephoto lens so that smaller parts of the body such as arms and legs are visible, as well as a step cycle and other soft biometric traits that can be useful for the gait recognition task.

The rest of the paper is structured as follows. In Section II. the basics of infrared thermo-vision with emphasis on the LWIR part of the infrared spectrum is described. A list of widely used thermal imaging datasets that contain the LWIR segment is given in Section III. The process of creating a dataset, from defining scenarios, capturing images, processing videos up to the annotation of the image is described in Section IV. The paper ends with the conclusion and plan for future research.

II. Infrared Thermal Imaging Basics

Infrared (IR) thermal sensors have the capability of imaging scenes and objects based either upon the IR light reflectance or upon the IR radiation emittance. The IR radiation is an EM radiation emitted in proportion to the heat generated or reflected by an object and, therefore, IR imaging is referred to as thermal imaging. The wavelengths of IR are longer than those of visible light (Fig. 1), so IR is not visible to humans [2].

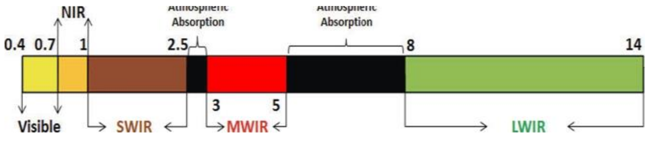


Fig. 1 Electromagnetic spectrum with illustrated IR segments wavelengths in μm [2]

Given the wavelength, the IR spectrum is divided into the following bands [2]: Near-Infrared - NIR, ranges from

0,7 to 1 µm; Short-Wave infrared - SWIR, ranges from 1 to 3 µm; Mid-Wave infrared - MWIR, ranges of 3 to 5 µm; Long-Wave infrared - LWIR, ranges from 8 to 14 µm; Very Long-Wave infrared - VLWIR, is greater than 14 µm.

The NIR and SWIR bands are sometimes referred to as "the reflected IR radiation" and MWIR and LWIR bands as "thermal IR radiation". The latter does not require an additional source of light or heat since the thermal radiation sensors can form an image of the environment or object solely by detecting the emission of thermal energy of observed objects. Since IR sensors depend on the amount of emission of the thermal energy of a recorded object, they are, unlike the visible light cameras, invariant to illuminating conditions, robust to a wide range of light variations [3, 4] and weather conditions, and so can operate in total darkness.

III. THERMAL IMAGING DATASETS

To set up an image dataset, necessary is to define a scenario, select an appropriate environment that meets the required conditions for selected research task, taking into account lighting and weather conditions, and include volunteers who will perform the defined activities. In the case of thermal imaging, the task is even more complicated because it involves shooting in night time or darkness. In our case, this task is further complicated by the need for shooting in fog and rain.

Researchers commonly create the dataset according to their goals and tasks they intend to solve [1, 5, 6]. Table 1 summarizes the thermal datasets, with a brief description of the images they contain, the IR bands that include and concerning the research in which have been used.

When using existing databases such as OTCBVS Benchmark Database [7], KAIST Multispectral Pedestrian Dataset [8], CASIA Night Gait Dataset [9] and CASIA Infrared Night Gait Dataset [10], the research goal needs to be adjusted to the available data, which is not the case when the dataset is created from the scratch. In our case, we wanted to create a dataset that will simulate the real-life situation of unauthorized movement in the monitored area, sneaking around protected objects and crossing the state border in the night time at favorable and unfavorable weather conditions. Such recordings none of the existing datasets include.

Table 1. Thermal imaging datasets for human detection

IR band	*Dataset*	*Ref.*
LWIR	OTCBVS Benchmark Database; OSU Color-Thermal Database; image resolution 240x320	[7]
LWIR	CASIA Night Gait Dataset	[9]
LWIR	CASIA Infrared Night Gait Dataset; human silhouette image is normalized to the resolution of 129×130 NLPR Database - daytime	[10]
SWIR	Controp Fox 720 thermal camera, video sequence resolution 640 x 480 pixels at distance 3000 m	[11]
LWIR	520 images captured with FLIR ThermoVision A320 InfraRed camera; resolution 320x240	[12]
LWIR	OTCBVS datasets	[13]
MWIR LWIR	OTCBVS; 360 × 240 pixels; 768 images captured with ICI 7320 thermal camera; resolution 320x240	[14]
LWIR	Video sequence 25 FPS at resolution 640x512	[15]
LWIR	OTCBVS dataset 01; OSU Thermal Pedestrian Database\00001\00001; OSU Thermal Pedestrian Database\00002\00002; Terravic Motion Infrared Database	[16]
LWIR	Tetravision image database	[17]
LWIR	ANTID dataset; Resolutions range from 320×240 to 920×480 both 8 and16 bit pixel; PNG file type. Images are recorded with various FLIR cameras and images representing a different type of objects: humans, animals, cars, static objects…	[18]
LWIR	IR Tau2 640x512, 13mm f/1.0 (HFOV 45°, VFOV 37°) FLIR BlackFly (BFS-U3-51S5C-C) 1280x1024, Computar 4-8mm f/1.4-16 megapixel lens (FOV set to match Tau2). Conditions: Day (60%) and night (40%) driving, streets, and highways, clear to overcast weather. Frame Annotation Label Totals: 10,228 total frames and 9,214 frames with bounding boxes: Person (28,151), Car (46,692), Bicycle (4,457), Dog (240), Other Vehicle (2,228); Video Annotation Label Totals: 4,224 total frames and 4,183 frames with bounding boxes.: Person (21,965), 2. Car (14,013), Bicycle (1,205), Other Vehicle (540).	[19]
LWIR	FLIR ThermoVision A-20M; 10.000 thermal images at resolution 320x240 pixels from height 1,5 m	[20]
LWIR	NEC-C200 infrared thermal camera with 320x240 pixel image resolution up-sampled to 640x480 using cubic kernel; 17.000 images in testing dataset	[21]
LWIR	MoviRobotics S.L. indoor autonomous mobile platform mSecuritTM with FLIR thermal camera	[22]
LWIR	MTIS - thermal imaging system launched on a mobile phone, resolution 64 × 62; distance: 2 - 9 m, max 29 m	[23]
SWIR	WVU Outdoor SWIR Gait (WOSG) Dataset	[24]
LWIR	1,440 walking samples collected from 8 subjects in different walking conditions at distance 2,6 m	[25]

IV. DATASET DESCRIPTION

The main goal was to create a dataset for learning a deep learning model for person detection in realistic conditions that can happen during surveillance and monitoring protected area. All scenarios are recorded with thermal cameras in night conditions and during the winter period. Recordings were done in clear weather, fog, and heavy rain.

Recorded people had to mimic people who intentionally but unauthorized enter the monitored area, therefore should move at the different distance from the camera, with different movement speeds and body positions, from crawling on all fours, hiding, hunched walking to running.

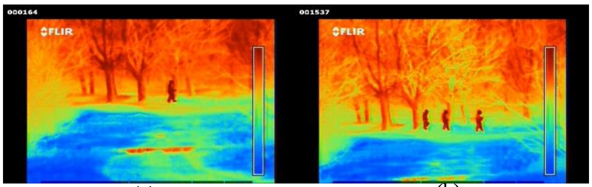
A. Dataset design criteria

The basic scenarios that should be met:

- **Clear weather:**

1. It is used to estimate the maximum (reference) distance of the camera on which the person on the record can be detected with a naked eye. The reference distance of 110 m is selected, and the recording with the standard lens is carried out at a distance of 165 m, Fig 2.

2. Primarily used for training the person detection model
3. Record people while walking, running, sneaking (hunched walking).



(a) (b)

Fig. 2 Comparison of images taken at the clear weather and the distance of 165 m: (a) standard thermal camera lens, (b) using telephoto lens

- **Dense fog weather:**
  1. Evaluation of the camera's ability to record in fog and conditions with low visibility - reference distances of 30 and 50 m
  2. It is used to prepare a test set to examine the robustness of a model for person detection: the objective is to learn a model for person detection at clear weather footage and observe its ability to be used in conditions that simulate realistic conditions in protected areas such as fog or rain.
  3. Record people while walking, running, sneaking (hunched walking).
- **Heavy rain weather:**
  1. Estimate the camera's ability to record in rainy weather and weather with high humidity - distances from 30 to 215 m
  4. It is used during the test phase to test the robustness of the model for person detection trained on clear weather images and to examine the ability to be used under real-world surveillance conditions.
  2. Record people while walking, running, sneaking (hunched walking).

B. Data collection

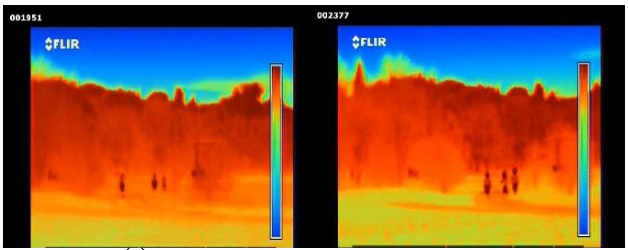
We used the FLIR ThermaCam P10 [26] LWIR thermal imaging camera for recording with a standard lens and FLIR 7 degrees Telephoto Lens (P/B series) [27] mounted on a tripod at the height of about 140 cm. The camera captures a thermal image of resolution of 320 x 240 pixels, and we used an external video recorder to enhance the quality of the recordings, which upscale images at a resolution of 1280 x 960 pixels. For the measurement of distance, we used the ViewRanger application [28] installed on the CAT S60 [29] smartphone. The CAT S60 smartphone was also equipped with a thermal camera, but we do not include images taken using this smartphone in this dataset.

1) Clear weather

The recording was carried out on a meadow with a small forest, so it was possible to simulate the conditions of hiding people to determine the camera's potential in terms of "whether and how much it can see through trees or bushes". The shooting location was approximately 180 m in length, and the air temperature was about 2 ° C under night conditions, with good visibility and without affecting atmospheric conditions.

Using a standard lens, the person is 110 m away, recorded at a normal pace individually and in a group (three people).

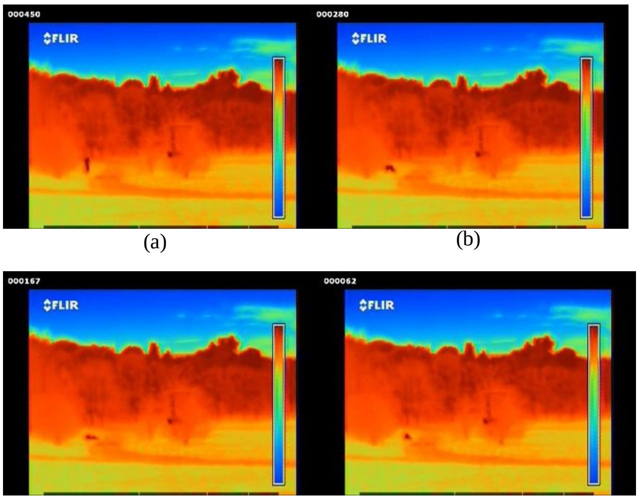
We recorded one and three persons so that they walked away from the camera (50 to 110 m) and then walked over the camera's field of view (at a distance of 110 and 165 m). In this scenario, people have changed body positions, so they were recorded during upright walking at the distances of 50 to 165 meters while the reference distance was 110 m. The recording was done using the standard lens of FLIR Thermal Cam P10 and using the telephoto lens. Using a standard lens, the person was recorded 110 meters away, walking at a normal pace, individually and in a group (three people). At the same distance, people were recorded using the standard lenses during running, as well as using a telephoto lens. Fig. 3 shows three normal walking persons and presents different image quality between images taken using standard (a) and telephoto lens (b).



(a) (b)

Fig. 3 Comparison of images taken at the clear weather and the distance of 110 m: (a) standard camera lens, (b) telephoto lens

Also, we recorded a person in clear weather at a distance of 110 meters with changed body positions - from normal to walk on all fours, crawling and lying on the ground (Fig. 4).



(a) (b)

(c) (d)

Fig. 4 Comparison of images taken at the clear weather and the distance of 165 m using standard lenses – one person – changing positions: (a) normal walking, (b) fourleg walking; (c) lying on the ground - left side of the person; (d) lying on the ground - head in the direction of the camera

Then the shooting was continued with a standard lens so that people walk away from the distance of 110 m to the distance of 165 m.

At this distance (165 m), people were recorded while walking over the camera's view field at upright normal speed walking and upright running and the hunched normal speed walk and hunched running, Fig. 4. At this distance, we used only telephoto lens since the images recorded using a standard lens were low quality, and it was almost impossible to detect and recognize a person with a naked eye.

2) *Foggy weather*

Due to the high density of water particles in the air in fog conditions, LWIR radiation is highly dispersed, so the visibility for the thermo-visual camera is significantly reduced compared to other atmospheric conditions. This fog condition was the most demanding scenario as it is stated in the reviewed literature [29, 30].

Although it was primarily intended to repeat shooting scenarios as well as at clear weather, the same was not possible. Concerning available equipment and extremely dense fog, people at distances of over 50 meters were not visible at all. After the initial visual inspection of the recorded images, we selected a distance of 0 to 30 m and at 50 m, using only the telephoto lens, Fig. 5.

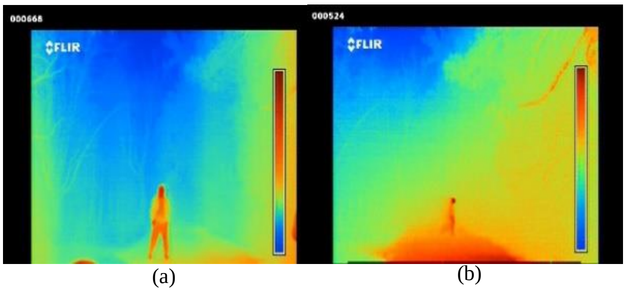


Fig. 5 Comparison of images taken at the foggy weather and the distance of: (a) telephoto lens at 30 m, (b) telephoto lens at 50 m

We recorded up to three persons that changed the way of movement and position of the body. The recording was done in the part of the woods on the asphalt road, and the people moved away from the camera and crossed the camera view field with a normal upright walk, hunched and running. The air temperature was about 2 ° C, and the visibility was significantly reduced due to dense fog (less than five meters of visibility).

3) *Rainy weather*

For shooting over the rainy weather, we choose the location that provided the ability to shoot at distances of over 165 meters. We have made a recording using standard and telephoto lenses at distances of 30, 70, 110, 140, 170, 180, and 215 m. Two persons were recorded at all distances - individually and together at normal walking and running in upright and hunched positions, Fig. 6. Shooting at a distance of 215 m was possible using the telephoto lens only.

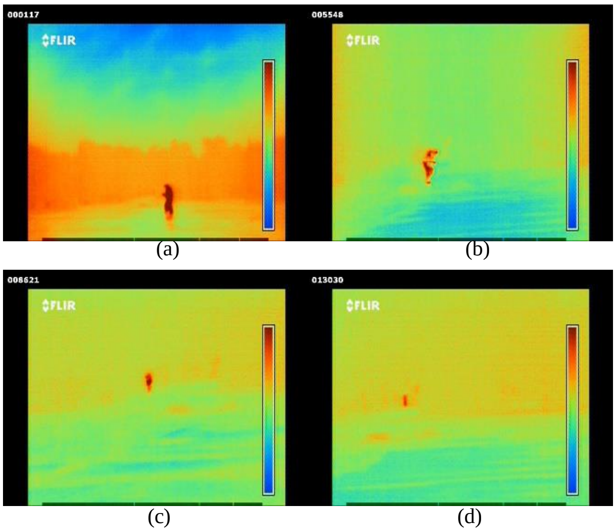


Fig. 6 Comparison of images taken in heavy rain weather and the distance of: (a) telephoto lens at 30 m, (b) telephoto lens at 70 m (hunched); (c) telephoto at 140 m; (d) telephoto at 215 m

C. *Data preprocessing*

After recording and video pre-processing, we got about 20 minutes of videos recorded on the clear weather, about 13 minutes on fog and about 15 minutes on rainy weather. The resulting video material needed additional processing to be prepared for training, so we cut all the videos into smaller parts that best fit the pre-defined scenarios. At this stage, we used the VSDC Video Editor's free software tool to cut video materials [31].

After cutting the videos into smaller parts, it was necessary to extract frames from all the videos we have got. We performed the image extraction using MATLAB and corresponding scripts for extracting video frames. The result was a total of 11,900 frames in 1280x960 resolution (for clear weather at all distances, all body positions, and motion speeds). For fog, we obtained a total of 4,905 images for both recording distances and all body positions. For rainy weather, there are 7,030 frames for all distances and all body positions.

For supervised learning of a model for person detection in thermal images, there should be available annotated frames with tagged persons in the learning set as well as in the test set that will serve as ground truth. That is why the next phase was the process of frame annotation and tagging of all persons present on each image. As the manual annotation and objects tagging is long lasting and tedious work, we tried to choose representative images that will be manually annotated. Considering that human walking or running is a repetitive activity, we have selected the frames at different distances and angles from the camera, that would maximally show the differences in body positions. Therefore, several scenarios for choosing candidate images have been applied. The first step was the visual inspection of all the frames, and then the corresponding frames were selected according to the defined criterion (maximum difference in the motion and position of the body, as well as visibility of the recorded persons). After this phase, the initial set selected for manual annotation and person tagging was reduced to a total of 7,412 frames. In a formed set, 3,174 frames were obtained for clear weather, 1,460 images for fog, and 2,778 images for rain.

D. *Dataset annotations*

Annotation and tagging of objects in images are one of the most time-consuming procedures in the preparation of the dataset. There are several free tools available for the image annotation, such as the VGG Annotator [32], which runs from the Internet browser and has several useful features, then Labelbox [33], LabelMe [34], Annotorious [35].

The choice of image annotation tool depends on the method to be used for model learning, so in our case, we have chosen the tool available on the GitHub repository called Boobs - Yolo BBox Annotation Tool [36]. Fig. 7 shows the user interface of tool that we used for annotation, and Fig. 8 shows the same tool when images are loaded and objects (persons) annotated with bounding boxes.

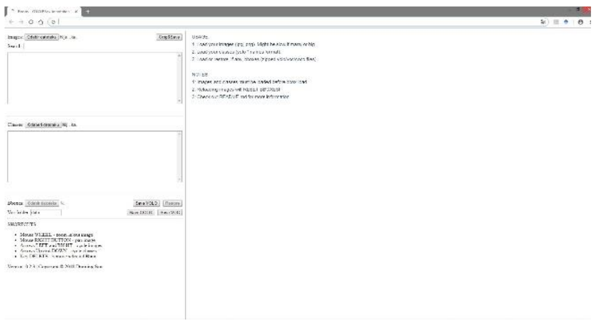


Fig. 7 Boobs – YOLO BBox Annotation Tool user interface

Fig. 8 Boobs – YOLO BBox Annotation Tool user interface with annotated image

This tool runs within the web browser and supports all popular browsers (Firefox, Chrome, Safari, Opera). The key reason for using this tool is its option to simultaneously store annotations into the three most popular formats for machine learning - YOLO [37], VOC [38], and MSCOCO [39]. This saves time for later phases of model learning because no subsequent conversion is required for a specific format.

In our case, we saved the image annotations in all three formats for the later research, but for now, we used only YOLO format because we have already used YOLO detector [40] for person detection on video materials and we got state-of-the-art detection results [41, 42]. In the YOLO format image annotation is saved as a .txt file for each image so at the end of annotation the number of txt files is the same as the number of annotated images. Annotation in YOLO format comprises of five labels in a line for each annotated object in the image, Fig. 9. The first value is an integer number representing an object class, so if there are several object classes in the image, they are sequentially numbered starting from 0. The second value in the line represents the center of the selected object on the X-axis [*object center in X*]. The third value represents the center of the object on the Y-axis [*object center in Y*]. The fourth value represents the width of the object on X [*object width in X*], and the fifth value represents the width of the object at Y [*object width in Y*] [43].

If there are multiple tagged objects in the image, each

```
0 0.703515625 0.6541666666666667 0.03203125 0.13125
0 0.5921875 0.6578125 0.034375 0.11770833333333333
0 0.523828125 0.6520833333333333 0.02578125 0.122916666666666
```

Fig. 9 YOLO format annotations of three annotated objects shown in the Fig. 8.

is described in a new line within the same txt file corresponding to the image on which the objects are located. For example, Fig. 9 is an example of the three annotated objects shown in Fig. 8. Since all objects represent people and belong to the same class, each annotated object is numbered with 0, and the txt file has three lines that begin with zero.

After the annotation of objects or persons in the image is finished, the next step is to train the machine learning model for person detection. For the training and testing purposes, we used the YOLO detector [40] within the Darknet framework [44], available in the GitHub repository [45]. The preliminary results have shown that the prepared set is worthwhile to continue researching and learning a model for person detection on thermographic images [46, 47].

V. Conclusion

This paper presents a new dataset of thermal images designed for learning a model of human detection. A particular concern was paid to simulate realistic conditions in the protected area at different weather conditions. The shooting of people made under harsh weather conditions - by dense fog and heavy rain, is an additional feature of this set of data.

Recorded persons moved in the way that is common for people who intentionally but illegally enter the supervised area: at a different distance from the camera and different positions, different speeds of movement and in different movement types, from creeping, sneaking to walking and running.

Motion captured by a telephoto lens provides enough details to be used for recognizing individual activities (running, walking, hunched walking, hunched running, etc.) and for the gait recognition and identifying people by way of walking.

The formed dataset will be used to train a person detection model on thermal images using deep neural network such as YOLO detector. The plan is to use formed dataset for gait recognition and recognition of movement types in adverse weather conditions.

ACKNOWLEDGMENTS

This research was fully supported by Croatian Science Foundation under the project IP-2016-06-8345 “Automatic recognition of actions and activities in multimedia content from the sports domain” (RAASS).

REFERENCE

[1] M. Kristo, M. Ivasic-Kos. An Overview of Thermal Face Recognition Methods // Proceedings of 2018 41st International Convention on Information and Communication Technology, Rijeka: MIPRO, 2018. 174-200

[2] Chris Solomon and Toby Breckon; Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab; John Wiley & Sons, Ltd. 2011.

[3] Bourlai, Thirimachos, and Bojan Cukic. "Multi-spectral face recognition: identification of people in difficult environments." Intelligence and Security Informatics (ISI), 2012 IEEE International Conference on. IEEE, 2012.

[4] Zheng Wu, Nathan Fuller, Diane Theriault, Margrit Betke, "A Thermal Infrared Video Benchmark for Visual Analysis", in Proceeding of 10th IEEE Workshop on Perception Beyond the Visible Spectrum (PBVS), Columbus, Ohio, USA, 2014.

[5] M. Buric, M. Pobar, M. Ivasic-Kos. An overview of action recognition in videos, 2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO) /Rijeka: IEEE, 2017. 1310-1315

[6] M. Ivasic-Kos, M. Pobar. Building a labeled dataset for recognition of handball actions using mask R-CNN and STIPS // Proceedings of 2018 7th European Workshop on Visual Information Processing (EUVIP)/ Tampere, Finska: IEEE, 2018. 1-6

[7] J. Li, W. Gong, "Real Time Pedestrian Tracking using Thermal Infrared Imagery", *JCP, Vol. 5, No. 10*, pp. 1606 - 1613, 2010.

[8] Hwang, Soonmin, et al. "Multispectral pedestrian detection: Benchmark dataset and baseline." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015.

[9] T. Bourlai, N. Kalka, B. Čukić, et al., "Ascertaining human identity in night environments", *Book Chapter, Distributed Video Sensor Networks, London, England, Springer*, pp. 451-467, 2011.

[10] D. Tan, et al., "Efficient Night Gait Recognition Based on Template Matching", in *Pattern Recognition, ICPR 2006, 1*

[11] E. Chen, O. Haik, Y. Yitzhaky, "Classification of moving objects in the atmospherically degraded video", *Optical Engineering, vol. 51, no. 10*, p. 101710, 2012.]

[12] E. T. Gilmore, C. Ugbome, C. Kim, "An IR-based Pedestrian Detection System Implemented with Matlab-Equipped Laptop and Low-Cost Microcontroller", *International Journal of Computer Science and Information Technology, Vol. 3, No. 5*, pp. 79 - 87, 2011.]

[13] R. Miezianko, D. Pokrajac, "People detection in low-resolution infrared videos", in IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Honeywell Labs, Minneapolis, MN, pp. 1 - 6, 2008.]

[14] E. Jeon et al., "Human Detection Based on the Generation of a Background Image by Using a Far-Infrared Light Camera", *Sensors, Vol. 15, No. 3*, pp. 6763 - 6788, 2015.]

[15] F. Coutts, S. Marshall, P. Murray, "Human detection and tracking through temporal feature recognition", in *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European, Lisbon*, pp. 2180 - 2184, 2014.]

[16] U. Dias, M. Rane, "Motion Based Object Detection And Classification For Night Surveillance", *ICTACT Journal on Image and Video Processing, Vol. 3, No. 2*, pp. 518 - 521, 2012

[17] B. Besbes, et al., "Pedestrian Detection in Far-Infrared Daytime Images Using a Hierarchical Codebook of SURF", *Sensors, Vol. 15, No. 4*, pp. 8570 - 8594, 2015.]

[18] Berg, Amanda, Jörgen Ahlberg, and Michael Felsberg. "A thermal infrared dataset for evaluation of short-term tracking methods." Swedish Symposium on Image Analysis. 2015.

[19] https://www.flir.eu/oem/adas/adas-dataset-form/

[20] W. Wong, et al., "Omnidirectional Thermal Imaging Surveillance System Featuring Trespasser and Faint Detection", *International Journal of Image Processing (IJIP), Vol. 6, No. 6*, pp. 518 - 538, 2011.

[21] I. Riaz, J. Piao, H. Shin, "Human detection by using centrist features for thermal images", *IADIS International Journal on Computer Science and Information Systems*, Vol. 8, No. 2, pp. 1 - 11, 2013

[22] [J. Castillo, et al., "Segmenting Humans from Mobile Thermal Infrared Imagery", in *3rd International Work-Conference on The Interplay Between Natural and Artificial Computation: Part II: Bioinspired Applications in Artificial and Natural Computation (IWINAC '09)*, Springer-Verlag, Berlin, Heidelberg, pp. 334 - 343, 2009.

[23] F. Lee, F. Chen, J. Liu, "Infrared Thermal Imaging System on a Mobile Phone", *Sensors, Vol. 15, No. 5*, pp. 10166-10179, 2015.

[24] B. DeCann, A. Ross, J. Dawson, "Investigating gait recognition in the short-wave infrared (SWIR) spectrum: dataset and challenges", *in SPIE 8712, Biometric and Surveillance Technology for Human and Activity Identification X*, 2013.

[25] J. Yun, S. Lee, "Human Movement Detection and Identification Using Pyroelectric Infrared Sensors", *Sensors, Vol. 14, No. 5*, pp. 8057-8081, 2014.

[26] https://www.termogram.com/pdf/therma_cam_p_series/thermacam_p10/P10_datasheet.pdf

[27] https://www.pass-thermal.co.uk/flir-131-mm-7-degree-telephoto-p-b-series-lens

[28] https://play.google.com/store/apps/details?id=com.augmentra.viewranger.android&hl=hr

[29] Gim, LEE Cheow, EE Kok Tiong, and HENG Yinghui Elizabeth. "Performance challenges for high-resolution imaging sensors for surveillance in tropical environment.", DSTA HORIZONS, 2015, pp. 80 – 88.

[30] Holst, Gerald C. *Common sense approach to thermal imaging*. Washington: SPIE Optical Engineering Press, 2000.

[31] http://www.videosoftdev.com/free-video-editor

[32] http://www.robots.ox.ac.uk/~vgg/software/via/

[33] https://www.labelbox.com/

[34] http://labelme.csail.mit.edu/Release3.0/

[35] https://annotorious.github.io/

[36] https://github.com/drainingsun/boobs

[37] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016., https://pjreddie.com/darknet/yolo/

[38] http://host.robots.ox.ac.uk/pascal/VOC/

[39] http://cocodataset.org/#home

[40] J. Redmon and A. Farhadi, “Yolov3: An incremental improvement,” arXiv preprint arXiv:1804.02767, 2018.

[41] M. Buric, M. Pobar, and M. Ivasic-Kos, “Adapting YOLO network for Ball and Player Detection”, In ICPRAM 2019.

[42] Buric, Matija; Pobar, Miran; Ivasic-Kos, Marina. Ball detection using YOLO and Mask R-CNN // Proceedings of the 5th Annual Conf. on Computational Science & Computational Intelligence (CSCI’18) / Las Vegas, USA: 2018.

[43] *https://timebutt.github.io/static/how-to-train-yolov2-to-detect-custom-objects/*

[44] https://pjreddie.com/darknet/yolo/

[45] https://github.com/pjreddie/darknet

[46] M. Kristo, M. Pobar and M. Ivasic-Kos, Human detection in thermal imaging using YOLO, ICCTA 2019, Istanbul (In press)

[47] M. Ivasic-Kos, M. Kristo, M. Pobar, Person Detection in thermal videos using YOLO, IntelliSys 2019, London (In press)