Deep Learning Based Object Recognition in Real Time Images Using Thermal Imaging System

Rohini Goel [a, 1] , Avinash Sharma [a], and Rajiv Kapoor [b]

[a] *CSE Department, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala, Haryana, India*
[b]*Professor, Department of ECE, Delhi Technological University Delhi, India*

Abstract. An efficient driver assistance system is essential to avoid mishaps. The collision between the vehicles and objects before vehicle is the one of the principle reason of mishaps that outcomes in terms of diminished safety and higher monetary loss. Researchers are interminably attempting to upgrade the safety means for diminishing the mishap rates. This paper proposes an accurate and proficient technique for identifying objects in front of vehicles utilizing thermal imaging framework. For this purpose, image dataset is obtained with the help of a night vision IR camera. This strategy presents deep network based procedure for recognition of objects in thermal images. The deep network gives the model understanding of real world objects and empowers the object recognition. The real time thermal image database is utilized for the training and validation of deep network. In this work, Faster R-CNN is used to adequately identify objects in real time thermal images. This work can be an incredible help for driver assistance framework. The outcomes exhibits that the proposed work assists to boost public safety with good accuracy.

Keywords. Thermal Imaging, HSV color space Segmentation, Faster Region-Convolution Neural Network (Faster-RCNN), Region Proposal Network, Intersection over Union (IoU).

1. Introduction

Driver assistance frameworks are probably the important applications object recognition. Individual in front of vehicles face a high danger of injury or demise during the night or because of restricted perceivability. Having a warning framework [1] to hinder interlopers is pivotal for both the safety of public and to avoid monetary losses. To forestall the mishaps, several intelligent frameworks have been created utilizing close circuit TV cameras combined with machine learning innovations [2]. However, most frameworks [3] work along with database of images captured in proper light, while the detection performance in front of vehicle in the restricted perceivability [4] is extensively more awful. In thermal imaging system, the radiations as infrared light are detected by the device called thermal camera to capture the thermal images. To recognize individuals in the restricted perceivability, several investigations have

---

[1] Rohini Goel, CSE Department, Maharishi Markandeshwar (Deemed to be University), Mullana

Email:27rohinisharma@gmail.com

adopted machine learning based methods [5] utilizing thermal IR images. The machine learning based approaches can be distinguished on the basis of feature descriptors & classifiers such as HOG[6], SIFT[7], INERTIA[8], SVM [9][10] or Adaboost [11].
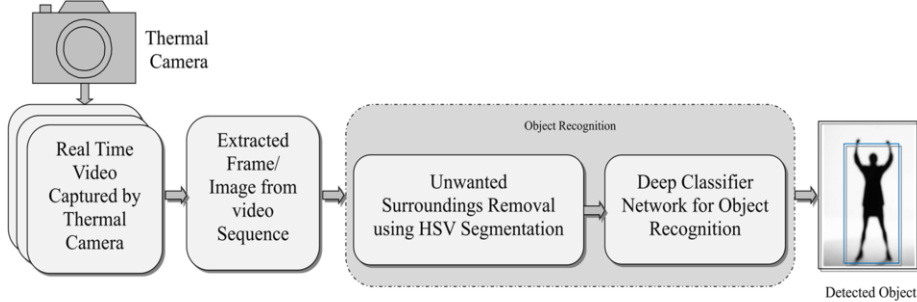

Figure 1. Basic block diagram of the proposed method used to recognize the object.

In some human recognition applications haar feature based classifiers [12][13] is used for thermal IR images. Although the machine learning methods [14] performs satisfactorily but for images having noisy/ ambiguoused features[15] suffer from degraded performance. To handle this problem convolutional neural network [16] [17] based deep learning methods [18] have been used for object recognition in thermal images. The recognition performance for thermal images can be upgraded using deep network based methodologies [19]. It has observed that human detection methods based on CNN [20][21] give better recognition results as compared to methods based on SVM. The selection of proper training data affects the recognition performance of deep network methods[22]. Several investigations have created thermal image datasets, e.g. CVC-14 [24], LSI [25] and KAIST [23] for person recognition in limited visibility. The work presents a precise as well as proficient strategy to recognize objects in real time thermal images. In this paper, a system is intended to segment the moving objects from the image using color HSV segmentation. Finally, the Faster-RCNN is utilized to distinguish the objects in the real time thermal images.This paper is organized in various subsections as: section II presents the proposed procedure in detail including collection steps using thermal imaging, HSV segmentation and deep network [26] (Faster R-CNN) for recognition of object. The outcomes and discussions are explained in section III. Finally, the conclusion of paper is discussed in the section IV.

2. Proposed method

The proposed method is used to recognize the object on it from scene sequence captured by the thermal camera [27]. The basic block diagram of the proposed work is illustrated in figure 1. The "VINCENT HD and Night vision IR camera" as shown in figure 2 acquires the video sequence with the resolution of 720p in night vision IR camera mode. The frames are extracted from the video sequences for the further processing. The various steps of the proposed approach for the recognition of the objects are illustrated in algorithm 1 and discussed in the following subsections.

2.1   Object Recognition

For the real time identification of the objects in the thermal images extracted from the video sequence by the "VINCENT HD and Night vision IR camera: Initially the noise (unwanted surroundings) is removed from the image and then objects are identified using Faster R-CNN [28] network.

Figure 2. The "VINCENT HD and Night vision IR camera" used for the thermal imaging

- Noise (unwanted surrounding) Removal Using HSV Segmentation

The unwanted surroundings of the object are considered as noise. In the thermal video frames the living objects appear brighter as they emit greater infrared radiations than other. Hence HSV segmentation [29] is applied to remove the unwanted surroundings. The image is converted to HSV color space form RGB color space as given below:

$$H = \begin{cases} 1 + (G - B)/\Delta \ (if \ R = max(R, G, B)) \\ 3 + (B - R)/\Delta \ (if \ G = max(R, G, B)) \\ 5 + (R - G)/\Delta \ (if \ B = max(R, G, B)) \end{cases}$$
$$S = [max(R, G, B) - min(R, G, B)]/\max(R, G, B)$$
$$V = max(R, G, B) \tag{1}$$

Where $\Delta = [max(R, G, B) - min(R, G, B)]$ and (R, G, B) is the RGB color space at pixel location *(x,y)* in the image. At that point, the lower limit l_T and upper edge h_T range of the white color is characterized in HSV color space. The thresholding of the HSV color space is done to separate the white color in the image that addresses living objects.

$$I_{Seg}(x, y) = \begin{cases} 1 & (if \ l_T \leq I_{HSV}(x, y) \leq h_T) \\ 0 & else \end{cases} \tag{2}$$

Where, '1' and '0' addresses white and black tones respectively, I$_{HSV}$(x,y) input image in HSV color space and I$_{Seg}$(x,y) is output image after noise removal. Finally, the objects are segmented and the surrounding objects is taken out in the output image.

- Deep Network for Object Recognition

Faster R-CNN [28] is utilized to recognize objects in the real time thermal images. The Region Proposal Network (RPN) utilizes 'attention' system to tell the Fast R-CNN [30] detector network where to look. In this work, VGG-16 network model [31] is utilized in Faster R-CNN model [28]. In default arrangement, there are anchors at position of image having 3 scales and 3 ratios. The anchors with an IOU more than 0.7 are classified as foreground and the anchors don't cover any ground truth object (IOU under 0.3) are classified as background. The loss function F_{Loss} to be limited is given as:

$$F_{Loss}(\{m_v\}, \{n_v\}) = \frac{1}{C}\sum_v L_c(m_v, m_v^*) + \gamma \frac{1}{R}\sum_v m_v^* L_r(n_v, n_v^*) \tag{3}$$

Where, m_v is the predicted anchor probability of anchor v being an object. Vector n_v signifies the parameterized coordinates of predicted bouncing box. L_r and L_c are the

regression and classification loss respectively. For classification, Cross Entropy Loss is utilized and for Bounding Box Regression smooth L1 loss is utilized. The normalization parameters C and R are mini batch size (i.e. C=256) and anchor locations (i.e. R=2400). The term 'γ' is the weight balancing parameter set to 10. The Fast R-CNN network utilizes cross entropy loss in between numerous object classes with end to end back propagation and Stochastic Gradient Descent with momentum having value of 0.9.

3.   Result and discussion

For this work, VINCENT HD and Night vision IR camera is used to record the real time thermal video that has an approximate range of 1.5 km and resolution up to 720p but the video is recorded at a resolution of 360*450. The camera is set at channel number 101 assigned to night vision IR camera mode and IP address 192.168.1.65 is assigned to access camera remotely. The 749 frames are used evaluate the performance of proposed work as shown in Fig. 3. Faster R-CNN implementation is performed with configuration having Intel(R) Core(TM) i5-7200U CPU @ 2.50 GHz, 2701 MHz.

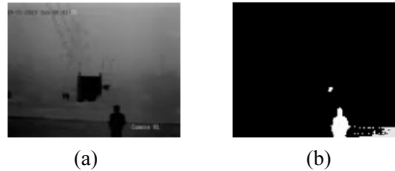**Figure 3.** Extracted frames from the video sequence captured by thermal Night vision IR camera.

(a)   (b)

Figure 4 (a). Input frame **(b)** Surrounding removal performed at threshold values l_T =0.85 and h_T =1.

3.1 Object Recognition

HSV segmentation is used for the removal of the unwanted surroundings. As already discussed, living objects appear more white than other objects in thermal images because the living objects radiates more infrared rays than other objects. For the masking of white color space two thresholds (l_T and h_T) are used. As the obstacles appear whiter than others, so the upper threshold h_T is set to '1' but the lower threshold l_T is varied between 0.0-1.0 to get the optimum result. These results are recorded for HSV segmentation at threshold l_T=0.85, h_T =1 as shown in figure 4(b). Finally the Faster R-CNN network is applied to the output of the HSV segmentation to recognize objects in the thermal images. The performance of the network is revamped to enhance the accuracy by optimizing the parameter IoU threshold. The initial results of the proposed algorithm are obtained using threshold value 0.7 as shown in figure 5(c) but at the shorter distance from the camera. The lower IoU values provides better results in terms of timely identification but faced lower accuracy. The IoU threshold value '0.5' is used which timely recognize the objects with good accuracy as shown in figure 5(g). The final results for the IoU=0.7 and 0.5 are shown in Fig. 5(d) and 5(h) respectively.
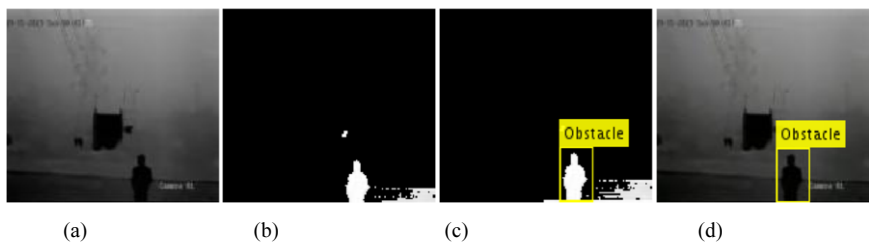
| (a) | (b) | (c) | (d) |

Figure 5 (a). Input frame **(b)** HSV segmented image after Surrounding removal **(c)** Faster R-CNN result with at IoU=0.7 **(d)**The final results of object recognition with Faster R-CNN at IoU=0.7

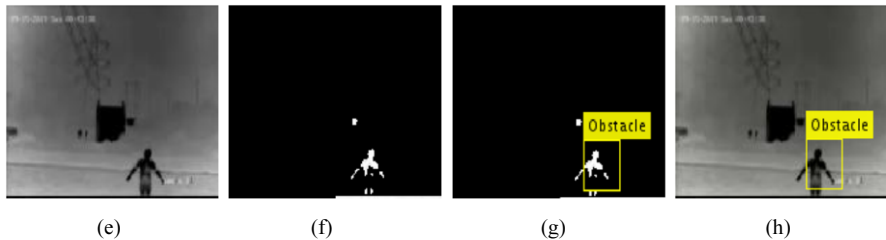| (e) | (f) | (g) | (h) |

Figure 5 (e). Input frame **(f)** HSV segmented image after Surrounding removal **(g)** Faster R-CNN result with at IoU=0.5 **(h)**The final results of object recognition with Faster R-CNN at IoU=0.5

The proposed work performance is estimated with the help of parameters like accuracy, precision and recall. These parameters are evaluated for different values of the Intersection over Union (IoU) as shown in Table 1. The visual results and Table 1 concludes that the proposed system has better performance at IoU=0.5. Figure 6 illustrates the graphical comparison which also shows that the performance of this with accuracy of approximately 84.7%.

Table 1. The proposed method performance for different IoU

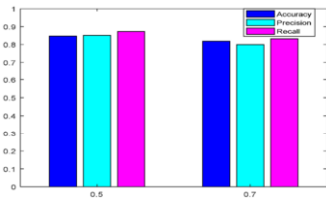| IoU	Accuracy	Precision	Recall
0.5	0.847	0.851	0.873
0.7	0.818	0.799	0.832

Figure 6. Performance at different values of IoU

4.   Conclusion and future scope

The proposed strategy turns out sufficiently for the object recognition real time thermal images. This paper introduced a novel methodology dependent on deep learning network to distinguish objects in images caught utilizing thermal imaging framework. This work can be robust technique to build up an early warning framework to forestall mishaps for the public safety upgrade. This strategy will be likewise cost effective as it doesn't need any huge change in the architecture as well as will decrease the monetaryin terms of accidental compensation. The framework will be extended for recognition of other object classes near the vehicle to improve the framework.

References

[1] Y. Wu., X. He, and T.Q. Nguyen, Moving Object Detection with a Freely Moving Camera via Background Motion Subtraction, *IEEE Trans. on Circuits and Systems for Video Technology* **27**(2), (2017), 236–248.

[2] R. Goel, A. Sharma, and R. Kapoor, State-of-the-Art Object Recognition Techniques: A Comparative Study. *Advances in Intelligent Systems and Computing (Springer)* (2020).

[3] J. Lee, J. Park, I. Kim, D.Y. Kang, Application of vision-based safety warning system to haeundae beach, *Korea. J. Coast. Res.* **91**(2019), 216–220.

[4] N. Hautiere, J. Tarel, J. Lavenant, and D. Aubert, Automatic fog detection and estimation of visibility distance through use of an onboard camera. *Machine Vis. and Applications (Springer)*, **17**(2006), 8–20.

[5] F. Suard, A. Rakotomamonjy, A. Bensrhair, A. Broggi, Pedestrian detection using infrared images and histograms of oriented gradients. *In Proceedings of the 2006 IEEE Intelligent Vehicle Symposium, Tokyo, Japan*, (2006), 206–212.

[6] N. Dalal, and B. Triggs, Histograms of oriented gradients for human detection. *In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA*, (2005), 886–893.

[7] K. Mikolajczyk, C. Schmid, A. Zisserman, Human detection based on a probabilistic assembly of robust part detectors. *In Proceedings of the 8th European Conference on Computer Vision, Prague, Czech Republic*, (2004), 69–82.

[8] Y. Fang, K. Yamada, Y. Ninomiya, B.K.P. Horn, I. Masaki, A shape-independent method for pedestrian detection with far-infrared images. *IEEE Trans. Veh. Technol.* **53**(2004), 1679–1697.

[9] H.K. Ragb, V.K. Asari, T.H. Aspiras, Human detection in infrared imagery using intensity distribution, gradient and texture features. *In Proceedings of the Mobile Multimedia/Image Processing, Security, and Applications 2018, Orlando, FL, USA*, (2018).

[10] F. Xu, X. Liu, K. FujiMura, Pedestrian detection and tracking with night vision. *IEEE Trans. Intell. Transp. Syst.* **6**(2005), 63–71.

[11] L. Zhang, B. Wu, R. Nevatia, Pedestrian detection in infrared images based on local shape features. *In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA*, (2007), 1–8.

[12] C.H. Setjo, B. Achmad, F. Djoko, Thermal image human detection using Haar-cascade classifier. *In Proceedings of the 7th International Annual Engineering Seminar (InAES), Indonesia*, (2017), 1–6.

[13] A. Neves, and R. Ribeiro, Algorithms for face detection on infrared thermal images. *Int. J. Adv. Softw.* **10**(2018), 499–512.

[14] S. Guennouni, A. Ahaitouf, A. Mansouri, A comparative study of multiple object detection using haar-like feature selection and local binary patterns in several platforms. *Model. Simul. Eng.*, **8**(2015).

[15] M. Hasan, S. Ullah, M.J. Khan, K. Khurshid, Comparative analysis of SVM, ANN and CNN for classifying vegetation species using hyperspectral thermal infrared data. *In Proceedings of the International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, Enschede, The Netherlands*, (2019), 1861–1868.

[16] O. Yardimci, B.C. Ayyıldız, Comparison of SVM and CNN classification methods for infrared target recognition. *In Proceedings of the Automatic Target Recognition XXVIII, Orlando, FL, USA*, (2018).

[17] J.H. Kim, H.G. Hong, K.R. Park, CNN-based human detection in nighttime images using visible light camera sensors. *Sensors*, **17**(2017).

[18] H. Qu, L. Zhang, X. Wu, X. He, X. Hu, X. Wen, Multiscale object detection in infrared streetscape images based on deep learning and instance level data augmentation. *Appl. Sci.*, **9**(2019).

[19] B. Khalid, A.M. Khan, M.U. Akram, S. Batool, Person detection by fusion of visible and thermal images using convolutional neural network. *In Proceedings of the 2019 2nd International Conference on Communication, Computing and Digital systems (C-CODE), Islamabad, Pakistan*, (2019).

[20] C. Herrmann, M. Ruf, J. Beyerer, CNN-based thermal infrared person detection by domain adaptation. *In Proceedings of the Autonomous Systems: Sensors, Vehicles, Security and the Internet of Everything, Orlando, FL, USA*, (2018).

[21] J. Liu, S. Zhang, S. Wang, D.N. Metaxas, Multispectral deep neural networks for pedestrian detection. *arXiv* (2016).

[22] A.D'Acremont, R. Fablet, A.Baussard, G. Quin, CNN-based target recognition and identification for IR imaging in defense systems. *Sensors*, **19**(2019).

[23] S. Hwang, J. Park, N. Kim, Y. Choi, I.S. Kweon, Multispectral pedestrian detection: Benchmark dataset and baseline. *In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), Boston, MA, USA*, (2015), 1037–1045.

[24] A. Gonzalez, Z. Fang, Y. Socarras, J. Serrat, D. Vazquez, J. Xu, A.M. Lopez, Pedestrian detection at day/night time with visible and FIR cameras: A comparison. *Sensors*, **16**(2016).

[25] A. Khellal, H. Ma, Q. Fei, Pedestrian classification and detection in far infrared images. *In Proceedings of the International Conference on Intelligent Robotics and Applications, UK*, (2015), 511–522.

[26] R. Goel, A. Sharma, R. Kapoor, Object Recognition Using Deep Learning. *Journal of Computational and Theoretical Nanoscience*, **16**(2019).

[27] S. Singh, A. Khosla, R. Kapoor, Object Tracking with a Novel Visual- Thermal Sensor Fusion Method in Template Matching. *International Journal of Image, Graphics & Signal Processing*, **7** (2019), 39-47.

[28] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans.on Pattern Analysis and Machine Intelligence*, **39**(2017), 1137–1149.

[29] D. D. Burdescu, M. Brezovan, E. Ganea, L. Stanescu, A New Method for Segmentation of Images Represented in a HSV Color Space. *Proceedings of International Conference on Advanced Concepts for Intelligent Vision Systems (Springer)*, (2009), 606-617.

[30] R. Girshick, Fast R-CNN. *Proceedings of IEEE International Conference on Computer Vision*, (2015).

[31] S. Tammina, Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images. *International Journal of Scientific and Research Publications*, **9**(2019).