مجلة

كلية التـــراث الجامعة

مجلة علمية محكمة

متعددة التخصصات نصف سنوية

العدد السابع والثلاثون

15 حزيران 2023

ISSN 2074-5621

رقم الايداع في دار الكتب والوثائق 719 لسنة 2011

مجلة كلية التراث الجامعة معترف بها من قبل وزارة التعليم العالي والبحث العلمي بكتابها المرقم (ب 3059/4) والمؤرخ في (2014/ 4/7)

Review of pedestrian detection of low light based on fusion of multispectral image

Yusra T. Nasser **Rafid A. Haleot**

College of Education/ Mustansiriyah University

Abstract:

Object detection is a basic problem of machine vision and deep learning, Pedestrian detection is a specific application of the object detection problem, and it has become one of the research hotspots in recent years. In the past many techniques have been proposed to Pedestrian detection is a specific application of object detection. However, the variations in images such as body attire and pose, occlusion, different illumination parameters in different scenarios and clutter present in the background poses challenges in attaining high accuracy. This paper provides a review of some well-known existing techniques for pedestrian detection in a multi-spectral method via image fusion and convolutional neural networks.

Keywords: Pedestrian detection**,** multispectral image, Infrared and visual image.

1. Introduction

Pedestrian detection is a critical task in many applications that rely on visual information such as self-driven vehicles, aerial surveillance, rescue and so on. Though significant progress has been made in the computer vision community, pedestrian detection is still far from being solved in challenging circumstances such as nighttime, occlusion, low resolution and so on. Such defects make challenging the deployment of vision-based methods in applications requiring around-the-clock operations, e.g., autonomous driving[1]. Pedestrian detection, as the first and most fundamental step in many real-world tasks, e.g., human behavior analysis[2], gait recognition, intelligent video surveillance[3] and automatic driving, has attracted massive attention in the last decade[4]. Compared with other types of object detection, pedestrian detection puts forward stricter requirements on accuracy and real time performance, In recent years, large quantities of reviews of general object detection have been published [5].

Although numerous efforts have been made for this problem and significant improvement has been achieved in recent years [6],there still exists an insurmountable gap between current machine intelligence and human perception ability on pedestrian detection [7].

Cross-spectral fusion of visible light and infrared thermal images has become a research focus for all-day pedestrian detection since multi-modal information is intuitively considered to be complementary[8].

Most of them are confined to detecting pedestrians during daytime using visible (VI) cameras. However, the performance of VI cameras depends on good illumination conditions and can be affected when illumination is poor[9]. Recently, multispectral detectors that employ a fusion

of infrared (IR) and visible cameras have been developed to achieve robust and reliable pedestrian detection in adverse illumination circumstances [10].

Thermal and visible cameras have different imaging characteristics under different conditions, visible cameras provide precise visual details (such as color and texture) in a well-lit environment, while IR sensors do not require external light but mainly rely on the radiant temperature of the object, which is extremely useful for nighttime or shadow detection[11].

The main challenges for robust IR pedestrian detection can be classified into two main types. First, IR images have some adverse properties, such as their noisy nature, low-resolution, and no visual detailed information. These adverse properties make the discriminative feature extraction of an object very difficult, and thus affect the detection performance. Second, IR images are susceptible to weather conditions since IR cameras detect the difference in the temperature of the environment. For instance, pedestrians look brighter than the background in cold weather, while their brightness looks like the background in hot weather[12]. To overcome these problems, multispectral pedestrian detectors, which take aligned color–thermal image pairs as input, are proposed. Color images have more detailed information under good lighting conditions than thermal images, such as information on edge, texture, and color. However, when color images are captured under weak illumination conditions or during the nighttime, the feature information is lost, and the detection performance is reduced; examples can be found in the second row of Figure1. Thermal sensors are not sensitive to light and can capture pedestrians' contours under the insufficient lighting conditions [13]; examples can be found in the first row of Figure 1

Although multispectral pedestrian detection is superior to the single modality setting (visible or thermal), it still suffers several limitations as follows:

Firstly, the quality of feature fusion is affected by the focalizing centrality of pedestrian in fused feature map. Particularly, it highlights the complementary information from multiple modalities. Classical feature fusion method usually concatenates or averages the two-modality features [14] which enhances local feature from spatial perspective, without making use of mask information. Besides, effective complementary information fusion from multiple modalities remains an open problem. Secondly, weak appearance objects (e.g., small, or occluded pedestrians) usually suffer with low confidence score due to the inconsistency between IoU and classification score[15]. To address these issues, it is crucial for alleviating the misalignment problem by cascade learning strategy. However, the object awareness in the feature map attracts less attention, which has a significant impact on the performance improvement[16].

Infrared (IR) and visual (VI) image fusion is designed to fuse multiple source images into a comprehensive image to boost imaging quality and reduce redundancy information, which is widely used in various imaging equipment to improve the visual ability of human and robot. The accurate, reliable, and complementary descriptions of the scene in fused images make these techniques widely used in various fields. In recent years, a large number of fusion methods for IR and VI images have been proposed due to the ever-growing demands and the progress of image representation methods[17]

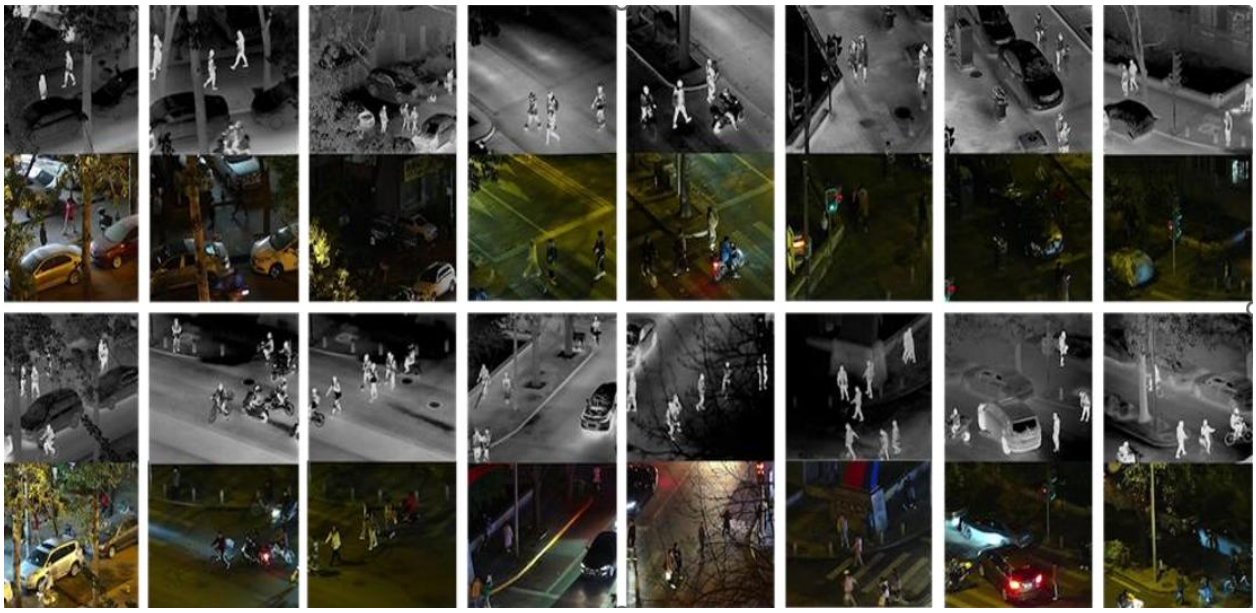

Figure 1. Examples of multispectral pedestrian detection images the first and second rows are thermal and color images[13]**.**

2.Related work

Despite significant progress in machine learning, pedestrian detection in the real-world is still regarded as one of the challenging problems, limited by occluded appearances, cluttered backgrounds, and bad visibility at night. This has caused detection approaches using multi-spectral sensors such as color and thermal which could be complementary to each other. The authors in[18] propose a novel sensor fusion framework for detecting pedestrians even in challenging real-world environments. design a convolutional neural network (CNN) architecture that consists of three-branch detection models taking different modalities as inputs. all detection probabilities from each modality in a unified CNN framework and selectively use them through a channel weighting fusion (CWF) layer to maximize the detection performance. An accumulated probability fusion (APF) layer is also introduced to combine probabilities from different modalities at the proposal-level.

Multispectral images consisting of aligned visual-optical (VIS) and thermal infrared (IR) image pairs are well-suited for practical applications like autonomous driving or visual surveillance. Such data can be used to increase the performance of pedestrian detection especially for weakly illuminated, small-scaled, or partially occluded instances. The current state-of-the-art is based on variants of Faster R-CNN and thus passes through two stages: a proposal generator network with handcrafted anchor boxes for object localization and a classification network for verifying the object category. The authors propose a method for effective and efficient multispectral fusion of the two modalities in an adapted single-stage anchor-free base architecture. aim at learning pedestrian representations based on object center and scale rather than direct bounding box predictions. In this way, can both simplify the network architecture and achieve higher detection performance, especially for pedestrians under occlusion or at low object resolution [19].

Pedestrian safety is one of the most significant issues in the development of advanced driver assistant systems and autonomous vehicles. DNN (Deep Neural Network) or deep learning has been effectively implemented through many applications, especially on object classification.

Most pedestrian detection research only dealt with detection during the daytime or nighttime. The authors evaluate and compare the performance of DNN-based pedestrian detection algorithm YOLO at both daytime and nighttime environment. The experiment result indicates that the performance of DNN-based pedestrian detection is significantly affected due to the lighting conditions [20].

In [13]the authors puts forward a robust multispectral feature fusion network (MSFFN) for pedestrian detection, which fully integrates the features extracted from visible light and infrared channels. multiscale semantic features were extracted by two core modules, namely, multiscale feature extraction of visible images (MFEV) and multiscale feature extraction of infrared images (MFEI) and fused by the improved YOLOv3 network for pedestrian recognition., it is proved that the MSFFN model can detect pedestrians more accurately than both MFEV and MFEI over daytime and nighttime images on multiple scales.

Multispectral pedestrian detection suffers from modality imbalance problems which will hinder the optimization process of dual-modality network and depress the performance of detector, in[21] The authors propose Modality Balance Network (MBNet) which facilitates the optimization process in a much more flexible and balanced.

Preliminary fusion strategies would fail to exploit informative features from cross-spectral images, may introduce additional interference. The authors propose an attention based multi-layer fusion network in the triple-stream deep convolutional neural network architecture for multispectral pedestrian detection. The effectiveness of multi-layer fusion is examined and verified, a channel-wise attention module (CAM) and a spatial-wise attention module (SAM) are developed and incorporated into the network aiming at more subtle adjustment to weights of multispectral features along both the channel and spatial dimensions respectively achieves state-of-the-art performance on all-day detection while outperforming compared methods at nighttime[22].

The existing vision based pedestrian detection algorithms still have two limitations that restrict their implementations, those being real-time performance as well as the resistance to the impacts of environmental factors, to address these issues, the authors propose a lightweight Illumination and Temperature aware Multispectral Network (IT-MN) for accurate and efficient pedestrian detection. The proposed IT-MN is an efficient one-stage detector. For accommodating the impacts of environmental factors and enhancing the sensing accuracy, thermal image data is fused by the proposed IT-MN with visual images to enrich useful information when visual image quality is limited [23].

The authors propose a pre-processing pipeline, which enables any existing normal-light pedestrian detection system to operate in low-light conditions. It is based on signal-processing and traditional computer-vision techniques, such as the use of signal strength of a depth sensing camera (amplitude images) and robust principal component analysis. The information in an amplitude image is less noisy, and is of lower dimension than depth data, marking it computationally inexpensive to process[24].

A problem pedestrian detection is the performance reduction at nighttime, the authors present multispectral image and deep neural network to improve the detection accuracy. In the evaluation, different image sources including RGB, thermal, and multispectral format are compared for the performance of the pedestrian detection. In addition, the optimizations of the architecture of the deep neural network are performed to achieve high accuracy and short processing time in the pedestrian detection task. The result implies that using multispectral images is the best solution for pedestrian detection at different lighting conditions[25].

the authors in [11] a novel and fully adaptive multispectral feature fusion approach, named Guided Attentive Feature Fusion (GAFF). By combining the intra- and intramodality attention modules, the proposed approach allows the network to learn the adaptive weighing and fusion of multispectral features. These two attention mechanisms are guided by the prediction and comparison of the pedestrian masks in the multispectral feature fusion stage. Specifically, at each spatial position, thermal or visible features are enhanced when they are in a pedestrian (intra-modality attention) or when they possess a higher quality than in the other modality (intramodality attention).

The authors in[26] believe the LLVIP dataset will contribute to the community of computer vision by promoting image fusion, , pedestrian detection, and image-to-image translation in very low-light applications, to improve the fusion quality of nighttime images, which suffer from low illumination, texture concealment, and color distortion, The authors in [27] first design a scene-illumination disentangled network (SID Net) to strip the illumination degradation in nighttime visible images while preserving informative features of source images. Then, a texture–contrast enhancement fusion network (TCEFNet) is devised to integrate complementary information and enhance the contrast and texture details of fused features., a color consistency loss is designed to mitigate color distortion from enhancement and fusion.

Although the research of single modal visible pedestrian detection has been very mature, it is still not enough to always meet the demand of pedestrian detection. Thus, a multi-spectral pedestrian detection method via image fusion and convolutional neural networks. The infrared intensity distribution and visible appearance features are retained with a total variation model based on local structure transfer, the results attest that method has superior detection performance, which can detect pedestrian targets robustly even in the case of harsh illumination conditions and cluttered backgrounds[28].

Multispectral pedestrian detection, which consists of a color stream and thermal stream, is essential under conditions of insufficient illumination because the fusion of the two streams can provide complementary information for detecting pedestrians based on deep convolutional neural networks (CNNs). To further improve the detection performance, the authors analyzed the existing multispectral fusion methods and proposed a novel multispectral channel feature fusion (MCFF) module for integrating the features from the color and thermal streams according to the illumination conditions[29]. Despite the rapid development of pedestrian detection algorithms, the balance between detection accuracy and efficiency is still far from being achieved due to edge GPUs (low computing power) limiting the parameters of the model. To address this issue, The researchers proposed YOLOv4-TP-Tiny based on the YOLOv4 model, which mainly includes two modules, two-dimensional attention (TA) and pedestrian-based feature extraction (PFM)[30].

3. Discussion The task of automatic encoding of images into a feature domain for classification purposes is executed by Convolutional Neural Networks (CNNs). In contrast Deep Neural Networks (DNNs) have demonstrated their efficiency through various practical implementations. Nevertheless. Prior studies have primarily concentrated solely on daytime or nighttime pedestrian detections while neglecting other instances when different lighting situations occur. Recent developments incorporating multispectral image pairs can enhance visual acquisition capabilities, which then creates more dependable and stable pedestrian detection systems. Results from conducted experiments highlighted the advantage of utilizing multispectral imagery as it provided the best solution to detect pedestrians under varying

lighting conditions. Optimum performance was achieved via multi spectral pedestrian detection methods based on image fusion and convolutional neural networks even in adverse illumination settings and undesired background circumstances.

Table 1 Summary of classifiers used for pedestrian detection of multispectral image.

Authors	algorithm	Dataset	Detection rate
Park et al. (2020)[19]	Faster R-CNN	KAIST	7.49%
Jason et al. (2020)[20]	DNN (YOLO)	KAIST	In the daytime 45% only 20% in the nighttime
Yongato et al.(2020)[22]	R-CNN (CS-RCNN)	KAIST	/
Yanlei et al. (2021) [25]	deep neural network	/	6.9%
Yang et al.(2021)[23]	(IT-MN)	/	4.19%
Zhou et al. (2022)[30]	YOLOv4-TP-Tiny	/	58.3%

4. Conclusion

Pedestrian detection is of great importance in the field of computer vision, and this field has witnessed great progress. However, pedestrian detection remains relatively difficult, especially in low light conditions and bad weather. The use of multispectral images with deep neural network techniques can provide a high pedestrian detection rate, by utilizing multispectral image pairs (infrared and visible images), which offer complementary visual information, and from these the pedestrian detection systems become more reliable and powerful. This review first introduces the content of general object detection, explains the multispectral images then analyzes the development of pedestrian detection.

ACKNOWLEDGEMENTS

The authors would like to thank Mustansiriyah University for supported (https://uomustansiriyah.edu.iq/) Baghdad- Iraq for its support in this work.

References

[1] L. Ding, Y. Wang, R. Laganière, D. Huang, and S. Fu, “Convolutional neural networks for multispectral pedestrian detection,” Signal Process. Image Commun., vol. 82, p. 115764, 2020, doi: https://doi.org/10.1016/j.image.2019.115764.

[2] P. Dollár, C. Wojek, B. Schiele, and P. Perona, “Pedestrian detection: A benchmark,” in 2009 IEEE conference on computer vision and pattern recognition, 2009, pp. 304–311.

[3] A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? the kitti vision benchmark suite,” in 2012 IEEE conference on computer vision and pattern recognition, 2012, pp. 3354–3361.

[4] S. Zhang, R. Benenson, and B. Schiele, “Are we ready for autonomous driving? the kitti vision benchmark suite,” in CVPR, 2015, vol. 1, no. 2, p. 4.
[5] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, and X. Liu, “Deep learning for generic object detection,” A Surv. [J]., 2018.
[6] R. Benenson, M. Omran, J. Hosang, and B. Schiele, “Ten years of pedestrian detection, what have we learned?,” in European Conference on Computer Vision, 2014, pp. 613–627.
[7] S. Zhang, R. Benenson, M. Omran, J. Hosang, and B. Schiele, “How far are we from solving pedestrian detection?,” in Proceedings of the iEEE conference on computer vision and pattern recognition, 2016, pp. 1259–1267.
[8] A. González et al., “Pedestrian detection at day/night time with visible and FIR cameras: A comparison,” Sensors, vol. 16, no. 6, p. 820, 2016.
[9] S. Hwang, J. Park, N. Kim, Y. Choi, and I. So Kweon, “Multispectral pedestrian detection: Benchmark dataset and baseline,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1037–1045.
[10] C. Li, D. Song, R. Tong, and M. Tang, “Illumination-aware faster R-CNN for robust multispectral pedestrian detection,” Pattern Recognit., vol. 85, pp. 161–171, 2019.
[11] H. Zhang, E. Fromont, S. Lefèvre, and B. Avignon, “Guided attentive feature fusion for multispectraal pedestrian detection,” in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021, pp. 72–80.
[12] Y. Chen and H. Shin, “Pedestrian detection at night in infrared images using an attention-guided encoder-decoder convolutional neural network,” Appl. Sci., vol. 10, no. 3, 2020, doi: 10.3390/app10030809.
[13] X. Song, S. Gao, and C. Chen, “A multispectral feature fusion network for robust pedestrian detection,” Alexandria Eng. J., vol. 60, no. 1, pp. 73–85, 2021, doi: https://doi.org/10.1016/j.aej.2020.05.035.
[14] L. Zhang et al., “Cross-modality interactive attention network for multispectral pedestrian detection,” Inf. Fusion, vol. 50, pp. 20–29, 2019, doi: https://doi.org/10.1016/j.inffus.2018.09.015.
[15] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask r-cnn,” in Proceedings of the IEEE international conference on computer vision, 2017, pp. 2961–2969.
[16] Z. Cai and N. Vasconcelos, “Cascade r-cnn: Delving into high quality object detection,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 6154–6162.
[17] X. Jin et al., “A survey of infrared and visual image fusion methods,” Infrared Phys. Technol., vol. 85, pp. 478–501, 2017.
[18] K. Park, S. Kim, and K. Sohn, “Unified multi-spectral pedestrian detection based on probabilistic fusion networks,” Pattern Recognit., vol. 80, pp. 143–155, 2018, doi: https://doi.org/10.1016/j.patcog.2018.03.007.
[19] A. Wolpert, M. Teutsch, M. S. Sarfraz, and R. Stiefelhagen, “Anchor-free small-scale multispectral pedestrian detection,” arXiv Prepr. arXiv2008.08418, 2020.
[20] J. Nataprawira, Y. Gu, K. Asami, and I. Goncharenko, “Pedestrian Detection in Different Lighting Conditions Using Deep Neural Networks.,” in IICST, 2020, pp. 97–104.
[21] K. Zhou, L. Chen, and X. Cao, “Improving multispectral pedestrian detection by addressing modality imbalance problems,” in European Conference on Computer Vision, 2020, pp. 787–803.
[22] Y. Zhang, Z. Yin, L. Nie, and S. Huang, “Attention based multi-layer fusion of multispectral

images for pedestrian detection," IEEE Access, vol. 8, pp. 165071–165084, 2020.
[23] Y. Zhuang, Z. Pu, J. Hu, and Y. Wang, "Illumination and temperature-aware multispectral networks for edge-computing-enabled pedestrian detection," IEEE Trans. Netw. Sci. Eng., vol. 9, no. 3, pp. 1282–1295, 2021.
[24] V. Shah, A. Agarwal, T. T. Verlekar, and R. Singh, "Adapting Deep Neural Networks for Pedestrian-Detection to Low-Light Conditions without Re-training," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 2535–2541.
[25] J. Nataprawira, Y. Gu, I. Goncharenko, and S. Kamijo, "Pedestrian detection using multispectral images and a deep neural network," Sensors, vol. 21, no. 7, p. 2536, 2021.
[26] X. Jia, C. Zhu, M. Li, W. Tang, and W. Zhou, "LLVIP: A visible-infrared paired dataset for low-light vision," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 3496–3504.
[27] L. Tang, X. Xiang, H. Zhang, M. Gong, and J. Ma, "DIVFusion: Darkness-free infrared and visible image fusion," Inf. Fusion, vol. 91, pp. 477–493, 2023, doi: https://doi.org/10.1016/j.inffus.2022.10.034.
[28] X. Chen, L. Liu, and X. Tan, "Robust Pedestrian Detection Based on Multi-Spectral Image Fusion and Convolutional Neural Networks," Electronics, vol. 11, no. 1, p. 1, 2021.
[29] Z. Cao, H. Yang, J. Zhao, S. Guo, and L. Li, "Attention fusion for one-stage multispectral pedestrian detection," Sensors, vol. 21, no. 12, p. 4184, 2021.
[30] H. Zhou, T. Wu, K. Sun, and C. Zhang, "Towards high accuracy pedestrian detection on edge gpus," Sensors, vol. 22, no. 16, p. 5980, 2022.

الملخص

الكشف عن الكائن هو مشكلة أساسية في رؤية الآلة والتعلم العميق. يعتبر كشف المشاة تطبيقًا محددًا لمشكلة كشف الكائن، وقد أصبح واحدًا من نقاط البحث الرئيسية في السنوات الأخيرة. في الماضي، تم اقتراح العديد من التقنيات لكشف المشاة كتطبيق محدد لكشف الكائن. ومع ذلك فإن الاختلافات في الصور مثل ملابس الجسم والوضعية والإغلاق ومعلمات الإضاءة المختلفة في سيناريوهات مختلفة والفوضى الموجودة في الخلفية تشكل تحديات في تحقيق الدقة العالية. تقدم هذه الورقة مراجعة لبعض التقنيات المعروفة الموجودة للكشف عن المشاة بطريقة متعددة الأطياف عبر دمج الصور والشبكات العصبية التلافيفية.