T2R2 東京科学大学 リサーチリポジトリ
Science Tokyo Research Repository

論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Learning-Based Pedestrian Detection on Visible and Thermal Images with Multiple Regressors Considering Misalignment
著者(和文)	WANCHAITANAWONGNapat
Author(English)	Napat Wanchaitanawong
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第12852号, 授与年月日:2024年9月20日, 学位の種別:課程博士, 審査員:田中 正行,塚越 秀行,中臺 一博,原 精一郎,川上 玲,奥冨 正敏
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第12852号, Conferred date:2024/9/20, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

Powered by T2R2 (Science Tokyo Research Repository)

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

系・コース： Department of, Graduate major in	Systems and Control Engineering Systems and Control Engineering	系 コース	申請学位（専攻分野）： Academic Degree Requested	博士 Doctor of	(Engineering)
学生氏名： Student's Name	Napat Wanchaitanawong		審査員主査： Chief Examiner	Masayuki Tanaka	

要旨（英文 800 語程度）
Thesis Summary（approx.800 English Words ）

Pedestrian detection is a crucial task in computer vision, particularly significant in applications like autonomous driving and surveillance. Accurate pedestrian detection is essential for ensuring the safety of both pedestrians and vehicle occupants, as well as enhancing security in public spaces. Traditional pedestrian detection techniques, which rely heavily on visible light cameras, face several challenges. These methods often struggle with varying illumination conditions, such as low light or nighttime scenarios, and cluttered environments with complex backgrounds. Additionally, traditional methods are sensitive to occlusions, where pedestrians are partially hidden by other objects, and variations in pedestrian poses, making detection less reliable.

To address these issues, researchers have explored multi-modal approaches that integrate data from different sensor types, specifically visible and thermal sensors. Thermal sensors capture heat signatures, which can reveal pedestrian presence even in low-light conditions, providing complementary information to visible light cameras. However, existing multi-modal methods often assume perfect alignment between the visible and thermal images. This assumption is rarely valid in real-world scenarios, where misalignment can occur due to differences in sensor viewpoints, calibration errors, and temporal asynchronization. Misalignment complicates the process of fusing data from both sensors. Algorithms that do not consider misalignment directly fuse the data, leading to worse performance due to inaccurate correspondences between the modalities. This makes it difficult to combine information from both modalities accurately, often resulting in detection errors such as false positives or false negatives. These discrepancies reduce the reliability and effectiveness of multi-modal pedestrian detection systems, highlighting the need for more robust solutions.

The primary focus of this research is to propose novel methods that address the misalignment problem in multi-modal pedestrian detection. The study begins with a comprehensive review of existing pedestrian detection methods, including both traditional and deep learning-based approaches. This review highlights significant challenges, particularly the issue of misalignment between visible and thermal modalities, guiding the development of novel detection methods. To tackle these challenges, this research proposes new evaluation metrics, multi-modal Intersection over Union (IoU^M) and Multi-Modal Log-Average Miss Rate (MR^M). These metrics provide a more accurate measurement of detection precision across different modalities. Moreover, we introduce two advanced multi-modal pedestrian detection frameworks: Multi-Modal Faster R-CNN and Multi-Modal Single Shot MultiBox Detector (SSD).

The Multi-Modal Faster R-CNN framework modifies both the Region Proposal Network (RPN) and the detector to incorporate a multi-modal regressor. This framework generates bounding box pairs that accurately locate pedestrians

in both modalities despite significant misalignment. The Multi-Modal SSD framework enhances real-time detection capabilities and improves accuracy and robustness in misaligned environments by incorporating multi-modal regressors, object-based training techniques, and shifting data augmentation. Both frameworks are validated through extensive experiments, including simulated disparity tests, demonstrating superior performance in misaligned scenarios using the KAIST Multispectral Pedestrian Dataset. Additionally, ablation studies verify the effectiveness of multi-modal regressors compared to traditional single-modal regressors and highlight the utility of IoU^M as a criterion.

The final chapters discuss the main findings and contributions in detail, providing a thorough analysis of the experimental results. The research emphasizes the strengths and limitations of the proposed methods, offering insights for future research directions. Ultimately, this work significantly contributes to the advancement of multi-modal pedestrian detection technology, fostering the development of safer and more reliable systems across various domains, including autonomous driving and surveillance. Future work will focus on further enhancing alignment techniques, integrating real-time processing capabilities, and exploring advanced fusion strategies to continue advancing this critical field. By improving the robustness and accuracy of pedestrian detection systems, this research aims to enhance the safety and reliability of applications that rely on accurate pedestrian detection.

In conclusion, this thesis provides a comprehensive and innovative approach to multi-modal pedestrian detection, addressing the key challenge of misalignment and paving the way for future advancements in the field. The proposed evaluation metrics and frameworks offer practical solutions that enhance the accuracy and robustness of pedestrian detection systems, making significant contributions to the advancement of computer vision technology.

備考 ： 論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。
Note : Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).

注意：論文要旨は、東工大リサーチリポジトリ(T2R2)にてインターネット公表されますので、公表可能な範囲の内容で作成してください。
Attention: Thesis Summary will be published on Tokyo Tech Research Repository Website (T2R2).