ENG 573 Capstone Proposal: GE Epistemic-Classifier

Stanley Gu Shgu2
Hongqing Liu hl85
Alec Petersdorf alecmp2
Weichen Liu wl45
Sanjit Arunkumar sanjita3

1. Introduction

In the realm of autonomous systems, robust and reliable detection capabilities are essential for interacting with the physical world. Cameras capturing Red-Green-Blue (RGB) images are extensively used for this purpose. However, the reliability of RGB images is often compromised under adverse conditions such as inclement weather or poor illumination. An alternate solution involves the use of infrared (IR) imaging, which offers more consistent performance under challenging conditions. The primary goal of this research is to enhance the reliability of individual object detection accuracy in autonomous systems through the incorporation of both IR and RGB data. This will be achieved by adapting an Epistemic Classifier for IR imaging and then fusing the output with that from a modified Epistemic Classifier for visible light.

1.1. Problem definition

The central issue arises from the limitations of detection mechanisms that are solely dependent on RGB images, primarily due to their vulnerability to environmental variables such as weather and light conditions. In contrast, while IR images are significantly less sensitive to these environmental factors, they typically exhibit lower resolution and less detailed textures. As such, these two imaging modalities can provide complementary information to each other. Our proposed solution involves modifying an existing Epistemic Classifier to accommodate IR data, and fusing this with the RGB data, thereby leveraging the robust performance of IR imaging under adverse conditions. A fundamental requirement for our system is its capability to discern when to rely on RGB, IR, or a combination of both modalities. This necessitates that the system possesses an epistemic understanding of the confidence level associated with each classification. Initially, this research will focus on single-object detection, serving as a proof of concept. Subject to successful outcomes in this phase, our aim is to progress to multi-modal detection.

1.1.1 Outline of the proposed method

Our proposed methodology encompasses several stages, starting with the acquisition of a suitable dataset. We will initially strive to utilize publicly available RGB and IR datasets, such as the RoadScene and FLIR ADAS datasets. If these datasets are unsuitable, we are prepared to develop a custom dataset. The chosen network architecture involves the use of a Convolutional Neural Network (CNN) for RGB data and a Fully Convolutional Network (F-CNN) for IR data, with the initial focus on single-object detection as a proof of concept. If necessary, domain adaptation techniques will be considered for further network fine-tuning.

One of the pivotal facets of our methodology is the intermediate fusion strategy for data processing. To enact this, we propose to train the IR and RGB networks independently, followed by a fusion at the feature level, thereby harnessing the distinct strengths of both modalities.

As we transition from single-object detection to multi-object detection, precise calibration of both RGB and IR cameras becomes essential to ensure pixel overlap. This necessitates the determination of both intrinsic and extrinsic parameters for each camera type. Furthermore, the multi-object detection will employ Mask R-CNN to augment the sophistication of object detection in various complex scenarios.

1.1.2 Desired outcome & Minimum outcome

The optimal outcome of this research would be a fully functioning multi-object detection system that seamlessly integrates the two sensor modalities. This would entail achieving consistent classifications in the ”I Know” region, thereby ensuring high-confidence detection across both RGB and IR data streams.

At the very least, we aim to establish a fully functioning single-object detection system that effectively fuses data from both RGB and IR sensors. This minimal goal serves as a critical foundation upon which the more ambitious multi-object detection system can be built.

2. Personal Objectives

Stanley Gu: My goal is to grasp multimodal sensor fusion, specifically the integration of RGB and IR data for effective object detection. Additionally, I wish to master Mask R-CNN, including its theory, practical use, and customization for image processing tasks.

Hongqing Liu: I want to learn how to evaluate reliability of different modality and fuse data based on these reliability

in this project. I also want to learn some object detectors in RGB and IR field.

Alec Petersdorf: I hope to broaden my perspective with this project, especially with exposure to a different light spectrum and to learn how to fuse sensor data together.

Weichen Liu: I'm a second-year master's student and I'm currently interested in computer vision on IR and RGB image field.

Sanjit Arunkumar:I want to gain more experience dealing with computer vision techniques using RGB as well as IR imaging, as well as apply machine learning concepts that i have learnt to practice.

3. Literature review

3.1. Epistemic Classifier

Virani et. al. [2] proposed a way to validate machine learning predictions using nearest neighbors to determine support for predictions. The support is used for justification and labels the prediction IK, IMK, and IDK, depending on how similar the support is for each class prediction.

3.2. Camera Calibration

Zhang [7] proposed a calibration method in five steps which involve printing a pattern onto a plane, capturing its image at various orientations, extracting feature points, estimating intrinsic and extrinsic parameters, and then optimizing the parameters by minimizing a maximum likelihood estimate. OpenCV also has some built-in functions that work with a chessboard which may be useful for quick calibration.

3.3. RGB image detection network

RGB image detection networks utilize the Red, Green, and Blue color channels to extract features and obtain spatial information for accurate object localization. Prominent models such as R-CNN, Faster R-CNN, and SSD have been developed to improve detection accuracy and speed. R-CNN introduced a two-step process of region proposal and feature extraction, while Faster R-CNN incorporated a Region Proposal Network (RPN) for end-to-end learning. SSD is another concept that has achieved real-time detection using multiple convolutional feature maps.

3.4. Infrared image detection

Infrared image detection algorithms aim to extract meaningful information from infrared images, which are captured in the non-visible spectrum and provide valuable insights beyond what is visible to the human eye. In recent years, deep learning-based methods have gained significant attention for their ability to automatically learn and extract complex features from infrared images. Convolutional neural networks (CNN's) have been widely employed in infrared image detection tasks, demonstrating superior performance compared to traditional methods. These deep learning models are trained on large-scale annotated datasets, enabling them to learn discriminative features and achieve state-of-the-art results in infrared image detection.

3.5. RGB-IR Image Fusion

The objective of RGB-IR image fusion is to merge data from RGB and IR images into a solitary image, enhancing its value as a data source for various applications. In recent years, many fusion methods based on deep learning have merged. In 2018, Xu et al. introduced a fusion tracking algorithm that utilizes visible and infrared images based on CNN [4]. It is a pixel-level fusion and simply concatenates RGB and IR images before processing them with CNN. Zhang et al. [6] introduced a fusion method based on MDnet. This method is a feature-level fusion, as it performs fusion based on joint features obtained from CNN. There are also some works that take the reliability of each modality into account. For example, Li et al. [8] proposed FANet, a quality-aware feature aggregation network for fusion tracking, and a modality weight computation method was proposed based on a response map of each modality in Zhang's work [5]. Deep learning can help to solve important problems in the image fusion field by providing better features and better adaptive weights for fusion.

4. Resources

The potential datasets we use to pretrain both networks is the FLIR video dataset, which contains 9,711 thermal and 9,233 RGB training/validation images with a suggested training/validation split. Includes 16-bit pre-AGC frames. There is a RoadScene dataset [3] derived from the FLIR video dataset, which contains 221 aligned Vis and IR image pairs containing rich scenes such as roads, vehicles, pedestrians, and so on. We will use these datasets to train the fusion model.

We plan to use the provided code from GE [1] and mainly the Tensorflow Python library. The group will use GitHub for version control.

5. Plan

Timeline	Task
Week 1-3	Single Object Detection on each Sensory Modality Separately
Week 4	Add Epistemic Classifier to Each Network
Week 5-6	Fusion Algorithm
Week 7-8	Multi-object detection and Camera Calibration
Week 9-10	Object Tracking(optional)

Table 1. Research Timeline and Corresponding Tasks

References

[1] Naresh Iyer, Nurali Virani, and Zhaoyuan Yang. Eepistemic-classifier, 2021. https://github.com/GE-Research-Machine-Learning/Epistemic-Classifier.

[2] Nurali Virani, Naresh Iyer, and Zhaoyuan Yang. Justification-based reliability in machine learning, 2021.

[3] Han Xu, Jiayi Ma, Junjun Jiang, Xiaojie Guo, and Haibin Ling. U2fusion: A unified unsupervised image fusion network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[4] Ningwen Xu, Gang Xiao, Xingchen Zhang, and Durga Prasad Bavirisetti. Relative object tracking algorithm based on convolutional neural network for visible and infrared video sequences. In *Proceedings of the 4th International Conference on Virtual Reality*, pages 44–49, 2018.

[5] Xingchen Zhang, Ping Ye, Shengyun Peng, Jun Liu, Ke Gong, and Gang Xiao. Siamft: An rgb-infrared fusion tracking method via fully convolutional siamese networks. *IEEE Access*, 7:122122–122133, 2019.

[6] Xingming Zhang, Xuehan Zhang, Xuedan Du, Xiangming Zhou, and Jun Yin. Learning multi-domain convolutional network for rgb-t visual tracking. In *2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pages 1–6. IEEE, 2018.

[7] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.

[8] Yabin Zhu, Chenglong Li, Bin Luo, and Jin Tang. Fanet: Quality-aware feature aggregation network for robust rgb-t tracking. *arXiv preprint arXiv:1811.09855*, 2018.