A Multimodal Fusion-Based LNG Detection for Monitoring Energy Facilities (Student Abstract)

Junchi Bin[1], Choudhury A. Rahman[2], Shane Rogers[2], Shan Du[1], Zheng Liu[1]

[1]School of Engineering, University of British Columbia, Okanagan Campus, Kelowna, BC, Canada
[2]Intelliview Technologies Inc., Calgary, AB, Canada
{junchi.bin, zheng.liu}@ubc.ca

Abstract

Fossil energy products such as liquefied natural gas (LNG) are among Canada's most important exports. Canadian engineers devote themselves to constructing visual surveillance systems for detecting potential LNG emissions in energy facilities. Beyond the previous infrared (IR) surveillance system, in this paper, a multimodal fusion-based LNG detection (MFLNGD) framework is proposed to enhance the detection quality by the integration of IR and visible (VI) cameras. Besides, a Fourier transformer is developed to fuse IR and VI features better. The experimental results suggest the effectiveness of the proposed framework.

Introduction

The energy department of Canadian government reported that the export of natural gas made CAD$6.1 billion of net profits in 2017 (Bin et al. 2021). In order to meet the rising demands of natural gas, Canadian companies have been continuously constructing refineries for increasing capability of export. In refineries, the raw natural gas is cooled and compressed to liquefied natural gas (LNG) which enables large vessel to transport it overseas with numerous quantities. In recent energy industries, infrared (IR) cameras are usually installed to monitor the potential LNG emission around the pipeline. The majority of the early works (Bin et al. 2021; Wang et al. 2020) focus on implementing deep learning techniques to classify if LNG is in the regions of interest from background subtraction. Although these methods have demonstrated their effectiveness in LNG emission detection, the applied background subtraction technique limits the system in surveillance scenarios (fixed cameras are fixed on the plant). Besides, the system may be unstable if the camera jitters which damages the effectiveness of background subtraction. Therefore, it is recommended to developing a more stable and portable detection framework such as Faster RCNN (Ren et al. 2015). On the other hand, with advances in sensory technology, the visible (VI) and infrared (IR) cameras are usually ensembles on a surveillance system. Although the IR camera is sensitive to LNG, the lack of context information in IR images makes the LNG detection less precise. On the contrary, the VI images have rich context information to compensate for the limitation of IR images. However, there is still a lack of investigation on developing a multimodal LNG emission detection system based on VI and IR cameras. Based on the aforementioned concerns, a multimodal fusion-based LNG detection framework, i.e., MFLNGD, is proposed to address the aforementioned problems in this study as shown in Figure 1.

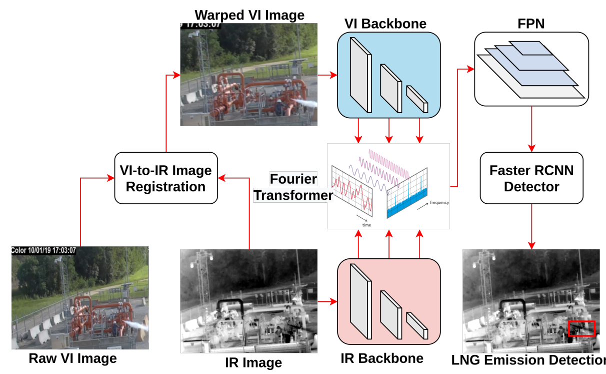

Figure 1: The illustration of the proposed multimodal fusion-based liquefied natural gas detection (MFLNGD)

Methodology

Overview. First, the raw VI images are aligned and warped to match with IR images in a pixel-wise manner. Then, the warped VI images and IR images are fed into the corresponding backbone for feature extraction. In this study, the geometric fusion (Prakash, Chitta, and Geiger 2021) is adopted for VI-IR feature fusion. Moreover, to reduce the memory occupation from the original vision transformer (Prakash, Chitta, and Geiger 2021), the Fourier transformer is proposed to achieve efficient VI-IR feature fusion. Finally, the feature pyramid network (FPN) and Faster RCNN detector are used for final LNG detection.

VI-to-IR Image Registration. In order to unsupervised match, the VI images to IR images, the normalized total gradient (NTG) as target function is implemented to solve the transformation function optimized by stochastic gradient descent (Chen et al. 2018). The NTG aims to achieve VI-IR registration based on image gradient consistency.

Backbone. The backbone is the primary network for

Backbone	Modality	Fusion Operators	Scene Dependence		Scene Independence		Overall	
			mAP	AP50	mAP	AP50	mAP	AP50
R18	VI	–	18.24	57.97	4.18	18.87	11.21	38.42
	IR	–	19.63	55.16	1.47	7.66	10.55	31.41
	VI + IR	Embedding	33.68	79.31	**12.83**	30.74	23.25	55.02
R18	VI + IR	FT	42.35	90.30	11.98	**40.10**	27.16	**65.20**
V19	**VI + IR**	**FT**	**51.94**	**94.48**	10.94	35.21	**31.44**	64.84

Table 1: Detection results of the proposed MFLNGD.

feature extraction. In this study, two popular backbones, ResNet-18 (R18) and VovNet-19 (V19) are employed (Ren et al. 2015; Lee et al. 2019). In our case, the outputs of stages from VI and IR backbones are fused by fusion operators accordingly.

Fourier Transformer for Fusion. Although vision transformer can achieve a better result in multimodal fusion, such fusion operator may be too large to train with images of higher resolution. Inspired by the recent NLP technique (Lee-Thorp et al. 2021), the multi-head attention is replaced by a fast Fourier transform to fuse multimodal information without increasing memory load.

FPN and Detector. Based on the geometric fusion (Prakash, Chitta, and Geiger 2021), there is a set of features from multiple stages of the backbone. Then, these features are fed into FPN for enriching multi-scale features. Finally, the Faster RCNN detector localizes and classifies the regions of LNG from the fused multimodal information (Ren et al. 2015).

Experiments

Data Description. The experimental dataset is collected from an existing industrial natural gas facility for training models. The dataset consists of 5352 8-bit 512 $\times$ 640 IR and VI video frames, which are derived from thermal surveillance devices. In order to comprehensively evaluate the model's performance, there are two test sets, i.e., scene dependence for evaluating model's capability (dataset is randomly split by 70% and 30%), and scene independence for evaluating model's robustness (dataset is split by half based on recording scenarios). The evaluation metrics are in COCO format (Bin et al. 2021)

Experimental Results and Discussion. For validating the feasibility of the proposed MFLNGD, only IR and VI video frames are used with R18 as the backbone. Although both IR and IV can perceive the LNG in the scene dependence set, they are lack generalization for the unfamiliar scenes resulting in low mAP and AP50 in scene independent sets. After combining VI and IR images, the performance is significantly improved with an embedding network (3 $\times$ 3 Convolution) after concatenation of VI and IR features. However, the embedding network cannot extract the global information, which is important to the perception of the non-rigid object such as gas. Using the proposed Fourier transformer (FT) as fusion operators can fuse VI and IR features concerning the global information. The results are further improved after using FT in the proposed framework. Moreover, using V19 as backbone can further improve the model's capability. It is noticeable that the gap is large between mAP and AP in the scene independent set. The reason is the LNG object is a non-rigid body that changes over time. Therefore, it isn't easy to have an extremely precise localization result. However, the values of AP50 are satisfactory for industrial demand in visual surveillance of LNG.

Conclusion

This paper proposes a multimodal fusion-based LNG detection (MFLNGD) framework to detect the non-rigid gas object with hybrid information from VI and IR cameras after calibration. Moreover, a fusion operator, Fourier transformer, is proposed to enhance detection quality. The experimental suggests the effectiveness of the proposed framework. In the future, the motion analysis module such as optical flow to further enhance the robustness of the LNG detection.

References

Bin, J.; Rahman, C. A.; Rogers, S.; and Liu, Z. 2021. Tensor-Based Approach for Liquefied Natural Gas Leakage Detection From Surveillance Thermal Cameras: A Feasibility Study in Rural Areas. *IEEE Transactions on Industrial Informatics*, 17(12): 8122–8130.

Chen, S.-J.; Shen, H.-L.; Li, C.; and Xin, J. H. 2018. Normalized Total Gradient: A New Measure for Multispectral Image Registration. *IEEE Transactions on Image Processing*, 27(3): 1297–1310.

Lee, Y.; Hwang, J.-w.; Lee, S.; Bae, Y.; and Park, J. 2019. An Energy and GPU-Computation Efficient Backbone Network for Real-Time Object Detection. In *CVPR Workshops*.

Lee-Thorp, J.; Ainslie, J.; Eckstein, I.; and Ontanon, S. 2021. FNet: Mixing Tokens with Fourier Transforms. arXiv:2105.03824.

Prakash, A.; Chitta, K.; and Geiger, A. 2021. Multi-Modal Fusion Transformer for End-to-End Autonomous Driving. In *CVPR*.

Ren, S.; He, K.; Girshick, R.; and Sun, J. 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In *NeurIPS*, volume 28.

Wang, J.; Tchapmi, L. P.; Ravikumar, A. P.; McGuire, M.; Bell, C. S.; Zimmerle, D.; Savarese, S.; and Brandt, A. R. 2020. Machine vision for natural gas methane emissions detection using an infrared camera. *Applied Energy*, 257: 113998.