Three-Channel Infrared Imaging for Object Detection in Haze

Beinan Yu, Yifan Chen, Si-Yuan Cao, Hui-Liang Shen, and Junwei Li

***Abstract*— Object detection is of wide application for its capability in recognizing and locating targets in scenes. Under heavy hazy conditions, however, the detection performance on RGB images will greatly degrade since the image contents are polluted. In this work, we propose an imaging system to acquire three-channel images in the shortwave infrared (SWIR) spectrum to facilitate object detection under hazy conditions. The system captures SWIR images in the form of pseudo color that are most suitable for detecting objects, such as pedestrians and vehicles. Two different types of filters, i.e., liquid crystal tunable filter (LCTF) and optical filters, are employed in our imaging system design. We use the LCTF to acquire narrowband hyperspectral images, which are fed into a band simulation model to generate wideband images for optimal band selection. We present a specific measure called recognition and localization (RL) score to evaluate the detection performance of three-band combinations. Based on the measure, optimal bands are determined using an efficient searching algorithm. Then, we customize three optical filters and install them on a filter wheel, with which we can acquire three-channel images in the SWIR spectrum. The effectiveness of our imaging system is evaluated on a self-collected RGB-SWIR image dataset. Experimental results indicate that, compared with the RGB images after haze removal, the three-channel SWIR images acquired by our imaging system are of higher quality and can achieve better object detection performance under hazy conditions.**

***Index Terms*— Band selection, dehazing, haze weather, hyperspectral imaging, infrared imaging, object detection, shortwave infrared (SWIR).**

I. Introduction

OBJECT detection, with its aim to locate targets in images, is an important task in computer vision. It has achieved great success in practical applications, such as surveillance [1], [2], defect inspection [3]–[5], and pedestrian detection [6], [7]. Recently, deep learning-based methods have come to dominate the field of object detection with the introduction of convolutional neural networks (CNNs) [8]. R-CNN [9], using a CNN as the classifier, yields considerable gains in accuracy, but its detection speed is slow. Fast R-CNN [10] and Faster R-CNN [11] improve the detection speed but still cannot meet the real-time requirement. YOLO [12] achieves real-time detection by employing a single network to predict both the bounding box and class label of each object in an image. The improved YOLO versions [13], [14] work in a similar way and produce better detection performance. However, although these detectors work well on clear images, they may perform quite poorly on hazy images that are of low contrast and faint color. Image fusion [15]–[17] and dehazing methods [18]–[21] have been introduced to enhance the quality of hazy images using image priors or neural networks. However, the dehazing effect is insignificant in the case of heavy haze and cannot obviously benefit object detection.

Recently, infrared images are increasingly adopted in object detection tasks when RGB information degrades or even disappears. In general, most captured infrared images are single-channel (panchromatic) or hyperspectral (more than 30 channels) ones. For example, single-channel infrared images have been employed to extract infrared edgelet [22] and thermal infrared (TIR) HOG [23] features under illumination insufficient conditions. Principal component thermography (PCT) [24] uses low-rank projection to extract thermal features for defect detection. The two variants, Sparse-PCT [25] and CCIPCT [26], can further improve detection accuracy and computational efficiency. Moreover, features from single-channel infrared data could also be combined into RGB features to form multispectral aggregated channel feature (MACF) [27]. The average miss rate can be reduced by 15% in this manner. UGC model [28] alleviates the modality discrepancy of RGB and TIR images and achieves a good performance in pedestrian detection by using features extracted from RGB and TIR images. STDFusionNet [29] fuses the thermal targets and the visible texture structures for better salient target detection. Hyperspectral images are also useful for object detection. For example, the RGBN-shortwave infrared (SWIR) system in [30] achieves 60%–70% accuracy when using infrared hyperspectral images in powder detection, whereas in RGB images, the powders are indistinguishable. The aforementioned works validate that the extracted infrared textural or spectral features are of great use in object detection.

For practical applications, band selection methods [31]–[37], which aim to find a small number of bands that are optimal for relevant vision tasks, have been introduced to

Manuscript received December 29, 2021; revised February 17, 2022; accepted March 19, 2022. Date of publication April 1, 2022; date of current version April 15, 2022. This work was supported by the National Natural Science Foundation of China under Grant 81973751. The Associate Editor coordinating the review process was Dr. Bardia Yousefi. *(Corresponding author: Si-Yuan Cao.)*

Beinan Yu, Yifan Chen, and Si-Yuan Cao are with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: mr_vernon@hotmail.com; chenyifan@zju.edu.cn; 11631012@zju.edu.cn).

Hui-Liang Shen and Junwei Li are with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China, and also with the Ningbo Research Institute, Zhejiang University, Ningbo 315100, China (e-mail: shenhl@zju.edu.cnn; lijunwei7788@zju.edu.cn).

Digital Object Identifier 10.1109/TIM.2022.3164062

Fig. 1. RGB and SWIR images of haze scenes acquired by our imaging system. The detected pedestrians and vehicles are marked with bounding boxes and confidence scores. Compared with the RGB images, the SWIR images are more clear and produce higher object detection precision.

improve imaging or computational efficiency. For example, OCF [34] clusters all bands according to their corresponding contributions and selects the most significant one in each cluster. E-FDPC [35] computes the product of local density and intracluster distance of each band and chooses the bands with anomalously large values. ONR [36] exploits the neighboring structure of hyperspectral images for band selection. FNGBS [37] chooses optimal bands according to the local intensity and information entropy. Continuum removal (CR), as a magnitude normalization approach, can explore spectral bands with distinctive absorption or reflectance characteristics [38], [39]. Based on CR, CIMDM [40] selects the most discriminable bands for object identification.

In this work, we propose a three-channel wideband imaging system in the SWIR spectrum for better object detection performance in common hazy scenarios, with the assistance of a multispectral imaging system. Three optimal widebands are determined for pedestrian and vehicle detection since these two targets are of most interest in visual surveillance tasks. As exemplified in Fig. 1, our acquired SWIR images are clearer than the RGB counterparts and offer higher accuracy in object detection. As will be shown in the experiments (Section IV), the sharpness of SWIR images can be improved after haze removal, which will further benefit object detection.

To solve the problem of data lack in optimal band selection, we present a linear model to faithfully render synthetic wideband images with various center wavelengths and bandwidths. The model assumes that a wideband image can be regarded as a linear combination of real narrowband (basis) images, where the coefficients (weights) of narrowband images are solved by simulating spectral transmittance. The synthetic images simulated using our model are close to real captured ones and thus can be used for data augmentation. We define a specific quantitative measure based on the recognition and localization (RL) accuracy of the YOLOv3 detector [13] to evaluate the performance of each three-band combination. Under its guidance, an efficient algorithm is employed to select optimal bands first locally and then globally based on our self-collected dataset.

In summary, the main contributions of this work are given as follows.

1) We introduce a design scheme of a three-channel wideband SWIR imaging system, with the assistance of a hyperspectral imaging system, for object detection in haze conditions.
2) We present a local-to-global band selection algorithm to choose the optimal widebands under the guidance of a specific measure for best object detection performance.
3) Our experiments validate that, compared with common RGB images, the three-channel SWIR images acquired by our imaging system are of high quality and yield improved object detection precision in haze conditions.

The rest of this article is organized as follows. Section II introduces the framework and relevant issues of the proposed three-channel SWIR imaging method. Section III elaborates the implementation details of the band selection algorithm. Section IV shows the experimental results, and finally, Section V concludes this article.

II. Propoesed Three-Channel SWIR Imaging Method

A. Motivation

It is known that the SWIR spectral characteristics benefit the object detection task since the spectrum varies with different objects. Recent studies have proved the advantages of hyperspectral images in object detection. However, the spectral "fingerprints" of objects may be covered by other unrelated spectral characteristics in the hyperspectral data. Moreover, it is time-consuming to detect objects using all spectral bands, and the hyperspectral data cannot meet the requirement of the common detectors designed for RGB images. Inspired by the red, green, and blue channels adopted for color imaging, we select three SWIR bands with distinctive characteristics for three-channel imaging. The three bands are selected with a wide bandwidth for high SNR of images, meanwhile with simple shapes (e.g., rectangle or Gaussian) to facilitate the manufacture of optical filters. Based on the selected bands, we can customize imaging systems with the specific color filter array for the actual object detection applications.

Current band selection methods, such as OCF [34] and CR [38], determine the optimal bands based solely on the spectral characteristics. In this work, we argue that spatial information is important to object detection and hence introduce a specific measure to guide the band selection process. Moreover, as our method is conducted on a simulated wideband image dataset, the selected bands will no longer be limited to the original narrow bands. This is beneficial to a practical three-band SWIR imaging system that requires fast imaging speed and good image quality.

B. Overview

As shown in Fig. 2, the spectral transmittances of the optical filters in the three-channel imaging system are optimally determined with the assistance of a hyperspectral imaging system. In the following, we introduce the details of the two imaging systems, as well as their relationship.

The hyperspectral imaging system consists of an SWIR camera and a liquid crystal tunable filter (LCTF).

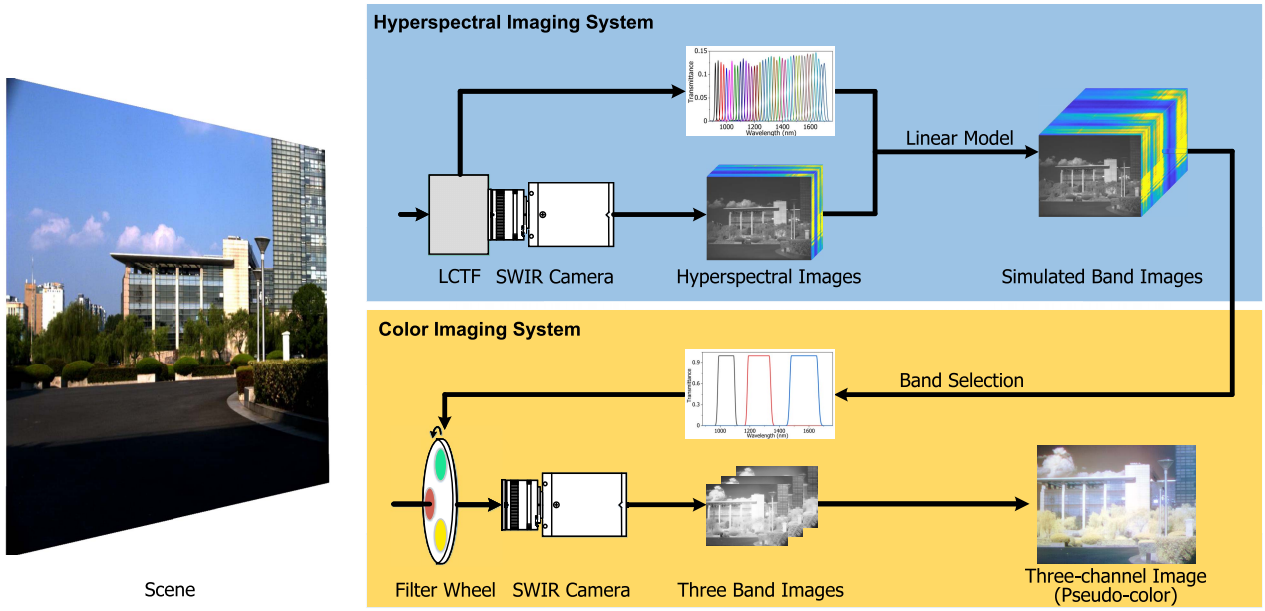

Fig. 2. Two SWIR imaging systems employed in this work. The hyperspectral imaging system acquires narrowband images using an LCTF. These images are further linearly combined to simulate plenty of wideband images for optimal band selection. Three custom-made optical filters are installed on a filter wheel, with which the three-channel imaging system captures three-channel SWIR images in form of pseudo color.

The full-width at half-maximum (FWHM) values of the LCTF are between 15 and 25 nm, varying with the center wavelengths of individual filters. The hyperspectral imaging system acquires 79 narrowband images of a scene in the SWIR spectrum from 920 to 1700 nm, by sequentially tuning the LCTF with an interval of 10 nm. With these narrowband images, we further simulate plenty of wideband images of the same scene with various bandwidths and center wavelengths. Then, three optimal bands are determined using a band selection algorithm guided by a specific measure of object detection precision.

The three-channel imaging system comprises an SWIR camera as well as a filter wheel installed with three optical filters. The optical filters are custom-made according to the specification of the optimal three-band combination. The imaging system acquires a three-channel image for the purpose of object detection. As the optical filters are of widebands, the three-channel imaging system can acquire high-quality three-channel images in a relatively short exposure time.

C. Wideband Image Simulation

We use a hyperspectral imaging system to acquire narrowband images in the SWIR spectrum for wideband image simulation. According to the imaging model [41], the ith narrowband image can be formulated as

$$I_i(p) = \sum_{\lambda} l(\lambda)\rho(\lambda, p)t_i(\lambda)s(\lambda)\Delta\lambda \tag{1}$$

where $l(\lambda)$ denotes the spectral energy distribution of the light source, $\rho(\lambda, p)$ denotes the spectral reflectance at pixel p, $t_i(\lambda)$ denotes the spectral transmittance of the ith filter, and $s(\lambda)$ denotes the spectral sensitivity of the sensor.

In this work, we assume that the candidate wideband images for optimal band selection can be simulated from narrowband spectral images using a simple linear model. This assumption is based on two observations. First, the spectral bands of the hyperspectral imaging system cover the whole SWIR spectrum. Second, the spectral transmittances of adjacent narrowband filters overlap in the spectrum. Specifically, we simulate the wideband image $\widetilde{I}$ by

$$\widetilde{I}(p) = \sum_i w_i \cdot I_i(p) \tag{2}$$

where p represents pixel position, I_i denotes the ith real narrowband image, and w_i is the corresponding weight.

We consider $l(\lambda)$ and $s(\lambda)$ to be constant during the imaging process. Denoting $m(\lambda) \triangleq l(\lambda)s(\lambda)$, the imaging model (1) can be simplified as

$$I_i(p) = \sum_{\lambda} \rho(\lambda, p)t_i(\lambda)m(\lambda)\Delta\lambda. \tag{3}$$

Then, (2) becomes

$$\begin{aligned}\widetilde{I}(p) &= \sum_i w_i \cdot \left(\sum_{\lambda} \rho(\lambda, p)t_i(\lambda)m(\lambda)\Delta\lambda\right) \\ &= \sum_i \sum_{\lambda} w_i\rho(\lambda, p)t_i(\lambda)m(\lambda)\Delta\lambda \\ &= \sum_{\lambda} \rho(\lambda, p)\left(\sum_i w_i t_i(\lambda)\right)m(\lambda)\Delta\lambda. \end{aligned} \tag{4}$$

By denoting

$$\widetilde{t}(\lambda) = \sum_i w_i t_i(\lambda) \tag{5}$$

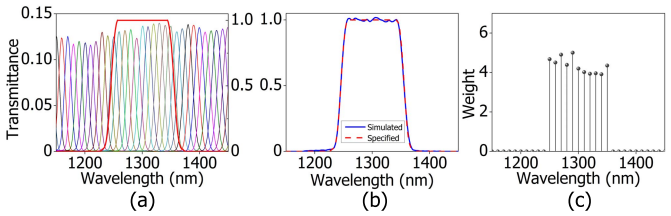

Fig. 3. Illustration of wideband simulation. (a) Spectral transmittances of the LCTF narrowband filters and a target wideband filter. (b) Comparison of the target and simulated spectral transmittances. The IoU of these two transmittances is 0.965. (c) Weights of narrowband filters computed using (8).

(4) can be rewritten as

$$\widetilde{I}(p) = \sum_{\lambda} \rho(\lambda, p)\widetilde{t}(\lambda)m(\lambda)\Delta\lambda. \tag{6}$$

This means that the simulated wideband image $\widetilde{I}$ can be regarded as an image acquired using a virtual filter with spectral transmittance $\widetilde{t}(\lambda)$.

Consequently, to generate a specific wideband image $\widetilde{I}$, our aim becomes simulating $\widetilde{t}(\lambda)$ using the linear combination of the spectral transmittances $t_i(\lambda)$ of narrowband filters, as shown in Fig. 3. To simplify notation and computation, we rewrite (5) in its vector–matrix form as

$$\widetilde{\mathbf{t}} = \mathbf{T}\mathbf{w} \tag{7}$$

where $\widetilde{\mathbf{t}}$ denotes the spectral transmittance vector of virtual filters, $\mathbf{T}$ denotes the spectral transmittance matrix of the LCTF, and $\mathbf{w} = (w_1, w_2, \ldots, w_n)^{\mathsf{T}}$. The weight vector $\mathbf{w}$ should satisfy two practical conditions. First, with the computed weight $\mathbf{w}^*$, the simulated wideband spectral transmittance $\mathbf{T}\mathbf{w}^*$ should be nonnegative. Second, the wideband transmittance can be simulated by a limited number of narrowband ones. Hence, we compute the weight as follows:

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmin}} \ \|\widetilde{\mathbf{t}} - \mathbf{T}\mathbf{w}\|_1 \quad \text{s.t. } \mathbf{w} \succeq 0. \tag{8}$$

The constraint $\mathbf{w} \succeq 0$ ensures the nonnegativity of the simulated wideband transmittance, and the ℓ_1-norm $\|\cdot\|_1$ ensures the sparsity of the solution [see Fig. 3(c)]. With the computed weight $\mathbf{w}^*$, the wideband image can be simulated according to (2).

With the above simulation, we generate 1310 wideband images for each scene. In this work, we use $N = 200$ scenes, and thus, our dataset has 262 000 synthetic images for band selection. The bandwidths of virtual filters range from 50 to 240 nm with a step size of 10 nm, considering the image SNR and spectrum utilization. The center wavelengths vary with a step size of 5 nm, while the rising and falling edges of virtual filters are kept within the SWIR spectrum. For simplification, the transmittance of each virtual filter is designed to be in an approximately uniform shape. The middle part of spectral transmittance keeps constant, and the rising and falling edges are of Gaussian shape where the standard deviation σ is set following the FWHM of LCTF. The purpose of this arrangement is to make the simulation of edges more appropriate. We use root-mean-squared error (RMSE) and intersection over union (IoU) to evaluate the accuracy of simulated transmittance. The mean IoU is 0.946 and the mean RMSE is 0.086. These values verify the high similarity between real and simulated transmittances. Fig. 4 shows an example of a simulated wideband image, indicating that the simulated image is close to the real one.

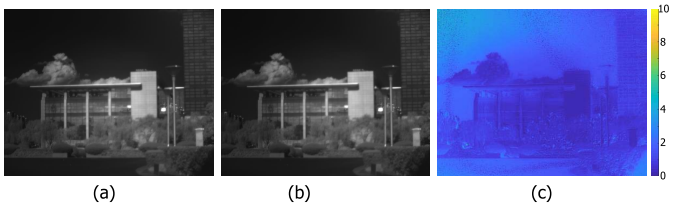

Fig. 4. Illustration of wideband image simulation. (a) Real image acquired using an optical filter with the target transmittance. (b) Simulated wideband image using a linear combination of narrowband images. (c) Difference map of real and simulated images. Average intensity error is 3.8%.

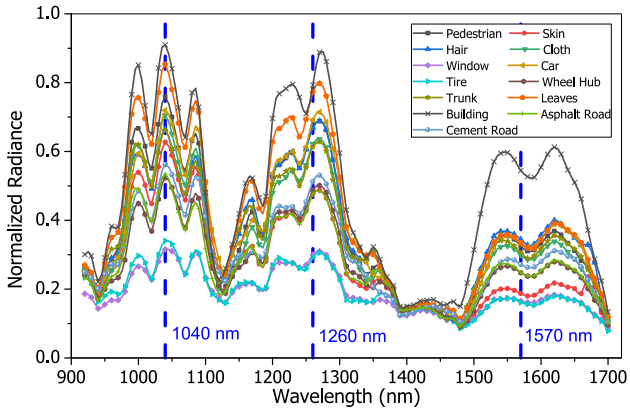

Fig. 5. Normalized radiance of some objects and common backgrounds, computed from hyperspectral images acquired by LCTF. The spatially averaged spectral radiance [42] of each object is normalized by the maximal radiance of an InfraGold whiteboard. The three dashed lines indicate the significant bands determined using the procedure introduced in Section III-B.

D. Band Selection Strategy

Spectral features are the fundamental characteristics of natural objects, and the spectral difference can benefit object detection. In some bands, the difference between spectral radiance of background (building, plant, and road) and objects (e.g., pedestrian and vehicle) is significant, as shown in Fig. 5. By exploiting spatial and spectral information in the wideband images, we select the three most representative bands that are optimal for object detection.

In this work, we adopt an object detection-oriented strategy that selects bands according to detection precision. To quantitatively guide the band selection, we introduce a specific measure related to object detection, namely, RL score. Given a three-channel image as input, the detector produces the predicted confidence and location of each object in the image. Confidence assesses the detection accuracy from the perspective of recognition. IoU of predicted and ground-truth locations evaluates the localization precision. The confidence and IoU

both range from 0 to 1, with a larger value representing higher accuracy. To directly connect object detection and band selection, we define the RL score of any three-band combination of the nth scene as

$$S'_n(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3) = \frac{1}{K_n} \sum_{k=1}^{K_n} c_k \cdot r_k \tag{9}$$

where $(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$ denotes a three-band combination. $\mathbf{b}_i$ represents the spectral transmittance of the ith band, which is a vector with massive elements. K_n denotes the number of objects of interest in the nth scene, and thus, the normalization using $1/K_n$ in RL score guarantees the same contribution of different scenes. For the kth object, its confidence is denoted as c_k and its IoU as r_k. The c_k is set to zero in case of misclassification or detection missing. We note that the RL score defined in (9) naturally incorporates the spectral and spatial information in the context of object detection.

A larger RL score means a better object detection precision. Band selection can be cast as the problem that finds the three bands $(\mathbf{b}_1^*, \mathbf{b}_2^*, \mathbf{b}_3^*)$ with maximal score

$$\left(\mathbf{b}_1^*, \mathbf{b}_2^*, \mathbf{b}_3^*\right) = \arg \max_{(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)} \frac{1}{N} \sum_{n=1}^{N} S'_n(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3) \tag{10}$$

where N denotes the number of scenes.

However, the direct use of transmittance will result in a huge computation burden. To cope with this issue, we use bandwidth Δ and center wavelength λ to simplify computation since the transmittance of simulated bands is designed in an approximate uniform shape that can be well described by these two parameters. Then, the RL score is of the form

$$S_n(\theta_1, \theta_2, \theta_3) = \frac{1}{K_n} \sum_{k=1}^{K_n} c_k \cdot r_k \tag{11}$$

where $\theta_i \triangleq (\Delta_i, \lambda_i)$ denotes the parameter pair of the ith band. Accordingly, the problem of finding optimal three bands is reformulated as

$$\left(\theta_1^*, \theta_2^*, \theta_3^*\right) = \arg \max_{(\theta_1, \theta_2, \theta_3)} \frac{1}{N} \sum_{n=1}^{N} S_n(\theta_1, \theta_2, \theta_3). \tag{12}$$

III. Details of Optimal Band Selection

As mentioned in Section II, the desired bands are optimally selected using a dataset with plenty of synthetic wideband images. Suppose that the dataset consists of N scenes, each with M synthetic wideband images. The exhaustive band searching process using all these images will be extremely time-consuming due to the computationally intensive operation of object detection. In the case that $N = 200$, $M = 1310$, and each object detection operation costs 0.04 s, the total computation time for all the $N \cdot \mathbf{C}_M^3 = 7.48 \times 10^{10}$ images would be around 19 years even using six graphic processing units (GPUs) NVIDIA RTX2080Ti. This is obviously impractical for real applications.

To deal with this problem, we introduce a two-stage algorithm for band selection. It is based on the assumption that the optimal three-band combinations would produce high RL scores on most scenes. In the first stage, we collect distinctive three-band combinations for individual scenes. In the second stage, we compute the distributions of parameters λ_i and Δ_i, $1 \le i \le 3$, and efficiently determine the optimal combination by narrowing the searching ranges from the distributions. With this treatment, the selection process can be accomplished in three days using six GPUs, which is much shorter than the exhaustive searching process.

A. Stage I: Selecting Bands for Individual Scenes

In the first stage, we select the well-performed three-band combinations from individual scenes. By adopting a threshold δ, we sequentially determine the three bands according to the RL score. More specifically, for each scene, we start with selecting the first band θ_i, then select the second band θ_j based on θ_i, and finally select the third band θ_k based on both θ_i and θ_j. The band overlap is restricted to be less than half of the smallest bandwidth. For the computation of RL score, we construct the three-channel SWIR images according to the specified bands, with band duplicating when necessary. The detailed band selection procedure for the nth scene is given as follows.

Step 1: We calculate the RL scores $S_n^{(1)}(\theta_i)$ for each single band θ_i and find the maximal score $S_{\max}$. Then, we form the single-band set $\Phi_n^{(1)}$ by collecting the significant single bands whose normalized scores $S_n^{(1)}(\theta_i)/S_{\max}$ are larger than the threshold δ.

Step 2: We fix the first band $\theta_i \in \Phi_n^{(1)}$ and compute the RL scores for each two-band combination (θ_i, θ_j). Similar to the computation of $\Phi_n^{(1)}$, we construct the two-band set $\Phi_n^{(2)}$ by collecting the band combinations whose normalized scores are larger than the threshold δ.

Step 3: For each two-band combination $(\theta_i, \theta_j) \in \Phi_n^{(2)}$, we could find the three-band combination $(\theta_i, \theta_j, \theta_k)$ with the maximal RL score. We construct the final three-band set $\Phi_n^{(3)}$ by collecting all these optimal three-band combinations.

With the above three steps, we can obtain the three-band set $\Phi_n^{(3)}$ for the nth scene. Considering that there are totally N scenes in the image dataset, the full three-band set is constructed as $\Phi^{(3)} = \{\Phi_1^{(3)}, \Phi_2^{(3)}, \ldots, \Phi_N^{(3)}\}$. In this work, we empirically set $\delta = 0.95$ and obtain $\Phi^{(3)}$ containing $C_0 = 172\,319$ three-band combinations.

B. Stage II: Selecting Final Bands

In the second stage, we compute the 1-D distributions of individual parameters (λ_i and Δ_i, $1 \le i \le 3$) in the full three-band set $\Phi_n^{(3)}$, assuming that the parameters are independent. In this way, the searching ranges of parameters can be considerably reduced when compared to the original 6-D searching space of $\Phi_n^{(3)}$.

Fig. 6 shows the distributions of the parameters, each normalized with respect to its corresponding maximal count of occurrence. It is observed that the distributions of center wavelengths λ_i, $1 \le i \le 3$, form obvious peaks. The distributions of bandwidths Δ_i ($1 \le i \le 3$) form plains rather than obvious peaks. This means that bandwidths are

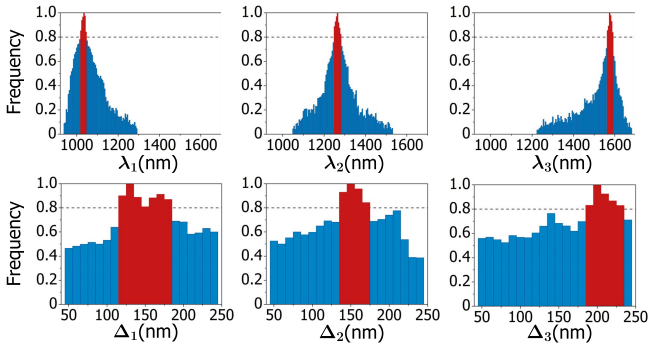

Fig. 6. Distributions of parameters λ_i and Δ_i $(1 \le i \le 3)$ of the three-band combinations in $\Phi^{(3)}$. The dashed lines indicate the positions specified by threshold $\eta = 0.8$. The searching ranges are marked in red. Note that the histograms corresponding to λ_1 and λ_3 are skewed. This is because the image counts near the two wavelength ends (920 and 1700 nm) are less than those in the middle range due to the setting of bandwidths.

Algorithm 1 Selecting the Optimal Three-Band Combinations

Input: Data of N scenes (M images per scene), band parameters $\Theta = \{\theta_1, \theta_2, \ldots, \theta_M\}$.
Output: The optimal three-band set $(\theta_i^*, \theta_j^*, \theta_k^*)$.
for $n = 1$ **to** N **do**
  Compute $\Phi_n^{(1)}$;
  Compute $\Phi_n^{(2)}$;
  Compute $\Phi_n^{(3)}$;
end
Compute the set $\Phi^{(3)} = \{\Phi_1^{(3)}, \Phi_2^{(3)}, \ldots, \Phi_N^{(3)}\}$;
Compute the distributions of λ_1, λ_2, λ_3, Δ_1, Δ_2, and Δ_3 from $\Phi^{(3)}$;
Compute $(\theta_i^*, \theta_j^*, \theta_k^*)$ using (13).

of lower sensitivity to the RL score when compared to center wavelengths.

We use a threshold η to restrict the searching ranges of parameters according to their distributions. Only the parameters with frequencies larger than η, marked in red in Fig. 6, are considered in the band selection process. Denoting the searching range of λ_i as $\Omega(\lambda_i)$ and that of Δ_i as $\Omega(\Delta_i)$, the band selection can be formulated as

$$\left(\theta_1^*, \theta_2^*, \theta_3^*\right) = \arg \max_{(\theta_1,\theta_2,\theta_3)} \frac{1}{N} \sum_{n=1}^{N} S_n(\theta_1, \theta_2, \theta_3) \qquad (13)$$

where $\lambda_i \in \Omega(\lambda_i)$ and $\Delta_i \in \Omega(\Delta_i)$, $1 \le i \le 3$. In this work, we empirically set $\eta = 0.8$, and consequently, the total number of three-band combinations reduces to $C = 35\,280$, only around 20% of C_0.

C. Algorithm Summary

To clarify, we summarize the algorithm for optimal three-band selection in Algorithm 1.

IV. Experiments

In the following, we first present the prototypes of the hyperspectral and three-channel imaging systems and introduce the acquired image dataset. We then discuss the results of our three-band selection algorithm. Finally, we demonstrate the effectiveness of our imaging system based on the evaluation of image quality and object detection in haze.

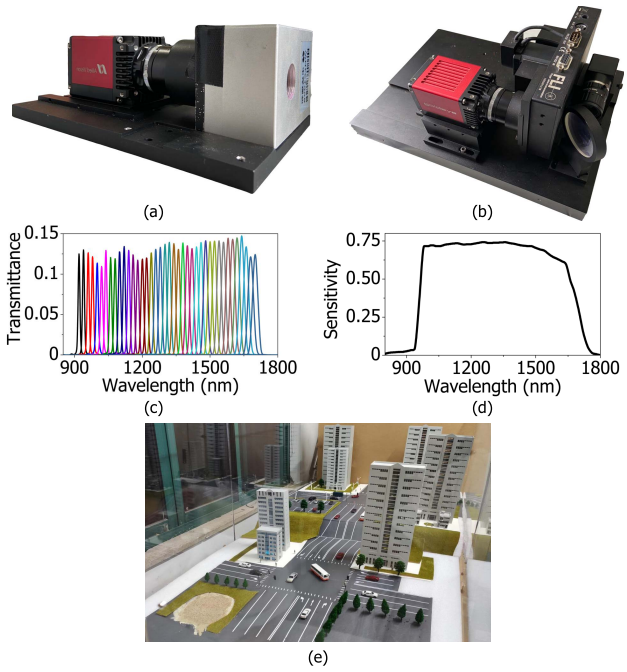

Fig. 7. Two SWIR imaging systems. (a) Hyperspectral imaging system consisting of an SWIR camera and an LCTF. (b) Three-channel imaging system consisting of an SWIR camera and a filter wheel installed with optimal optical filters. By adopting a beam splitter, RGB images can also be acquired using a color camera. (c) Spectral transmittances of the narrowband filters of LCTF. (d) Spectral sensitivity of the SWIR camera. (e) One of the toy scenes for image dataset collection.

To validate the effectiveness of our imaging system, we compare three-channel SWIR images with their original and dehazed RGB counterparts in terms of image quality. We employ the dark channel prior (DCP) algorithm[1] [18] and DehazeNet[2] [19] for image dehazing. We also employ the spectral edge (h) algorithm[3] [16] to fuse the RGB and panchromatic SWIR images, whose aim is to improve the scene visibility of RGB images.

A. Device and Dataset

Fig. 7 shows the hyperspectral and three-channel imaging systems that work in the SWIR spectrum from 920 to 1700 nm. The spectral transmittances of the LCTF (Wayho LCTF-S10) are of narrowband whose FWHMs are between 15 and 25 nm. The spectral sensitivity of the SWIR camera (AVT Goldeye-033) is relatively high in the majority of the spectrum range. As mentioned in Section II, the acquired narrowband SWIR images are used to simulate the wideband ones for optimal three-band selection, based on which the

[1]Source code available at https://github.com/noise-margin/Dehaze-Algorithms
[2]Source code available at https://github.com/caibolun/DehazeNet
[3]Source code available at https://github.com/entropyzeroo/ImageFusion/tree/master/SE

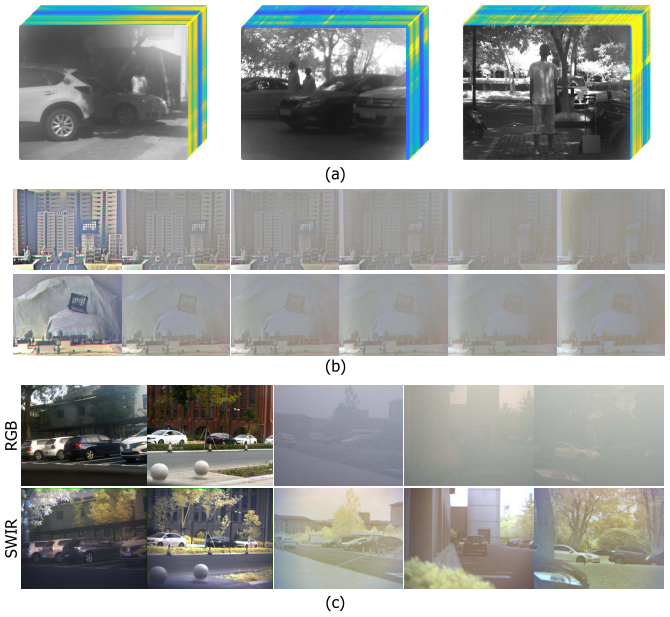

Fig. 8. Example images in the dataset. (a) SWIR hyperspectral images. (b) Three-channel SWIR images of two toy scenes. (c) RGB images and three-channel SWIR images of natural scenes. In (b), images in the first column are haze-free, while those in other columns are acquired under different haze levels from thin to thick. In (c), images in the first two columns are acquired under clear weather, while those in other columns are acquired in haze.

three-channel imaging system can acquire high-quality three-channel images with a filter wheel (FLI HS-625).

For the sake of image quality comparison, we install an additional RGB camera in the three-channel SWIR imaging system. By employing a beam splitter, the RGB and SWIR cameras can acquire scene images at the same viewpoint. We further align the RGB and three-channel SWIR images on a pixel-to-pixel basis using a homography transformation. To remove color cast, we adjust the brightness of SWIR images such that the average intensities of individual channels are identical.

We set up several toy scenes in our laboratory for image dataset collection, with one example shown in Fig. 7(e). The scenes include 13 urban and 11 rural ones, all placed in an acrylic container of size $0.8 \times 1 \times 2.4$ m^3 during the imaging process. A fog machine is used to generate hazy conditions from thin to thick by controlling the spray time. Each toy scene has a clear image and hazy images under five levels of haze, as shown in Fig. 8(b).

Our image dataset contains three types of images, i.e., SWIR hyperspectral images, three-channel SWIR images, and RGB images, acquired under clear or haze conditions. Table I shows the properties of the SWIR hyperspectral and three-channel images. A hyperspectral image comprises 79 narrowband images acquired using LCTF, while a three-channel image consists of three wideband images acquired using custom-made optical filters. The acquisition time of a hyperspectral image is longer than that of a three-channel image. This is because a large number of narrowband images need to be sequentially acquired for a hyperspectral image and also

TABLE I

PROPERTIES OF THE SWIR HYPERSPECTRAL AND THREE-CHANNEL IMAGES ACQUIRED USING OUR SYSTEMS

	Hyperspectral image	Three-channel image
Total scenes	200	500
Hazy scenes	80	300
Filters	LCTF	Optical
Band width	Narrow-band	Wide-band
SWIR channels	79	3
Acquisition time (s/scene)	$\approx$12	$\leq$0.1
Image size (pixels)	512 $\times$ 640	512 $\times$ 640

TABLE II

IMAGES USED IN THE EXPERIMENTS

Experiments	Images	
	Format	Property
Optimal Band Selection	Wide-band SWIR	Simulated
Evaluation of Image Quality	3-band SWIR, RGB	Real Toy Scenes
Comparison of Detectors	3-band SWIR, RGB	Real-World Scenes
Object Detection in Haze	3-band SWIR, RGB	Real-World Scenes

that the spectral transmittance of LCTF is relatively low. Fig. 8 shows some example images in the dataset. It is observed that, under clear weather, the sharpness of three-channel SWIR images is close to that of RGB images, despite some color shift. Under the hazy condition, the three-channel SWIR images are still of high contrast, while the RGB images become very dim.

Table II lists the images used in the following four experiments. The experiment *optimal band selection* uses the wide-band images simulated using hyperspectral SWIR images. The experiment *evaluation of image quality* employs the real toy scene images for the quantitative evaluation of image quality in haze. The experiments *comparison of detectors* and *object detection in haze* use real-world RGB and SWIR images to validate the performance of object detection.

B. Optimal Band Selection

As noted in Section III-B, the total number of three-band combinations is $C = 35\,280$ after restricting the parameter searching ranges. The optimal three-band combinations are determined using Algorithm 1 under the criterion of RL scores of YOLOv3 [13], each of which is computed from $N = 200$ scenes. Fig. 9 shows that the mean RL scores of the combinations are in the range $(0.76, 0.89)$, and the means and variances of RL scores are negatively correlated. This implies that the top-ranked combinations can perform comparatively well on various scenes, which is desired for practical applications.

Table III lists the parameters of the Top-5 combinations. It is observed that the center wavelengths of individual filters are very close and the bandwidths are exactly the same. Accordingly, their corresponding RL scores are very close, all above 0.885. Based on this observation, we design and manufacture

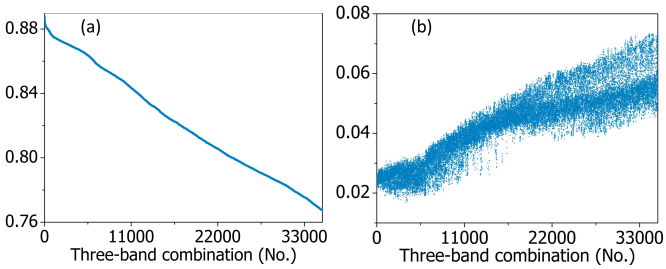

Fig. 9. Evaluation of the total 35 280 three-band combinations. (a) Mean RL scores of all three-band combinations, sorted in descending order. (b) Variances of the RL scores of the sorted three-band combinations.

TABLE III
PARAMETERS OF TOP-5 THREE-BAND COMBINATIONS

Top.	1	2	3	4	5
λ_1 (nm)	1040	1045	1040	1045	1040
λ_2 (nm)	1260	1260	1260	1260	1265
λ_3 (nm)	1570	1570	1575	1575	1575
Δ_1 (nm)	120	120	120	120	120
Δ_2 (nm)	160	160	160	160	160
Δ_3 (nm)	210	210	210	210	210
Mean RL score	0.8877	0.8867	0.8860	0.8857	0.8855
Variance ($\times 10^{-2}$)	2.08	2.25	2.17	2.23	2.18

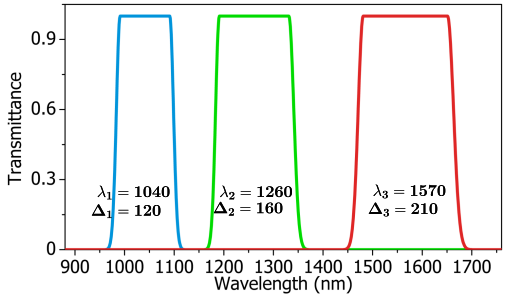

Fig. 10. Spectral transmittances of the Top-1 three-band combination. The transmittances are used in designing and manufacturing optical filters, which are installed in the three-channel SWIR imaging system.

three optical filters for the SWIR three-channel imaging system, using the parameters of the Top-1 combination (1040, 1260, and 1570 nm). Fig. 10 shows the corresponding spectral transmittances of the filters. Fig. 5 shows that these bands are the most distinctive ones for a wide range of objects, not limited to pedestrians and vehicles. This indicates that although the optimal bands are selected under the guidance of our specific RL score, they will also work well for common object detection.

We compare our three-band selection algorithm with the state-of-the-art approaches, including OCF[4] [34], E-FDPC[5] [35], ONR[6] [36], FNGBS[7] [37], and CR [38] in the object detection tasks. For CR, we first remove the continuum of

[4]Source code available at https://github.com/tanmlh/Optimal-Clustering-Framework-for-Hyperspectral-Band-Selection

[5]Source code available at https://github.com/senjia1980/EFDPC

[6]Source code available at https://github.com/tanmlh/Optimal-Neighboring-Reconstruction-for-Hyperspectral-Band-Selection

[7]Source code available at https://github.com/qianngli/FNGBS

TABLE IV
SELECTED BANDS AND MAP VALUES OF DIFFERENT BAND SELECTION ALGORITHMS

Method	Center Wavelength of Selected Bands (nm)			mAP↑
E-FDPC [35]	1270	1450	1530	0.806
CR [38]	1000	1040	1270	0.798
ONR [36]	1010	1460	1600	0.768
OCF [34]	990	1470	1630	0.749
FNGBS [37]	930	1270	1630	0.787
Ours	1040	1260	1570	**0.821**

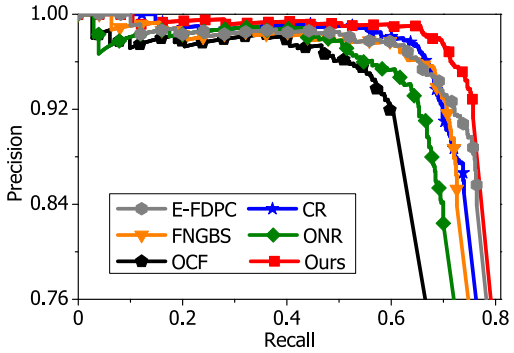

Fig. 11. Precision–recall curves of object detection for OCF [34], E-FDPC [35], ONR [36], FNGBS [37], CR [38], and our three-band selection algorithm.

the spectral radiance of different objects and then select three bands with the largest spectral differences. Table IV lists the selected bands by these algorithms and the corresponding mean average precision (mAP) of object detection. It is observed that our algorithm produces the best mAP. Fig. 11 shows the precision–recall curves of the different band selection algorithms. Both the mAP and the precision–recall curves indicate that our algorithm outperforms the others.

The detailed analysis shows that the clustering-based algorithms, OCF [34] and E-FDPC [35], select a band located at the atmospheric absorption area (near 1470 nm) that has little discriminative capability. ONR [36] is also affected by the absorption area during band selection. FNGBS [37] selects bands with a significant spectral difference, though it is not optimal for object detection. CR [38] selects the bands according to the spectral characteristics of the objects. The two selected bands (1000 and 1040 nm) are close and possess duplicating spectral information, and consequently, its detection precision is limited. In comparison, our algorithm produces the highest precision due to its capability of exploiting both spatial and spectral information under the guidance of the RL score.

C. Evaluation of Image Quality in Haze

The three-channel SWIR images are compared with RGB images and their dehazed and fused counterparts, both qualitatively and quantitatively. Fig. 12 shows the SWIR and RGB images of two urban and two rural toy scenes under clear

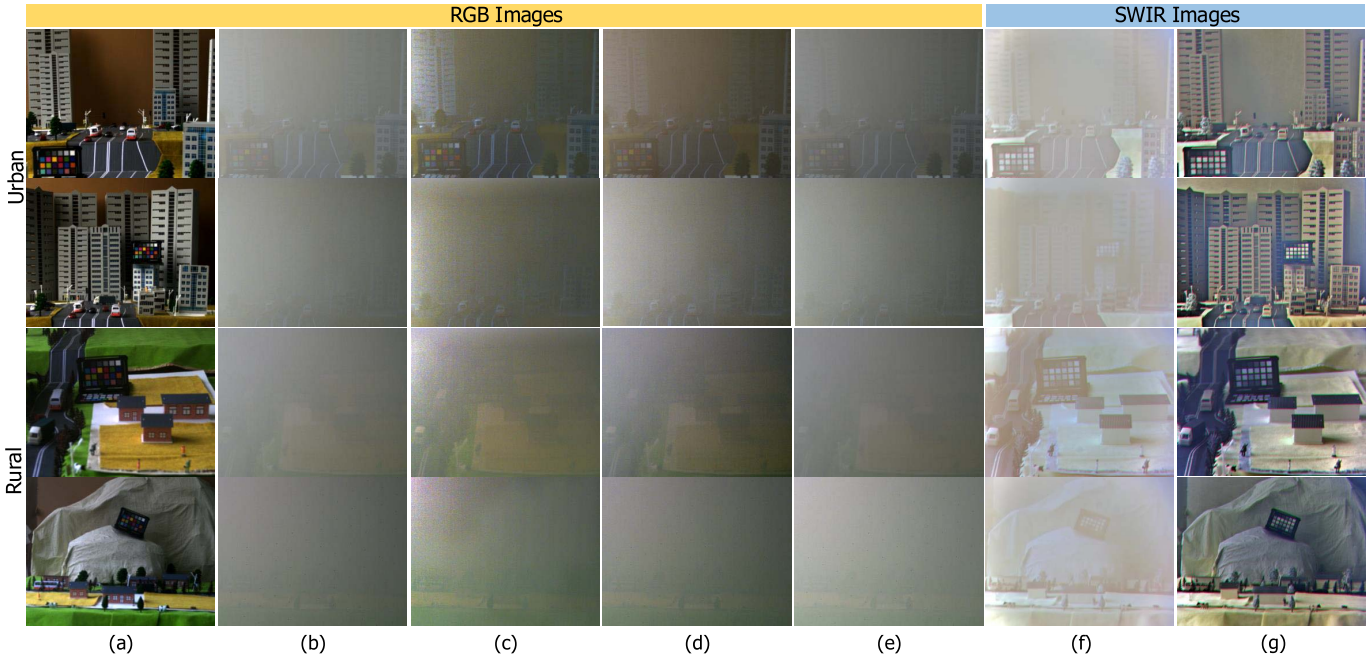

Fig. 12. Images of toy scenes. The first two rows are urban scenes. The last two rows are rural scenes. (a) RGB ground truth. (b) Original hazy RGB images. (c) Dehazed RGB images using DCP algorithm [18]. (d) Dehazed RGB images using DehazeNet [19]. (e) RGB images after SpE fusion [16]. (f) Hazy three-channel SWIR images. (g) SWIR ground truth.

TABLE V
AVERAGE PSNR, SSIM, RMSE, SAM, AND GMSD VALUES OF RGB AND SWIR IMAGES OF THE URBAN AND RURAL TOY SCENES

	Urban toy scenes					*Rural* toy scenes				
	PSNR↑	SSIM↑	RMSE↓	SAM↓	GMSD↓	PSNR↑	SSIM↑	RMSE↓	SAM↓	GMSD↓
Original RGB images	18.95	0.473	28.77	13.18	0.0983	19.04	0.489	28.48	14.66	0.165
DCP [18]	18.77	0.489	29.38	12.04	0.0651	18.85	0.513	29.11	12.33	0.139
DehazeNet [19]	19.56	0.519	26.83	13.17	0.0798	20.17	0.527	25.01	11.73	0.142
SpE fusion [16]	19.23	0.517	27.86	11.12	0.0915	19.21	0.508	27.93	12.40	0.157
Three-channel SWIR images	**25.02**	**0.774**	**14.31**	**7.13**	**0.0195**	**24.42**	**0.793**	**17.20**	**7.37**	**0.0422**

and haze conditions. It is observed that the RGB images are of low contrast even after dehazing [18], [19] or image fusion [16]. In comparison, the three-channel SWIR images are less affected by haze and the structural details are clearly visible.

Table V lists the average PSNR, SSIM, RMSE, SAM, and GMSD values of RGB and SWIR images of 13 urban and 11 rural toy scenes. In metric computation, we use the images acquired under clear weather as ground truths. It is observed that the three-channel SWIR images are of higher quality compared to all the RGB ones.

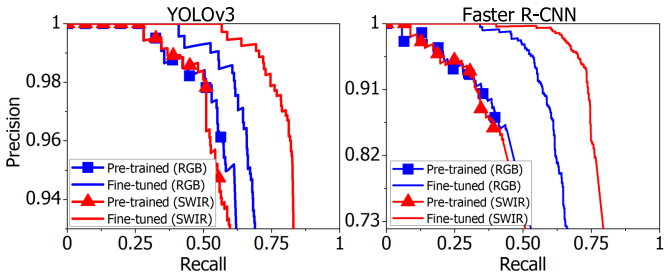

Fig. 13. Precision–recall curves of different detectors, either pretrained or fine-tuned, on our three-channel image dataset.

D. Comparison of Different Detectors

We evaluate the performance of YOLOv3[8] [13] and Faster R-CNN[9] [11] on our three-channel data. YOLOv3 has been employed for optimal band selection, and Faster R-CNN is deployed to validate whether our three-channel SWIR imaging system also performs well when using other detectors. Based on the pretrained weights, we fine-tune the detectors using RGB-SWIR images of 200 scenes (100 clear and 100 hazy ones) and test the detectors on the other 100 clear and 100 hazy scenes.

Fig. 13 shows the precision–recall curves of YOLOv3 and Faster R-CNN on RGB and SWIR images. YOLOv3

[8]Source code available at https://pjreddie.com/darknet/yolo
[9]Source code available at https://github.com/endernewton/tf-faster-rcnn

TABLE VI
MAP OF DIFFERENT DETECTORS ON RGB AND SWIR IMAGES UNDER CLEAR AND HAZY CONDITIONS

Detectors		YOLOv3 [13]				Faster-RCNN [11]			
Images		RGB		SWIR		RGB		SWIR	
		Clear	Hazy	Clear	Hazy	Clear	Hazy	Clear	Hazy
mAP	Pre-trained	0.690	0.485	0.675	0.543	0.521	0.316	0.508	0.347
	Fine-tuned	0.793	0.560	0.833	0.792	0.728	0.497	0.743	0.696

and Faster R-CNN produce a similar performance. On RGB images, the pretrained detectors produce slightly higher precision than on the SWIR images since the pretrained detectors are trained using RGB images. After fine-tuning the detectors, the precision on SWIR images is significantly higher than that on RGB images.

Table VI presents the mAP of YOLOv3 and Faster R-CNN under clear and hazy conditions. It is observed that both detectors produce higher mAP after fine-tuning, especially on those hazy images. Specifically, the pretrained detectors obtain a higher mAP on SWIR images than on RGB ones since SWIR images have more details for object detection under hazy conditions. Furthermore, the margin of the precision on SWIR and RGB images is enlarged after fine-tuning the detectors.

E. Object Detection on Hazy and Dehazed Images

We evaluate object detection on 300 three-band images acquired in real-world hazy scenes. As different detectors perform similarly on our dataset, hereafter, we employ the fine-tuned YOLOv3 to detect objects in RGB and SWIR images.

We evaluate the detection performance on both hazy and dehazed SWIR images. The DCP [18] dehazing algorithm is employed to process our three-channel SWIR images. To meet the condition of DCP that at least one color channel has very low intensity at some pixels, we preprocess the hazy SWIR image $\mathbf{H}$ as

$$\widetilde{\mathbf{H}} = \mathbf{H} - b \cdot \mathbf{1} \tag{14}$$

where b denotes the minimal element of $\mathbf{H}$. Then, we compute the dehazed image using the DCP algorithm as $\widetilde{\mathbf{J}} = \mathrm{DCP}(\widetilde{\mathbf{H}})$. Considering that $\widetilde{\mathbf{J}}$ appears quite dark due to the preprocessing (14), we compute the final dehazed SWIR image $\mathbf{J}$ by normalization

$$\mathbf{J} = \frac{\widetilde{\mathbf{J}}}{\widetilde{J}_m} \tag{15}$$

where $\widetilde{J}_m$ is the maximal element of $\widetilde{\mathbf{J}}$.

As object detection in panchromatic infrared images is quite common [22], [23], we also evaluate the detection performance on panchromatic SWIR images acquired in the SWIR spectrum ranging from 920 to 1700 nm. To meet the input requirement of detectors, we replicate the single-channel panchromatic SWIR image three times to form a three-channel image for object detection.

Fig. 14 shows the precision–recall curves of SWIR and RGB images. It is observed that the dehazing and image fusion

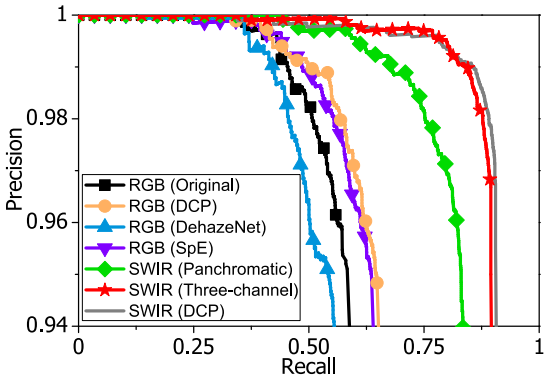

Fig. 14. Precision–recall curves of object detection for RGB and SWIR images. RGB images are the original ones and the processed ones by DCP [18] dehazing, DehazeNet [19], and SpE fusion [16]. SWIR images include the panchromatic images, three-channel images, and the dehazed three-channel images using DCP [18].

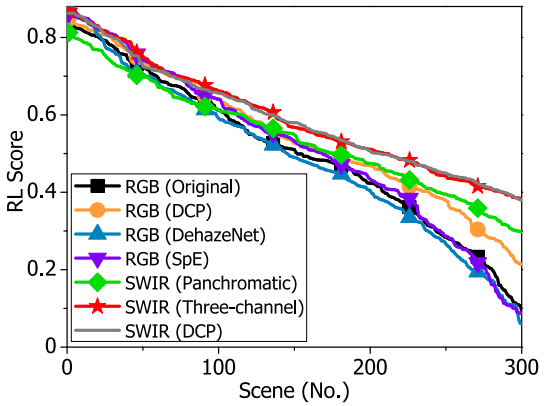

Fig. 15. RL scores of RGB and SWIR images. RGB images are the original ones and the processed ones by DCP dehazing [18], DehazeNet [19], and SpE fusion [16]. SWIR images include the panchromatic ones, and the original and dehazed three-channel ones.

algorithms for RGB images only achieve slight improvements in object detection. On the contrary, the detector produces higher precision on SWIR images. Due to the extraction of distinctive spectral information, we obtain higher precision on three-channel SWIR images when compared with the panchromatic ones. It is observed that the dehazing operation is beneficial to object detection on SWIR images.

We compare the object detection accuracy under different haze densities in terms of RL score. Generally, the RL score decreases when the haze level changes from thin to thick. Fig. 15 shows the RL scores of various hazy scenes, which are sorted in descending order for ease of comparison. It is observed that the RL scores of the three-channel SWIR images are higher than those of RGB and panchromatic SWIR images under all hazy conditions. In the case of thin haze, the RL scores of three-channel SWIR images are comparable to those of RGB images and higher than those of panchromatic SWIR images. When the haze becomes thicker, the RL scores of three-channel SWIR images decrease gradually but are always higher than those of other images. The reason is that the three-channel SWIR images have more textures than the RGB

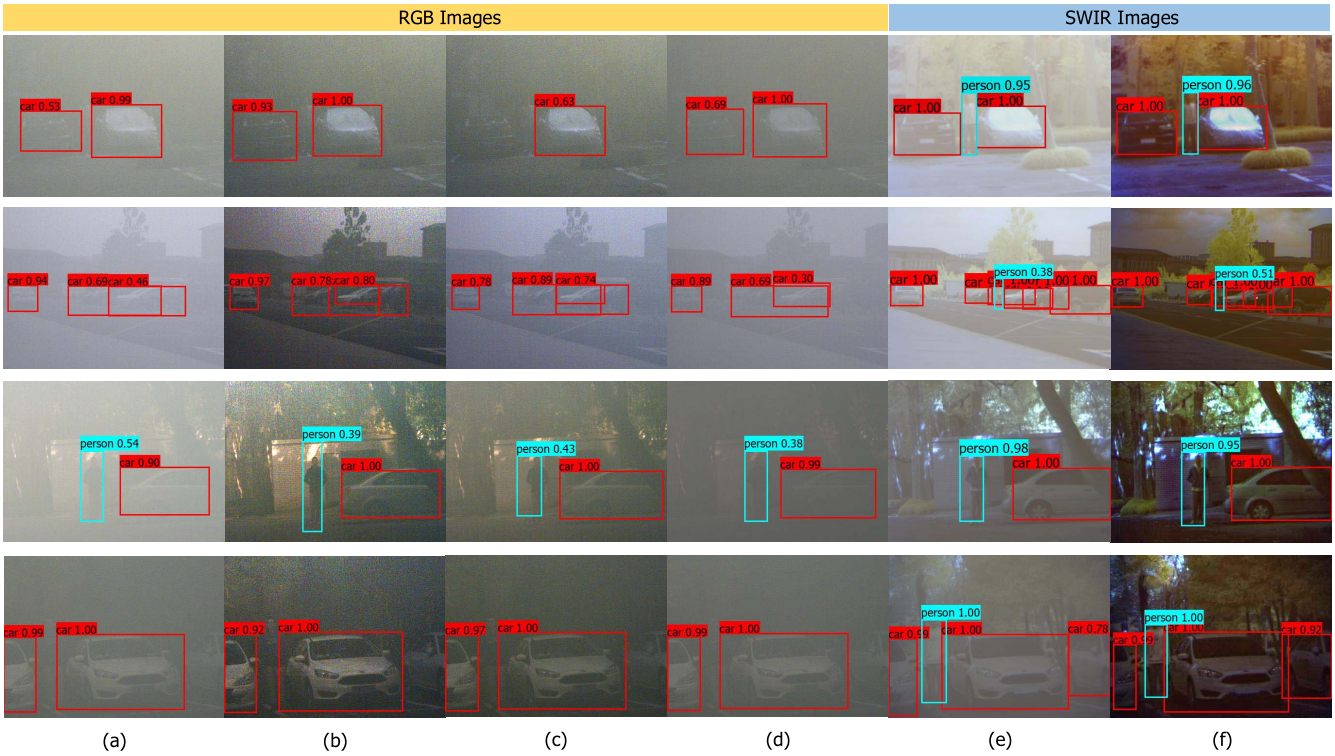

Fig. 16. Prediction of four different hazy scenes output by YOLOv3 [13]. (a) Original RGB images. (b) Dehazed RGB images using DCP [18]. (c) Dehazed RGB images using DehazeNet [19]. (d) RGB images after SpE fusion [16]. (e) Three-channel SWIR images. (f) Dehazed three-channel SWIR images using DCP [18].

TABLE VII
MAP AND AVERAGE RL SCORES OF VARIOUS SWIR IMAGES AND RGB IMAGES

Images		mAP↑	RL Score↑
RGB	Original	0.595	0.498
	DCP [18]	0.653	0.539
	DehazeNet [19]	0.579	0.506
	SpE fusion [16]	0.637	0.513
SWIR	Panchromatic	0.830	0.576
	Three-channel (Original)	**0.879**	**0.623**
	Three-channel (DCP)	**0.893**	**0.631**

ones and own more spectral information than the panchromatic ones.

Fig. 16 shows the detection results of four real-world scenes. For RGB images, the detection performance cannot be improved after either dehazing [18], [19] or image fusion [16] because the image quality has not been much improved. In comparison, the detectors identify more objects and produce higher confidence scores on original and dehazed SWIR images.

Table VII lists the mAP and the average RL scores of RGB and SWIR images of the 300 real-world hazy scenes. The mAP and RL scores of SWIR images are significantly higher than those of the RGB images. Compared with the conventional panchromatic SWIR images, we obtain better detection performance on the three-channel ones. Noticeable improvements in mAP and RL score are achieved on the three-channel SWIR images after applying DCP dehazing.

V. CONCLUSION

We have proposed a three-channel SWIR imaging system for object detection in haze with the assistance of a hyperspectral imaging system. For optimal band selection, we simulate plenty of wideband synthetic images using a number of narrowband images acquired by the hyperspectral imaging system. The optimal three-band combination is determined using an efficient searching algorithm, guided by a specific measure referred to as RL score. Experimental results validate that the three-channel SWIR images, which are acquired by our proposed imaging system, perform better than RGB images in object detection.

A limitation of our band selection algorithm is its slow computation speed since it needs to compute the RL score for all three-band combinations. In our future work, we plan to speed up the band selection process by incorporating both spectral and spatial information. In addition, we will also aim to recover the naturalness of SWIR images by learning the mapping between the SWIR and visible spectra.

REFERENCES

[1] B.-H. Chen, L.-F. Shi, and X. Ke, "A robust moving object detection in multi-scenario big data for video surveillance," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 4, pp. 982–995, Apr. 2019.
[2] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6479–6488.

[3] G. Song, K. Song, and Y. Yan, "EDRNet: Encoder–decoder residual network for salient object detection of strip steel surface defects," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 12, pp. 9709–9719, Dec. 2020.
[4] Y. He, K. Song, Q. Meng, and Y. Yan, "An end-to-end steel surface defect detection approach via fusing multiple hierarchical features," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1493–1504, Apr. 2020.
[5] Y. Tu, Z. Ling, S. Guo, and H. Wen, "An accurate and real-time surface defects detection method for sawn lumber," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 2501911.
[6] M. Braun, S. Krebs, F. Flohr, and D. Gavrila, "EuroCity persons: A novel benchmark for person detection in traffic scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 8, pp. 1844–1861, Aug. 2019.
[7] Y. Jiao, H. Yao, and C. Xu, "PEN: Pose-embedding network for pedestrian detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 3, pp. 1150–1162, Mar. 2021.
[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[9] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Oct. 2014, pp. 580–587.
[10] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
[12] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Oct. 2016, pp. 779–788.
[13] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
[14] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
[15] Y. Kudo and A. Kubota, "Image dehazing method by fusing weighted near-infrared image," in *Proc. Int. Workshop Adv. Image Technol. (IWAIT)*, Jan. 2018, pp. 1–2.
[16] D. Connah, M. S. Drew, and G. D. Finlayson, "Spectral edge image fusion: Theory and applications," in *Proc. Eur. Conf. Comput. Vis.*, Lecture Notes in Computer Science, vol. 8693. Cham, Switzerland, Springer, 2014, pp. 65–80.
[17] M. Awad, A. Elliethy, and H. A. Aly, "Adaptive near-infrared and visible fusion for fast image enhancement," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 408–418, 2020.
[18] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Sep. 2010.
[19] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: An end-to-end system for single image haze removal," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5187–5198, Nov. 2016.
[20] Z. Zhu, H. Wei, G. Hu, Y. Li, G. Qi, and N. Mazur, "A novel fast single image dehazing algorithm based on artificial multiexposure image fusion," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–23, 2021.
[21] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar, "Polarization-based vision through haze," *Appl. Opt.*, vol. 42, no. 3, pp. 511–525, Jan. 2003.
[22] L. Zhang, B. Wu, and R. Nevatia, "Pedestrian detection in infrared images based on local shape features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
[23] D.-E. Kim and D.-S. Kwon, "Pedestrian detection and tracking in thermal images using shape features," in *Proc. 12th Int. Conf. Ubiquitous Robots Ambient Intell. (URAI)*, Oct. 2015, pp. 22–25.
[24] N. Rajic, "Principal component thermography for flaw contrast enhancement and flaw depth characterisation in composite structures," *Compos. Struct.*, vol. 58, pp. 521–528, Dec. 2002.
[25] B. Yousefi, S. Sfarra, F. Sarasini, and X. P. V. Maldague, "IRNDT inspection via sparse principal component thermography," in *Proc. IEEE Can. Conf. Electr. Comput. Eng. (CCECE)*, May 2018, pp. 1–4.
[26] B. Yousefi, S. Sfarra, C. I. Castanedo, and X. P. Maldague, "Comparative analysis on thermal non-destructive testing imagery applying candid covariance-free incremental principal component thermography (CCIPCT)," *Infr. Phys. Technol.*, vol. 85, pp. 163–169, Dec. 2017.
[27] S. Hwang, J. Park, N. Kim, Y. Choi, and I. So Kweon, "MultiSpectral pedestrian detection: Benchmark dataset and baseline," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1037–1045.
[28] J. U. Kim, S. Park, and Y. M. Ro, "Uncertainty-guided cross-modal learning for robust multispectral pedestrian detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1510–1523, Mar. 2022.
[29] J. Ma, L. Tang, M. Xu, H. Zhang, and G. Xiao, "STDFusionNet: An infrared and visible image fusion network based on salient target detection," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, 2021.
[30] T. Zhi, B. R. Pires, M. Hebert, and S. G. Narasimhan, "MultiSpectral imaging for fine-grained recognition of powders on complex backgrounds," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8699–8708.
[31] C.-I. Chang, Q. Du, T.-L. Sun, and M. L. G. Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 6, pp. 2631–2641, Nov. 1999.
[32] C.-I. Chang and S. Wang, "Constrained band selection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1575–1585, Jun. 2006.
[33] S. Patra, P. Modi, and L. Bruzzone, "Hyperspectral band selection based on rough set," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5495–5503, Oct. 2015.
[34] Q. Wang, F. Zhang, and X. Li, "Optimal clustering framework for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5910–5922, Oct. 2018.
[35] S. Jia, G. Tang, J. Zhu, and Q. Li, "A novel ranking-based clustering approach for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 88–102, Jan. 2015.
[36] F. Zhang, Q. Wang, and X. Li, "Optimal neighboring reconstruction for hyperspectral band selection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2018, pp. 4709–4712.
[37] Q. Wang, Q. Li, and X. Li, "A fast neighborhood grouping method for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5028–5039, Jun. 2021.
[38] A. M. Filippi and J. R. Jensen, "Effect of continuum removal on hyperspectral coastal vegetation classification using a fuzzy learning vector quantizer," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1857–1869, Jun. 2007.
[39] B. Yousefi, S. Sojasi, C. I. Castanedo, X. P. Maldague, G. Beaudoin, and M. Chamberland, "Continuum removal for ground-based LWIR hyperspectral infrared imagery applying non-negative matrix factorization," *Appl. Opt.*, vol. 57, no. 21, pp. 6219–6228, 2018.
[40] X. Li, H. Wang, X. Li, Z. Tang, and H. Liu, "Identifying degraded grass species in inner Mongolia based on measured hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 5061–5075, Dec. 2019.
[41] H. Haneishi, T. Hasegawa, A. Hosoi, Y. Yokoyama, N. Tsumura, and Y. Miyake, "System design for accurately estimating the spectral reflectance of art paintings," *Appl. Opt.*, vol. 39, no. 35, pp. 6621–6632, Dec. 2000.
[42] L. Balick, A. Gillespie, A. French, I. Danilina, J. P. Allard, and A. Mushkin, "Longwave thermal infrared spectral variability in individual rocks," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 1, pp. 52–56, Jan. 2009.

Beinan Yu received the B.E. degree from Zhejiang University, Hangzhou, China, in 2019, where he is currently pursuing the Ph.D. degree with the College of Information Science and Electronic Engineering.

His research interests are multispectral/ hyperspectral imaging and image processing.

Yifan Chen received the B.E. degree from the Harbin Institute of Technology, Harbin, China, in 2019. He is currently pursuing the M.S. degree with the Polytechnic Institute, Zhejiang University, Hangzhou, China.

His research interests are image dehazing and image processing.

Si-Yuan Cao received the B.E. degree from Tianjin University, Tianjin, China, in 2016. He is currently pursuing the Ph.D. degree with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China.

His research interests are multispectral/multimodal image alignment and image processing.

Hui-Liang Shen received the B.Eng. and Ph.D. degrees in electronic engineering from Zhejiang University, Hangzhou, China, in 1996 and 2002, respectively.

He was a Research Associate and a Research Fellow with The Hong Kong Polytechnic University, Hong Kong, from 2001 to 2005. He is currently a Professor with the College of Information Science and Electronic Engineering, Zhejiang University. His research interests include multispectral imaging, image processing, and computer vision.

Junwei Li received the Ph.D. degree in optical engineering from the Beijing Institute of Technology, Beijing, China, in 2007.

He is currently a Professor with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China. His research interests include multispectral/hyperspectral imaging and image processing.