ORIGINAL PAPER

A high-performance approach to detecting small targets in long-range low-quality infrared videos

Chiman Kwan[1] · Bence Budavari[1]

Abstract

Since targets are small in long-range infrared (IR) videos, it is challenging to accurately detect targets in those videos. In this paper, we propose a high-performance approach to detecting small targets in long-range and low-quality infrared videos. Our approach consists of a video resolution enhancement module, a proven small target detector based on local intensity and gradient (LIG), a connected component (CC) analysis module, and a track association module known as Simple Online and Real-time Tracking (SORT) to connect detections from multiple frames. Extensive experiments using actual mid-wave infrared (MWIR) videos in range between 3500 and 5000 m from a benchmark dataset clearly demonstrated the efficacy of the proposed approach. In the 5000 m case, the F1 score has been improved from 0.936 without SORT to 0.977 with SORT.

Keywords Small target detection · Infrared videos · Connected component analysis · Long range · Video super-resolution · Deep learning

1 Introduction

Infrared (IR) videos in ground-based imagers contain a lot of background clutter and flickering noise due to air turbulence, etc. Moreover, the target size is quite small and hence, it is challenging to detect small targets from a long distance.

Small target detection for infrared images has been active in recent years [1–6]. Chen et al. [1] proposed to detect small IR targets by using local contrast measure (LCM), which is time-consuming and sometimes enhances both targets and clutters. To improve the performance of LCM, Wei et al. [2] introduced a multiscale patch-based contrast measure (MPCM). Gao et al. [3] developed an infrared patch-image (IPI) model to convert small target detection to an optimization problem. Zhang et al. [4] improved the performance of the IPI via non-convex rank approximation minimization (NRAM). Zhang et al. [5] proposed to detect small IR targets based on local intensity and gradient (LIG) properties, which has good performance and relatively low computational complexity. Recently, Chen et al. [6] proposed a new

and real-time approach for detecting small targets with sky background.

Parallel to the above small target detection activities, there are some conventional target tracking methods [7–9]. Furthermore, some target detection and classification schemes using deep learning algorithms (You Only Look Once (YOLO)) for larger objects in short-range infrared videos have been proposed in the literature [10–28]. Since YOLO uses texture information to help the detection, the use of YOLO is not very effective for long-range videos in which the targets are too small to have any discernible textures. Some of these new algorithms incorporated compressive measurements directly for detection and classification. Real-time issues have been discussed in [28].

In this paper, we summarize the investigation of an integrated approach to enhancing the performance of small target detection in long-range infrared videos. Our approach contains several modules. First, in order to improve the resolution of the low-resolution infrared videos, we propose to utilize proven video super-resolution (VSR) enhancement algorithms to improve the spatial resolution of the video. Three algorithms [29–31] were compared, and one method was proven to improve the target detection performance. Second, we propose to incorporate a proven target detector known as LIG [5] into our framework. LIG has shown good performance and computational speed in our recent paper [6]. Some customizations

✉ Chiman Kwan
   chiman.kwan@arllc.net

1   Applied Research LLC, Rockville, MD, USA

of LIG with respect to adaptive threshold selection have been carried out in this paper. Third, even with careful and robust threshold selection in LIG, there are still false positives in the detection results. Dilation within the connected component (CC) analysis module has been used to merge nearby detections together, which in turn mitigates the false positives. Through some rule analysis, the false positives have been further reduced. Finally, in order to further improve the overall detection performance, a simple and fast target association algorithm has been included in our framework. The method is known as Simple Online and Real-time Tracking (SORT) [31], which further enhances the target detection performance.

It is important to clarify the key difference between the [32] and this paper. Paper [32] applied optical flow techniques to extract target information based on motion information. Two frames separated by a certain frames are needed in [32]. The detection process then repeats across the video. In contrast, this paper focused on target detection using single frames. Although post-processing using SORT was used in both papers, the target detection modules in the two papers are completely different. Moreover, this paper is not meant to improve the results of [32], as the two papers contain quite different target detection approaches. The approach in [32] has a clear limitation in certain frames where there is no target motion. For example, if the target is moving toward the camera, then optical flow will not be able to extract any motion information. In such situations, the approach in this paper will be very useful. In other words, the two approaches are complementary.

Our contributions are as follows:

- An integrated framework for target detection in long-range videos

  We propose an integrated, flexible, and modular framework comprising video super-resolution, small target detection, connected component (CC) analysis, and target track association. We also improved the adaptive threshold selection in the LIG module.
- Unsupervised target detection approach

  For long range videos (3500 m and beyond), the target size is very small. Our proposed framework, unlike YOLO, does not require any training.
- Target detection enhancement using super-resolution images

  The objective here is to investigate whether super-resolution videos will help the small target detection performance. If super-resolution videos do help the detection, we would also like to quantify the detection performance using some well-known metrics such as precision, recall, and F1

score. Three super-resolution approaches were compared: one conventional and two deep learning. One of them was observed to yield improved target detection results.
- Extensive experiments using long-range infrared videos in a benchmark dataset known as SENSIAC [34] containing videos from 3500 to 5000 m away clearly demonstrated the efficacy of our proposed framework.

The rest of this paper is organized as follows. In Sect. 2, we will present the technical approach. Section 3 summarizes the experimental results using long-range videos from a benchmark dataset. Finally, Sect. 4 concludes the paper with some remarks and future directions.

2 Technical approach

As shown in Fig. 1, our proposed approach consists of several modules. First, since the long-range videos have low resolution, we propose to apply state-of-the-art video super-resolution algorithms to enhance the video resolution. The objective is to investigate how much performance gain one can achieve with video-super-resolution. Second, an unsupervised small target detection approach using the low intensity and gradient (LIG) [5] algorithm is applied to each frame for small target detection. Some improvements over the existing LIG method have been introduced by us. Third, since the LIG detection results may have some scattered false positives, we propose to apply connected component (CC) analysis to group neighboring pixels into clusters. Some empirical rules were formulated in this module. Finally, we propose a fast target associate algorithm known as Simple Online and Realtime Tracking (SORT) [32] to form target tracks. This step turns out to further improve the detection results.
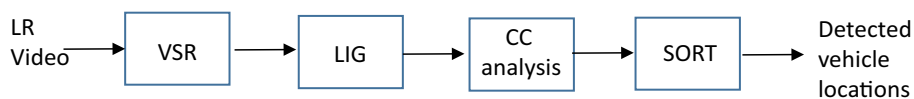
2.1 Video super-resolution algorithms

For long-range videos, the target size is quite small. Improving the resolution of the videos may help the target detection performance. In this research, we investigated three image resolution enhancement algorithms, which are briefly explained below.

2.1.1 Bicubic

Bicubic interpolation is a single frame super-resolution method. Pixels in the original frames are interpolated

Fig. 1 Proposed workflow for small target detection

LR Video → VSR → LIG → CC analysis → SORT → Detected vehicle locations

using 16 neighbors. It has been widely used in many practical applications due to its simplicity and computational efficiency.

2.1.2 Dynamic upsampling filter (DUF) algorithm for video super-resolution

Video Super Resolution using Dynamic Upsampling Filter (VSR-DUF) is a recent deep learning super resolution method developed by researchers at Yonsei University [29]. This particular model is able to utilize temporal information when generating a high-resolution frame. VSR-DUF uses frames before and after the current one in order to generate a single super resolution frame. The number of frames used before or after a given frame is known as the temporal radius. In our studies, we used a radius of 7. This is advantageous over single image super resolution methods because there are more relevant frames to draw information from.

### 2.1.3 Zoom slow-motion (ZSM) algorithm for video super-resolution

The Zooming-Slow Motion (ZSM) is a recent state of the art deep learning video super resolution method [30]. The method was developed by researchers from Purdue University, Northeastern University, and University of Rochester in 2020. The ZSM not only improves the resolution of the frames within a video but also improves the frame rate of the input videos (although the additional frames are currently not being utilized within our workflow).

ZSM can be broken down into three key components: feature temporal interpolation network, a deformable ConvLSTM, and a deep construction network [31]. The feature temporal interpolation network is used to interpolate missing temporal information between the input low-resolution frames. Next, the deformable ConvLSTM is used to align and aggregate the temporal information together. Lastly, the deep construction network predicts and generates the super resolution upsampled video frames.

The visual performance of the various super-resolution methods is shown in Fig. 2. It can be seen that ZSM method (two times and four times) yielded better results that DUF and bicubic methods. The edges around the bright spots are much sharper in the ZSM cases compared to the others. More detailed objective comparisons between the various VSR schemes are included in Sect. 3.3.

2.2 LIG principle and its improvements

Since the detection results of YOLO at the longer ranges (3500 m and above) were not as high as we would have liked [13], we investigated a traditional unsupervised small target detection method to see how it would perform on the
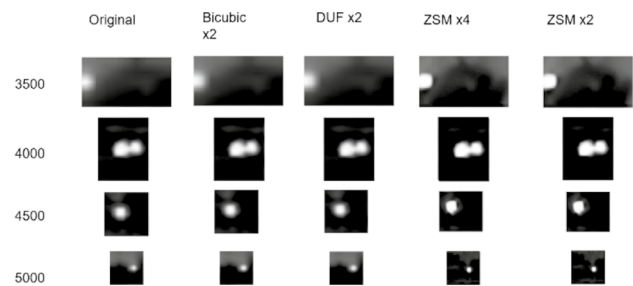
Fig. 2 Comparison of the image qualities of different video super-resolution algorithms. Videos from 3500 to 5000 m were used in our comparisons. The vehicle is BTR70

long-range videos. The algorithm of choice for this study was a local intensity gradient (LIG)-based target detector [5], specifically designed for infrared images. The LIG is relatively faster than other algorithms and is very robust to background clutter. As illustrated in Fig. 3, the algorithm scans through the input image using a sliding window, whose size depends on the input image resolution. For each window, the local intensity and gradient values are computed separately. Then, those values are multiplied to form an intensity-gradient (IG) map. An adaptive threshold is then used to segment the IG map and then the binarized image will reveal the target.

A major advantage of these unsupervised algorithms such as LIG is that they require no training, so there is no need to worry about customizing training data, which is the case with YOLO. A disadvantage of the LIG algorithm is that it is quite slow, taking roughly 30 s per frame of size 640×480 pixels.

There are several adjustments we made to the LIG algorithm to make it more suitable for the SENSIAC infrared dataset. First of all, we used differing patch sizes for the different resolution frames. For the original resolution frames, we used a patch size of 7×7 and for the $2 \times$ upsampled frames, we used a larger patch size of 19×19. This is to account for increased target size in high-resolution frames. Second, we adjusted the way in which the adaptable threshold T in LIG is calculated. In [5], the authors used the mean value of all nonzero pixels. For our dataset, this calculation produced a very small value due to the overwhelming amount of very low nonzero pixels. The left image in Fig. 4 highlights the significant role that the threshold plays for this algorithm. Third, we have implemented ways of speeding up the algorithm, such as incorporating multithreading within the script. We were able to speed up the computational time by close to three times.

Here, we would like to illustrate the importance of adaptive thresholding. For the example in Fig. 4, the mean value was 0.008. Using this threshold value for binarization as suggested by [5], we observe that roughly half the

Fig. 3 Illustration of the LIG principle

Input frame

Intensity

Gradient

I x G map

Segmentation results
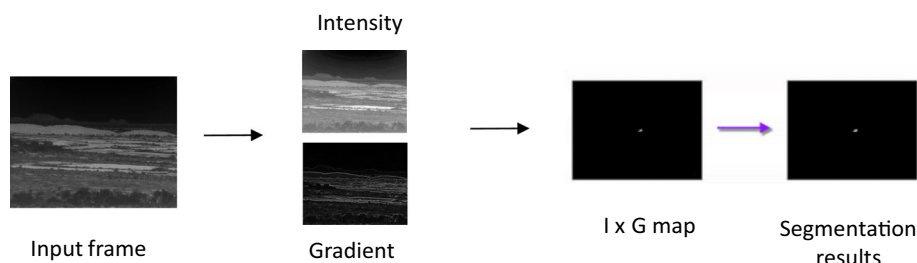
(a) Original adaptive threshold results [5]

(b) Our adaptive threshold results.

Fig. 4 Segmentation results using two different adaptive thresholds. False alarms have been reduced

nonzero pixels would be considered as detections, as seen on the left hand image of Fig. 4. This originally resulted in hundreds of false positives in the frames. So instead of using the mean of nonzero pixels in the LIG processed frame, we used the mean of the top 0.01% of pixels. A higher threshold is essential for eliminating false positives, as can be seen in the image on the right of Fig. 4.

2.3 CC analysis

We then perform connected component analysis on the segmented binarized image to find groupings of moving pixels between frames. Dilation of the binarized image can be used prior to the analysis in order to merge nearby pixels together. Moreover, the connected components are fed into a "Rule Analysis" block where there are several rule checks to determine whether the connected component is a valid detection or not. These rules are:

a. Check if the area of the connected component is reasonable. If the area is over 1 pixel and less than 100 pixels, then it is valid.
b. Compare the maximum intensity of pixels between the connected components. Out of the remaining connected components, the one with the pixel with the highest intensity is then chosen as the target.

2.4 Target association using SORT

Another tool we utilized is the SORT tracking association. SORT utilizes motion information as well as memory of past frames to associate targets from one frame to another. Each detected object, x, within a frame is modeled as follows [32]:

$$x = [u, \ v, s, r, \dot{u}, \dot{v}, \dot{s}]^{\mathrm{T}} \tag{1}$$

where u and v represent the horizontal and vertical pixel location of the object center, while s and r indicate the scale and aspect ratio of the object. These object states are then compared across frames to determine whether any of the states are correlated with any previous state.

3 Experimental results

3.1 Videos

We used four videos in the SENSIAC dataset [34]. We selected four daytime mid-wave infrared (MWIR) videos in the following ranges: 3500 m, 4000 m, 4500 m, and 5000 m. Although there are five different types of vehicles (BTR70, BMP2, BRDM2, ZSU23-4, and T72), the videos look similar because it is difficult to discern the various vehicles from such long ranges. Hence, we picked BTR70 videos in our experiments. In each video, there are 300 frames. The video frame rate is 7 frames/second, and the image size is 640×512. The images are in gray scale, and each pixel is represented by 8 bits. Figures 5, 6, 7 and 8 show several frames from each video. It can be seen that the vehicles (small bright spots) are hard to see and are quite small in size.
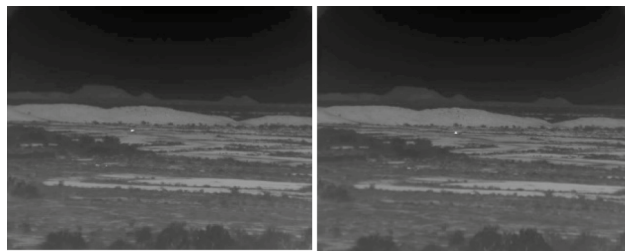
3.2 Performance metrics

A correct detection or true positive (TP) occurs if the binarized LIG result is within a certain threshold of the centroid of the ground truth bounding box. Otherwise, the detected object is regarded as a false positive (FP). Based on the
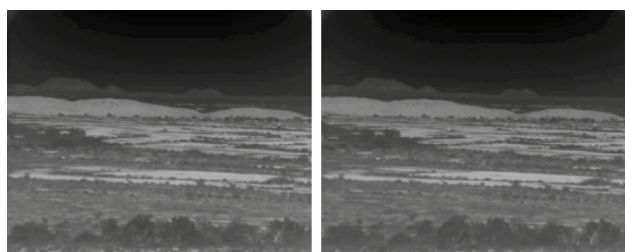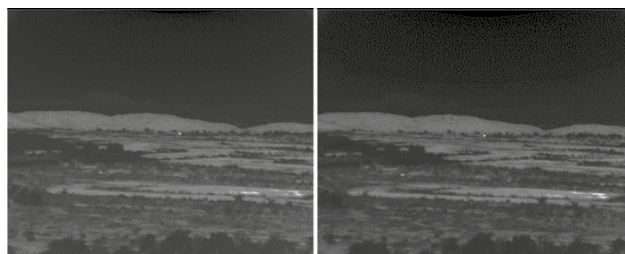
(a) Frame 1 **(b)** Frame 900
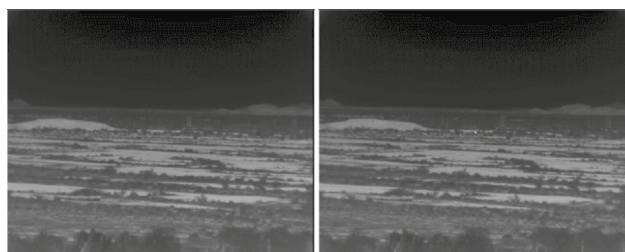
Fig. 5 Frames from the 3500 m video

(a) Frame 1 **(b)** Frame 900

Fig. 6 Frames from the 4000 m video

(a) Frame 1 **(b)** Frame 900

Fig. 7. 4500 m daytime videos

(a) Frame 1 **(b)** Frame 900

Fig. 8 Frames from the 5000 m video

correct detection and false positives counts, we can further generate precision, recall, and F1 metrics. The precision (P), recall (R), and F1 are defined as
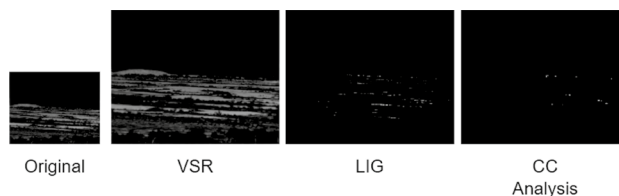

Fig. 9 Illustration of the workflow by combining ZSM and LIG. The video super-resolution (VSR) is done with ZSM

$$P = \frac{TP}{TP + FP} \tag{2}$$

$$R = \frac{TP}{TP + \text{Missed detections}} \tag{3}$$

$$F1 = \frac{2 \times P \times R}{P + R} \tag{4}$$

3.3 Performance of proposed approach without SORT

The proposed workflow without SORT is shown in Fig. 9. The VSR and LIG steps were mentioned earlier. The CC analysis involves several steps. First, it binarizes the LIG map with the adaptive threshold, meaning all values below the threshold are 0 and all those above are 1. Once this is done, dilation is performed on the resulting binarized image. The structuring element used is a square, and its size depends on the resolution of the images. At the original scale, we used a square with length 5 and at any other resolution we used a square with length 10. The dilation is crucial within the workflow because it merges groups of isolated pixels that correspond to the same target. In other words, without dilation, there would be significantly more false positives throughout the videos.

For our studies, we utilized the pretrained models provided by the authors of DUF and ZSM. The ZSM was trained on the large Vimeo-Septuplet dataset, which includes approximately 90,000 7-frame sequences from a variety of videos on Vimeo. Similarly to the DUF, the ZSM also has a temporal radius which uses frames before and after the current frame to better generate an upsampled frame. For our studies, we decreased their initial radius of 7 down to 3 due to hardware memory issues. Initially, we had concerns that the ZSM pretrained model may have similar difficulties in dealing with MWIR videos like the DUF, but the results proved otherwise. When testing, we focused on the longer range videos since YOLO struggled with detecting the targets in those videos. The results using various VSR algorithms are shown in Table 1. It should be noted that the ZSM×2 is a bicubically downsampled version of the

ZSM×4 (in our preliminary studies, this was done to get quick results since the LIG is time consuming for larger images).

"Original" means that we used the low-resolution videos as input to the workflow in Fig. 9. That is, no VSR module was used in the "Original" results. It can be seen in the 3500 m-5000 m video results in Table 2 that the ZSM outperforms or is comparable to the best performing methods in both precision and recall rates. In the 4000 m case, which is the most difficult case in terms of achieving high performance using super-resolution images, both ZSM×2

and ZSM×4 outperform all other cases by a slight margin. For the higher ranges like 4500 m–5000 m, the results were already quite good for the original, bicubic×2, and DUF×2, so there was not much room for improvement for the ZSM. But it still performs comparably to the top performing methods.

Although ZSM improves over the original resolution videos, the F1 scores are not improved by that much. So, in practical target detection applications, the value of VSR is somewhat limited. Moreover, the VSR increases the image size by two times in the ZSM×2 case, which consequently

Table 1 Performance of detection results without SORT using videos with different resolutions

(a) 3500 m vehicle 06				(b) 4000 m vehicle 06			
	P	R	F1		P	R	F1
Original	0.941	0.950	0.945	Original	0.943	0.943	0.943
Bicubic×2	0.947	0.953	0.950	Bicubic×2	0.940	0.940	0.940
DUF×2 [29]	0.944	0.950	0.947	DUF×2 [29]	0.940	0.937	0.938
ZSM×2 [30]	**0.957**	**0.957**	**0.957**	ZSM×2 [30]	0.960	0.963	0.962
ZSM×4 [30]	0.950	0.953	0.952	ZSM×4 [30]	**0.970**	**0.973**	**0.972**
(c) 4500 m vehicle 06				(d) 5000 m vehicle 06			
	P	R	F1		P	R	F1
Original	0.953	0.957	0.955	Original	0.928	0.943	0.936
Bicubic×2	0.957	0.960	0.958	Bicubic×2	0.928	0.950	0.939
DUF×2 [29]	**0.960**	0.957	0.958	DUF×2 [29]	0.928	0.947	0.937
ZSM×2 [30]	**0.960**	**0.963**	**0.962**	ZSM×2 [30]	**0.932**	**0.953**	**0.942**
ZSM×4 [30]	0.953	0.957	0.955	ZSM×4 [30]	0.922	0.950	0.936

Table 2 Performance of target detection with and without SORT

(a) 3500 m video			
	P	R	F1
Without SORT	0.941	0.950	0.945
With SORT	**1**	**0.95**	**0.974**
(b) 4000 m video			
	P	R	F1
Without SORT	0.943	0.943	0.943
With SORT	**0.990**	**0.953**	**0.971**
(c) 4500 m video			
	P	R	F1
Without SORT	0.953	**0.957**	0.955
With SORT	**0.990**	0.943	**0.966**
(d) 5000 m video			
	P	R	F1
Without SORT	0.928	0.943	0.936
With SORT	**0.997**	**0.957**	**0.977**

There are 300 frames in the videos.
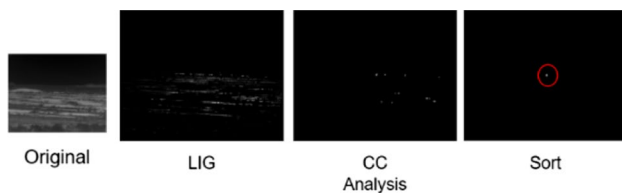
Fig. 10 Illustration of a lean workflow by combining LIG, CC, and Sort. The video super-resolution (VSR) is removed

increases the computational time of the LIG detection process.

3.4 Further detection improvement using SORT for track association

In this section, we would like to investigate the impact of SORT on the overall detection performance. From Sect. 3.3, we observed that although the VSR using ZSM improved the detection performance, the improvement is not very significant. As a result, we decided to focus on a lean version by removing the VSR in the above pipeline. Figure 10 illustrates the complete workflow of this lean version. This will speed up the processing with little loss of performance.

In our experiments, the SORT algorithm was implemented after the CC Analysis step. The bounding box information of each connected component is passed along to the SORT algorithm which then attempts to correlate the bounding boxes across frames. Since the detection rate is relatively high in the earlier stages of the workflow, we believed track association will help eliminate the majority of false positives. Our initial testing, presented in Table 2, has confirmed our hypothesis. On the 300 frames in the 3500 m video, the use of the SORT algorithm eliminated all the false positives. The results in other ranges also improved quite a lot. For instance, the F1 scores in the 5000 m improved from 0.936 without SORT to 0.977 with SORT. The SORT algorithm can help further distinguish between the background anomalies and the actual target. It should be noted that not all frames have a detection due to dim targets in some frames.

3.5 Computational complexity

The proposed framework is slow due to the use of LIG, which is the bottleneck. Even with parallel implementation for LIG using PARFOR in Matlab, it took approximately 75 s to process one frame, as can be seen in Table 3.

3.6 Subjective evaluation

Results for the workflow with and without SORT are included in Figs. 11, 12, 13 and 14 for comparison

Table 3 Computational times for the proposed target detection framework
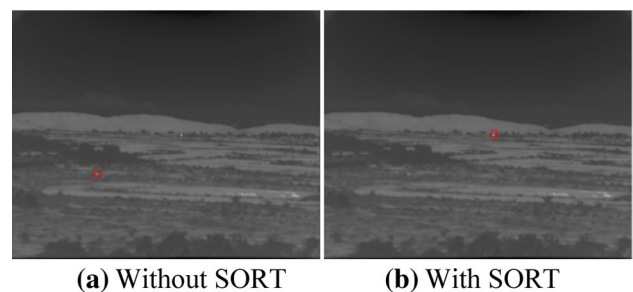
Method	Language	Time for 300 frames (s)
LIG	MATLAB	63,000
LIG Parallel	MATLAB	**22,000**

(a) Without SORT   **(b)** With SORT
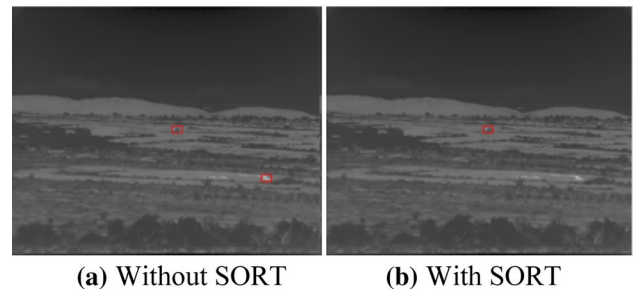
Fig. 11. 3500 m frame comparing with and without SORT

(a) Without SORT   **(b)** With SORT
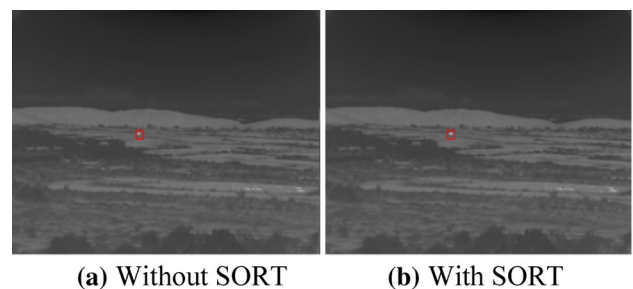
Fig. 12. 4000 m frame comparing with and without SORT

(a) Without SORT   **(b)** With SORT

Fig. 13. 4500 m frame comparing with and without SORT

purposes. In Fig. 11, we can see that there is a false positive in Fig. 11a where SORT was not used. In Fig. 13a, when there is no SORT, one can see that the target was missed and also a false positive was present. From those figures, one can easily observe that SORT helps eliminate the false positives and hence, can improve the overall target detection performance.
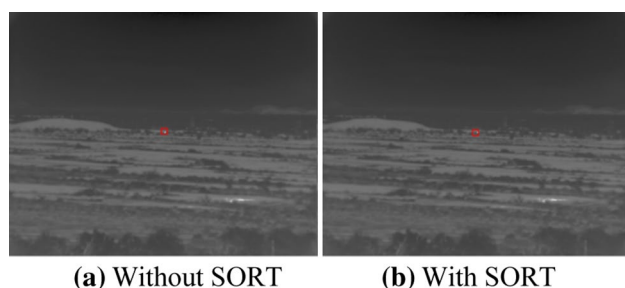
(a) Without SORT   **(b)** With SORT

Fig. 14. 5000 m frame comparing with and without SORT

4 Conclusions

In this research, we focus on small target detection in long-range infrared videos. A flexible and modular framework that contains a video super-resolution module, an unsupervised target detector, a connected component analysis module, and a track association module, was developed. The LIG method 5] was modified, sped up and applied to four MWIR videos that were collected from 3500 to 5000 m. Three VSR algorithms were compared, and one algorithm was found to improve the overall detection performance. However, the video super-resolution algorithm (ZSM) only improved the detection performance slightly. It was observed that the LIG algorithm worked quite well for target detection even at long ranges. Finally, the most dramatic improvement occurred when a simple target association algorithm known as SORT was incorporated into our framework.

One future direction is to further improve the computational complexity of the framework. The bottleneck is the LIG algorithm. Fortunately, the LIG uses a sliding window to scan through the whole image, implying that a parallel processing based on GPU may be a good research direction.

References

1. Chen, C.L.P., Li, H., Wei, Y., Xia, T., Tang, Y.Y.: A local contrast method for small infrared target detection. IEEE Trans. Geosci. Remote Sens. **52**(1), 574–581 (2014)
2. Wei, Y., You, X., Li, H.: Multiscale patch-based contrast measure for small infrared target detection. Pattern Recognit. **58**, 216–226 (2016)
3. Gao, C., Meng, D., Yang, Y., Wang, Y., Zhou, X., Hauptmann, A.G.: Infrared patch-image model for small target detection in a single image. IEEE Trans. Image Process. **22**(12), 4996–5009 (2013)
4. Zhang, L., Peng, L., Zhang, T., Cao, S., Peng, Z.: Infrared small target detection via non-convex rank approximation minimization joint l2,1 norm. Remote Sens. **10**(11), 1–26 (2018)
5. Zhang, H., Zhang, L., Yuan, D., Chen, H.: Infrared small target detection based on local intensity and gradient properties. Infrared Phys. Technol. **89**, 88–96 (2018)
6. Chen, Y., Zhang, G., Ma, Y., Kang, J. U., Kwan, C.: Small infrared target detection based on fast adaptive masking and scaling with iterative segmentation. IEEE Geosci. Remote Sens. Lett. (2020).
7. Kwan, C., Chou, B., Kwan, L. M.: A comparative study of conventional and deep learning target tracking algorithms for low quality videos. In: Proceedings of the 15th International Symposium on Neural Networks (2018)
8. Kwan, C., Budavari, B.: Enhancing small target detection performance in low quality and long range infrared videos using optical flow techniques. Remote Sens. **12**(24), 4024 (2020)
9. Demir, H. S., Cetin, A. E.: Co-difference based object tracking algorithm for infrared videos. In: IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, pp. 434–438. (2016)
10. Kwan, C., Chou, B., Yang, J., Tran, T.: Compressive object tracking and classification using deep learning for infrared videos. In: Proceedings of the SPIE 10995, Pattern Recognition and Tracking (Conference SI120) (2019)
11. Kwan, C., Chou, B., Yang, J., Tran, T.: Target tracking and classification directly in compressive measurement for low quality videos. In: SPIE 10995, Pattern Recognition and Tracking XXX, 1099505 (2019)
12. Kwan, C., Chou, B., Echavarren, A., Budavari, B., Li, J., Tran, T.: Compressive vehicle tracking using deep learning. IEEE Ubiquitous Computing, Electronics and Mobile Communication Conference, New York City. (2018)
13. Kwan, C., Gribben, D., Tran, T.: Multiple human objects tracking and classification directly in compressive measurement domain for long range infrared videos. In: IEEE Ubiquitous Computing, Electronics & Mobile Communication Conference, New York City. (2019)
14. Kwan, C., Gribben, D., Tran, T.: Tracking and classification of multiple human objects directly in compressive measurement domain for low quality optical videos. In: IEEE Ubiquitous Computing, Electronics & Mobile Communication Conference. New York City (2019)
15. Kwan, C., Chou, B., Yang, J., Tran, T.: Deep learning based target tracking and classification directly in compressive measurement for low quality videos. Signal Image Process. Int. J. (SIPIJ). (2019)
16. Kwan, C., Chou, B., Yang, J., Rangamani, A., Tran, T., Zhang, J., Etienne-Cummings, R.: Target tracking and classification directly using compressive sensing camera for SWIR videos. J. Signal Image Video Process. **13**, 1629–1637 (2019)
17. Kwan, C., Chou, B., Yang, J., Rangamani, A., Tran, T., Zhang, J., Etienne-Cummings, R.: Target tracking and classification using compressive measurements of MWIR and LWIR coded aperture cameras. J. Signal Information Processing. **10**, 73–95 (2019)
18. Kwan, C., Gribben, D., Rangamani, A., Tran, T., Zhang, J., Etienne-Cummings, R.: Detection and confirmation of multiple human targets using pixel-wise code aperture measurements. J. Imag. **6**(6), 40 (2020)
19. Kwan, C., Chou, B., Yang, J., Tran, T.: Deep learning based target tracking and classification for infrared videos using compressive measurements. J. Signal Inform. Process. (2019)
20. Kwan, C., Chou, B., Yang, J., Rangamani, A., Tran, T., Zhang, J., Etienne-Cummings, R.: Deep learning based target tracking and classification for low quality videos using coded aperture camera. Sensors **19**(17), 3702 (2019)

21. Lohit, S., Kulkarni, K., Turaga, P. K.: Direct inference on compressive measurements using convolutional neural networks. Int. Conf. Image Process. 1913–1917. (2016)
22. Adler, A., Elad, M., Zibulevsky, M.: Compressed learning: a deep neural network approach. (2016)
23. Xu, Y., Kelly, K. F.: Compressed domain image classification using a multi-rate neural network. (2019)
24. Wang, Z. W., Vineet, V., Pittaluga, F., Sinha, S. N., Cossairt, O., Kang, S. B.: Privacy-preserving action recognition using coded aperture videos. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. (2019)
25. Vargas, H., Fonseca, Y., Arguello, H.: Object detection on compressive measurements using correlation filters and sparse representation. In: Proceedings of the 26th European Signal Processing Conference (EUSIPCO). Pp. 1960–1964. (2018)
26. Değerli, A., Aslan, S., Yamac, M., Sankur, B., Gabbouj, M.: Compressively sensed image recognition. In: Proceedings of the 7th European Workshop on Visual Information Processing (EUVIP), Tampere. (2018)
27. Latorre-Carmona, P., Traver, V. J., Sánchez, J. S., Tajahuerce, E.: Online reconstruction-free single-pixel image classification. Image Vis. Comput. (2018)
28. Kwan, C., Gribben, D., Chou, B., Budavari, B., Larkin, J., Rangamani, A., Tran, T., Zhang, J., Etienne-Cummings, R.: Real-time and deep learning based vehicle detection and classification using pixel-wise code exposure measurements. Electronics, (2020)
29. Jo, Y., Oh, S. W., Kang, J., Kim, S. J.: Deep video super-resolution network using dynamic upsampling filters without explicit motion compensation. In: Proceedings of the IEEE Conference on Computer Vision and #pattern recognition (pp. 3224–3232). (2018)
30. Xiang, X.: Mukosame/Zooming-Slow-Mo-CVPR-2020. GitHub, 2020, github.com/Mukosame/Zooming-Slow-Mo-CVPR-2020
31. Xiang, X., Tian, Y., Zhang, Y., Fu, Y., Allebach, J. P., Xu, C.: zooming slow-mo: fast and accurate one-stage space-time video super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 3370–3379). (2020)
32. Kwan, C., Budavari, B.: Enhancing small moving target detection performance in low-quality and long-range infrared videos using optical flow techniques. Remote Sens. **12**(24), 4024 (2020)
33. Bewley, A., Ge, Z., Ott, L., Ramos, F., Upcroft, B.: Simple online and realtime tracking.
34. SENSIAC dataset, https://www.news.gatech.edu/2006/03/06/sensiac-center-helps-advance-military-sensing. Accessed 9 Nov 2020.