Performance Analysis of Siamese tracker through Preprocessing of Infrared Images

Tae Yoon Kim
Interdisciplinary Program in Intelligent Aerospace System
Seoul National University
Seoul, 08826, Republic of Korea
tykim512@snu.ac.kr

Chan Gook Park
Department of Aerospace Engineering
Seoul National University
Seoul, 08826, Republic of Korea
chanpark@snu.ac.kr

***Abstract*—Target tracking in infrared images is actively researched in computer vision. Unlike RGB images, infrared images rely on a single channel of thermal signals, making them vulnerable to issues such as non-uniform illumination and noise. In this study, we propose integrating CLAHE (Contrast-Limited Adaptive Histogram Equalization), a preprocessing technique, into the SiamRPN framework, a widely used Siamese network-based tracker. CLAHE enhances local contrast while maintaining computational efficiency, enabling robust tracking, even in resource-constrained environments. The performance of the proposed method was evaluated using challenging subsets of the LSOTB-TIR dataset, selected to represent conditions where standard models typically struggle. Experimental results demonstrate that the CLAHE-enhanced SiamRPN achieves significant improvements in accuracy and robustness compared to the baseline model, particularly in environments with low contrast and complex backgrounds. These findings suggest that CLAHE can be a practical solution for improving performance in environments with limited computational resources.**

Keywords—object tracking, infrared image, siamese network, preprocessing

I. Introduction

In the field of computer vision, extensive research has been conducted on leveraging deep learning for vision-based target tracking. Among these, methods based on Siamese networks have garnered significant attention for achieving remarkable performance and exceptionally fast inference speeds.

While most tracking algorithms are designed for RGB data, infrared (IR) imagery offers unique advantages by detecting thermal signals beyond the visible spectrum. This makes IR tracking particularly effective in low-light conditions, such as nighttime or in scenarios involving camouflage. In defense and aviation sectors, IR data can be combined with other modalities, such as radar, to enhance multispectral tracking accuracy. However, due to its single-channel composition, IR imagery is inherently limited in information compared to RGB, making it highly sensitive to shape-related features.

While much of the focus has been on feature extraction within neural networks, preprocessing techniques like CLAHE provide a complementary approach to enhance input quality, particularly for IR imagery. Existing methods often emphasize local information, which can lead to performance degradation when raw image quality is poor. Issues like saturation or complex backgrounds exacerbate these limitations, leading to target loss or reduced accuracy.

To address these challenges, this paper investigates the impact of applying CLAHE, a preprocessing technique, to enhance the input images used in existing Siamese trackers. CLAHE effectively improves tracking accuracy without introducing additional computational overhead, making it suitable for resource-constrained environments. The Siamese network chosen for this study is SiamRPN, which was fine-tuned using the LSOTB-TIR dataset based on a model pretrained on ImageNet. Performance evaluation was conducted on challenging subsets of the LSOTB-TIR dataset. These subsets were specifically chosen to reflect conditions where traditional trackers often fail, such as low contrast and complex backgrounds, to rigorously validate the proposed approach.

The results are analyzed to assess the performance improvements brought by the proposed preprocessing approach.

II. Proposed Method

This study aims to improve the performance of infrared-based target tracking by applying the CLAHE preprocessing technique to the existing SiamRPN tracker and comparing the performance before and after its application. CLAHE enhances local contrast, making the features of the target and background more distinct. This improves tracking robustness even in complex environments.

The proposed system consists of two stages. In the first stage, the CLAHE algorithm is applied to adjust the brightness distribution of the input image and emphasize details in low-contrast regions. This process mitigates common issues in infrared imagery, such as saturation or low-light conditions, while making the target's shape more prominent.

In the second stage, the preprocessed image is fed into the SiamRPN network to track the target. The enhanced features provided by CLAHE enable the network to more effectively learn and utilize detailed information about the target.

A. Preprocessing with CLAHE

CLAHE is a preprocessing technique based on adaptive histogram equalization that enhances the local contrast of an image. Unlike traditional histogram equalization, which applies adjustments uniformly across the entire image, CLAHE divides the image into small tiles and independently calculates and equalizes the histogram for each tile. This approach emphasizes details in low-contrast regions and enhances the distinction between the target and the background, even in complex environments.

In addition, CLAHE uses histogram clipping to prevent excessive enhancement of specific intensity levels, thereby minimizing visual distortions. The clipped histograms are then applied within each tile, and the boundaries between

979-8-3315-2040-3/25/$31.00 ©2024 IEEE

tiles are seamlessly connected using linear interpolation. This interpolation step plays a crucial role in eliminating artifacts caused by tile-based processing, ensuring that the resulting image has smooth brightness transitions. This is particularly important for infrared images, which often exhibit complex backgrounds or uneven illumination, as it ensures a more natural appearance while preserving key features.

The images below illustrate the changes before and after applying CLAHE. The left image shows the original input, while the right image demonstrates the result after CLAHE processing. CLAHE effectively enhances details in low-contrast regions, balances the brightness distribution, and makes the target's shape more distinguishable.

Fig. 1 Infrared Image Enhancement with CLAHE

B. SiamRPN

SiamRPN is a deep learning-based target tracking algorithm that utilizes a Siamese network to extract features of a target and predict its position. The algorithm takes a template image of the target and a search image as inputs, processes them through a shared feature extractor (backbone), and generates feature maps. It then calculates the cross-correlation between the two feature maps to measure the similarity with the target, which serves as the basis for position estimation.

One of the key characteristics of SiamRPN is its integration with a Region Proposal Network (RPN). While traditional Siamese networks were limited to predicting the position of the target, SiamRPN leverages RPN to simultaneously extract information about the target's location, size, and shape. This allows for robust tracking of targets with varying sizes and shapes, providing enhanced performance. The RPN efficiently predicts bounding boxes for the target, minimizing errors during the tracking process.

SiamRPN employs a backbone network to extract high-level features from the input images. These features are used in the cross-correlation process to compute the similarity between the template and the search region. The RPN then predicts the position and size of the target, ultimately determining the bounding box. This architecture achieves both high inference speed and accuracy, making it highly efficient compared to traditional tracking algorithms.

The figure below illustrates the structure of the proposed integrated method.

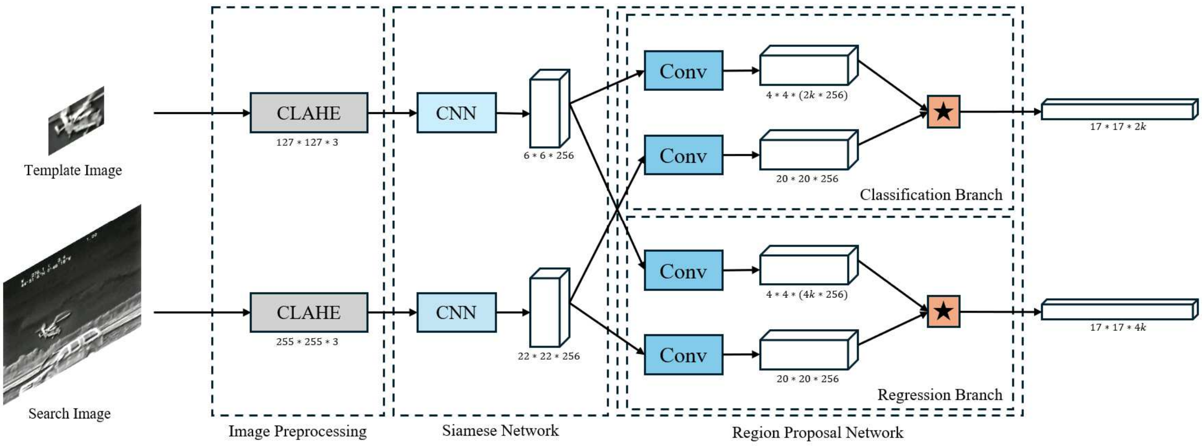

Fig. 2 Architecture of CLAHE-enhanced SiamRPN

III. Experiment

A. Experimental Setup

This study was conducted in an Ubuntu 22.04 environment using an RTX-4090 GPU, with implementation based on the PyTorch framework. The SiamRPN network used a pretrained backbone on ImageNet and was fine-tuned with the LSOTB-TIR dataset. The LSOTB-TIR dataset contains a variety of infrared scenarios, making it well-suited for evaluating tracking performance under challenging conditions, such as complex backgrounds and low-contrast environments.

The performance evaluation was carried out using the One-Pass Evaluation (OPE) protocol. Two key metrics were employed: Success Rate and Precision. The Success Rate measures the proportion of frames where the overlap ratio (Intersection over Union, IoU) between the predicted and ground truth bounding boxes exceeds a certain threshold, evaluating the overall robustness of the tracker. Precision measures the proportion of frames where the Euclidean distance (offset) between the predicted and ground truth target center falls below a certain threshold, focusing on localization accuracy. These metrics are visualized through cumulative curves, with the Success and Precision values derived from the Area Under Curve (AUC) or specific threshold values.

B. Performance Analysis

Table 1 summarizes the Success Rate and Precision results for SiamRPN with and without CLAHE, evaluated under different tile sizes (8, 12, 24). CLAHE-enhanced models consistently outperformed the baseline SiamRPN across all configurations. Notably, SiamRPN+CLAHE12 achieved the best performance, with a Success Rate of 0.661 and Precision of 0.808.

These results indicate that the tile size of 12 provides the optimal balance between enhancing local details and maintaining global context. Smaller tile sizes, such as 8, excessively emphasized local information, making the tracker more vulnerable to noise and artifacts like motion blur. Conversely, larger tile sizes, such as 24, diluted the representation of small targets, reducing the extraction of meaningful features and leading to performance degradation. CLAHE's ability to adaptively enhance contrast in low-visibility regions played a critical role in improving the distinction between the target and the background.

TABLE I. PERFORMANCE OF SIAMRPN WITH CLAHE

Tracker name	*Success*	*Precision*
SiamRPN+CLAHE12	0.661	0.808
SiamRPN+CLAHE24	0.630	0.765
SiamRPN+CLAHE8	0.626	0.768
SiamRPN	0.585	0.721

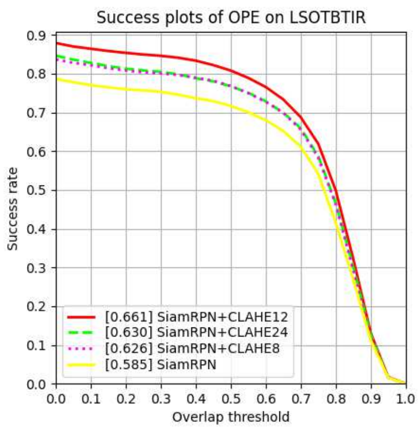

Fig. 3 Success Plot for OPE on LSOTB-TIR

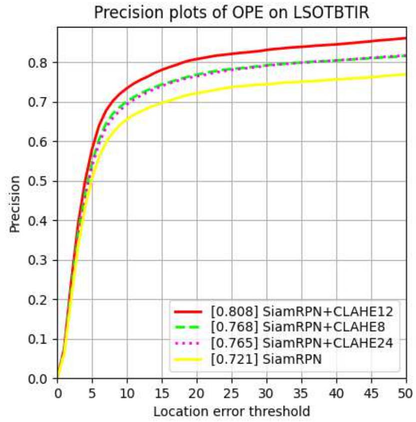

Fig. 4 Precision Plot for OPE on LSOTB-TIR

Above Figures show the Success and Precision plots, respectively, for SiamRPN with and without CLAHE. The Success Plot (Figure 3) demonstrates that models with CLAHE maintain higher Success Rates across the entire range of overlap thresholds. The performance gap becomes particularly prominent at overlap thresholds above 0.5, indicating that CLAHE enhances the accuracy and stability of bounding box predictions.

Similarly, the Precision Plot (Figure 4) shows that CLAHE-enhanced models achieve higher Precision values, especially at smaller offset distances. This suggests that CLAHE improves the model's ability to extract fine-grained features, resulting in better initial localization. Among the configurations, CLAHE with a tile size of 12 exhibited the most consistent performance in both Success and Precision metrics, while tile sized of 8 and 24 showed suboptimal results due to either excessive or insufficient contrast enhancement.

C. Qualitative Analysis

To visually demonstrate the impact of CLAHE on tracking performance, we compared scenarios where the tracker loses the target and where it successfully maintains the target (Figure 5). The left column shows results from the baseline SiamRPN where the target is lost under challenging conditions such as low contrast or complex backgrounds. In contrast, the right column shows results from the CLAHE-enhanced tracker, which successfully tracks the target in the same scenarios.

The improved performance of the CLAHE-enhanced tracker can be attributed to its ability to enhance local details in low-visibility regions. These adjustments allow the tracker to better distinguish the target from the background, even in challenging environments.

Fig. 5 Tracking Results Comparison with CLAHE

D. Discussion

The CLAHE-enhanced SiamRPN outperformed the baseline model in both Success Rate and Precision, demonstrating its effectiveness in improving local contrast and tracking robustness under challenging conditions such as motion blur and complex backgrounds. Success and Precision plots further confirmed that contrast enhancement positively impacted both bounding box prediction accuracy and localization precision.

While adjustments to the clipping limit had minimal effect, the choice of tile size significantly influenced performance. Smaller tile sizes increased noise, and larger tile sizes reduced critical target details. To achieve optimal results, the tile size should be adaptively adjusted based on the size of the target, highlighting the need for further research into adaptive parameter tuning strategies.

On the other hand, integrating denoising techniques led to a decrease in performance by reducing fine details crucial for tracking. This suggests that alternative methods specifically aimed at enhancing details, rather than suppressing noise, are needed to further improve performance in infrared target tracking.

In addition, combining CLAHE with other preprocessing techniques or applying it to multispectral data could unlock new possibilities for tracking systems, enhancing robustness and accuracy across diverse scenarios.

IV. Conclusion

This study demonstrated that integrating CLAHE as a preprocessing technique significantly improves the performance of the SiamRPN tracker in infrared target tracking. By enhancing input image quality, the CLAHE-enhanced SiamRPN achieved higher accuracy and robustness compared to the baseline, all while maintaining computational efficiency. This makes it particularly suitable for deployment in resource-constrained environments.

These findings highlight the potential of preprocessing techniques to enhance tracking performance without architectural changes. Future work could focus on adaptive parameter tuning and exploring synergies with other methods to further expand the capabilities of infrared target tracking systems.

Acknowledgment

This research was supported by the Agency for Defense Development, Defense Acquisition Program Administration as part of the Artificial Intelligence Research Laboratory for Flight Control Program (UD230014SD).

References

[1] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. S. Torr, "Fully-convolutional Siamese networks for object tracking," European Conference on Computer Vision (ECCV) Workshops, 2016, pp. 850–865.

[2] X. Zhu, H. Wang, B. Wu, and J. Yan, "LSOTB-TIR: A large-scale high-diversity thermal infrared single object tracking benchmark," IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2019, pp. 2003–2012.

[3] B. Li, B. Wu, Q. Wang, F. Zhang, J. Xing, and J. Yan, "SiamAtt: Siamese attention network for visual tracking," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 4854–4863.

[4] Z. Zhu, Q. Wang, L. Bo, B. Wu, J. Yan, and W. Hu, "UpdateNet: Learning the model update for Siamese trackers," IEEE International Conference on Computer Vision (ICCV), 2018, pp. 2512–2521.

[5] B. Li, J. Yan, W. Wu, Z. Zhu, and X. Hu, "High performance visual tracking with Siamese region proposal network," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 8971–8979.

[6] Y. Mei, N. Yan, H. Qin, T. Yang, and Y. Chen, "SiamFCA: A new fish single object tracking method based on Siamese network with coordinate attention in aquaculture," Computers and Electronics in Agriculture, vol. 216, p. 108542, Dec. 2023.

[7] J. Kim and B. Choi, "Research on channel-wise preprocessing for enhanced infrared object detection," Journal of The Korea Society of Computer and Information, vol. 29, no. 11, pp. 153–161, Nov. 2024.

[8] Byung-Jin Kang, Jaehyun Bae, Daehyeon Kim, and Kyounghoon Baek, "Performance evaluation of deep learning based object detection model on image infrared preprocessing," Journal of The Korean Society for Aeronautical and Space Sciences, vol. 52, no. 12, pp. 1055–1061, Nov. 2024.

[9] J.-Y. Baek et al., "Research on local and global infrared image pre-processing methods for deep learning-based guided weapon target detection," Journal of The Korea Society of Computer and Information, vol. 29, no. 7, pp. 41–51, July 2024.

[10] S. M. Pizer, R. E. Johnston, J. P. Ericksen, B. C. Yankaskas, and K. E. Muller, "Contrast-limited adaptive histogram equalization: Speed and effectiveness," Proceedings of the First Conference on Visualization in Biomedical Computing, 1990, pp. 337–345.