PREFIX-NORMALIZED BAND-GROUP RENDERING WITH A SENSOR-ROUTED HETEROGENEOUS TRACKER ENSEMBLE FOR HYPERSPECTRAL OBJECT TRACKING

Pengju Zhao

Independent Researcher, China

ABSTRACT

Reusing pretrained visual trackers for hyperspectral video requires mapping multiband measurements to three-channel inputs. We combine prefix-normalized band-group rendering with sensor-specific prediction selection, using adapted LoRAT-Giant and LoRAT-Large [1] alongside pretrained SAMURAI [2].

Raw mosaics are decoded into multichannel arrays, then averaged within three approximately equal index groups. Normalization bounds are estimated once from each sequence's first five image frames, without target annotations, and fixed thereafter. This startup look-ahead is confined to the initial five-frame block, not a rolling future-frame window; later images do not update the bounds. Preprocessing is therefore not strictly online from frame one. The representation compresses real spectral measurements rather than reconstructing physical RGB.

The three trackers independently predict boxes from the same first-frame initialization. For NIR, selection has two stages. First, a ridge regressor uses 47 appearance and trajectory features to choose Giant's or SAMURAI's box. Second, this box joins Large's and SAMURAI's boxes; the candidate with the greatest summed intersection-over-union (IoU) is retained, without coordinate averaging. This gated medoid returns SAMURAI whenever the first stage selects it, because its box occurs twice. RedNIR and VIS directly use Large and SAMURAI, respectively. Weights remain fixed during inference, with no selected-box feedback. These sensor-specific rules were chosen from labeled development comparisons (Table 1), not inferred from sensor physics.

Table 1. Development AUC (%) on public HOTC samples using identical group rendering. Bold: selected rule per sensor, not significance or independent validation.

Candidate rule	NIR (n=22)	RedNIR (n=13)	VIS (n=28)
LoRAT-Giant	72.842	43.791	61.574
LoRAT-Large	69.834	**48.402**	62.501
SAMURAI	74.191	41.656	**72.383**
Gated medoid	**74.650**	45.112	69.800

On public samples, replacing supplied false color yields matched-tracker AUC gains of 1.79, 5.68, and 3.46 percentage points for NIR/SAMURAI, RedNIR/Large, and VIS/SAMURAI. The NIR gain remains uncertain: its paired 95% bootstrap interval spans -0.20 to +4.27 points. A separate sequence-held-out selector analysis on training-workflow sequences gains 3.73 points over SAMURAI, conditional on development-selected models. As a limitation, these development comparisons neither independently validate the complete policy nor isolate individual preprocessing contributions.

The proposed method achieved third place in HOTC 2026's final second-stage evaluation, with organizer-reported AUC of 66.9679% and DP@20 of 84.7255% after code review and unified inference. This supports the method's overall performance; a limitation is that component contributions are not isolated.

***Index Terms*—** Hyperspectral object tracking, band-group rendering, tracker ensemble, sensor-specific selection

1. REFERENCES

[1] Liting Lin, Heng Fan, Zhipeng Zhang, Yaowei Wang, Yong Xu, and Haibin Ling, "Tracking meets LoRA: Faster training, larger model, stronger performance," in *Computer Vision – ECCV 2024*, 2025, pp. 300–318, doi: 10.1007/978-3-031-73232-4_17.

[2] Cheng-Yen Yang, Hsiang-Wei Huang, Wenhao Chai, Zhongyu Jiang, and Jenq-Neng Hwang, "SAMURAI: Adapting segment anything model for zero-shot visual tracking with motion-aware memory," arXiv:2411.11922, 2024.