RGB-D Tracking via Hierarchical Modality Aggregation and Distribution Network

Boyue Xu, Yi Xu, Ruichao Hou, Jia Bei ✉, Tongwei Ren, Gangshan Wu

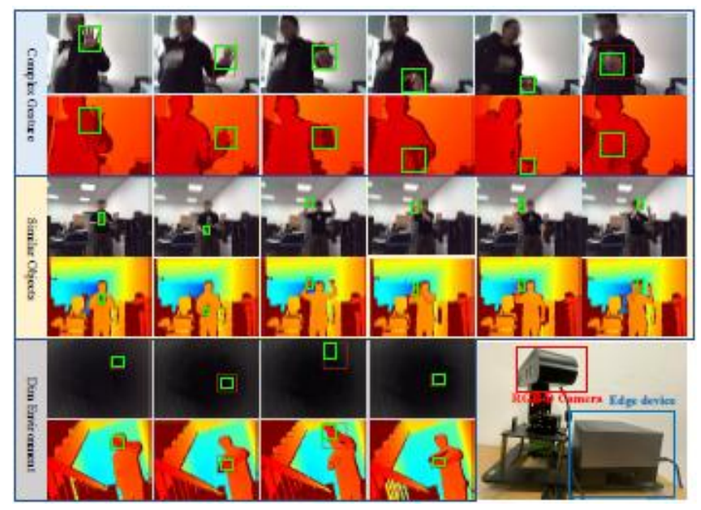
State Key Laboratory for Novel Software Technology, Nanjing University

MMAsia 2023

Introduction

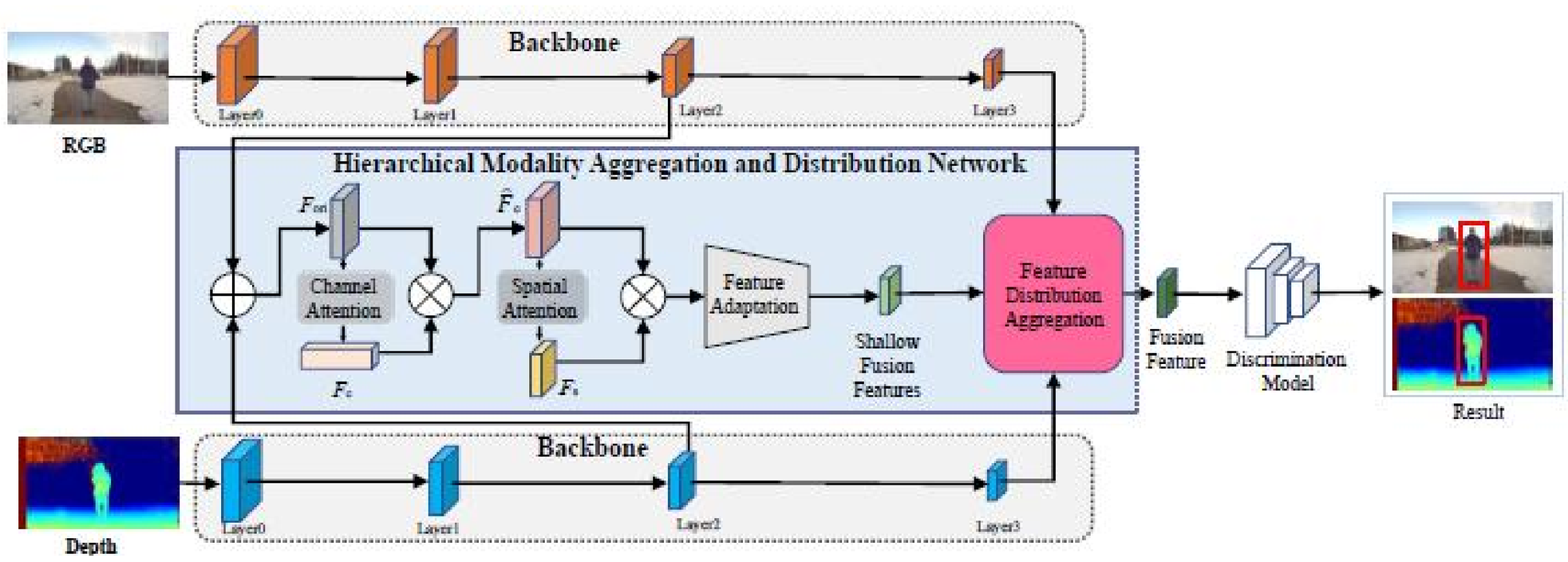
RGB-Depth(RGB-D) tracking combines RGB and depth data. RGB offers visual details like color and texture, while depth providing the distance from the camera to the object.

We proposed a RGB-D tracker named HMAD, which leverages the distinct feature representation strengths of RGB and depth modalities, giving prominence to a hierarchical approach for feature distribution and fusion, thereby enhancing the robustness of RGB-D tracking.



Method

The input of the HMAD is the corresponding RGB and depth images, which are merged through a hierarchical modality aggregation and distribution network. This process consists of two parts: the first part involves attention-based shallow feature extraction, while the second part involves feature distribution and fusion. The former is responsible for extracting the effective components in shallow features, while the latter is responsible for distributing and fusing shallow features with deep features. The fused features are then used for final target discrimination and tracking.



Experiments

Dataset: DepthTrack, RGBD1K

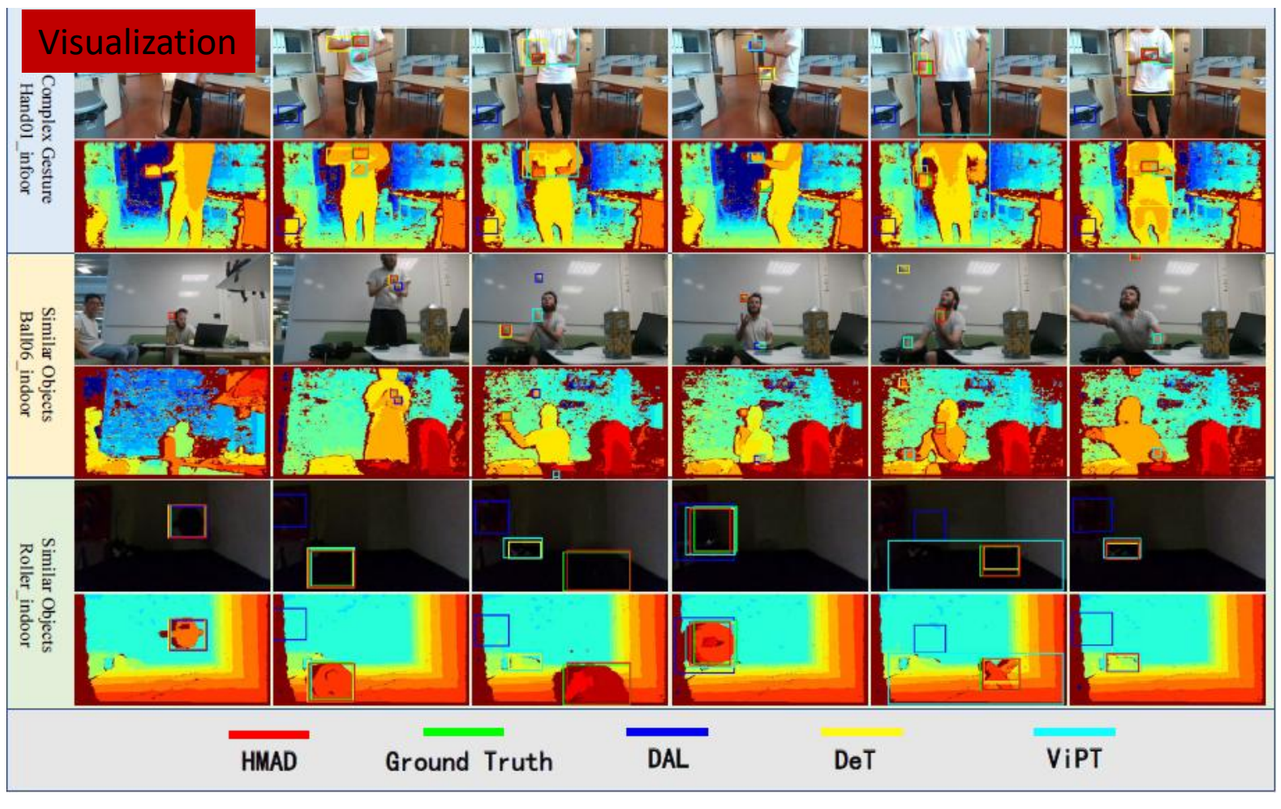
- Challenges: Small target, Similar targets Fast Movement, Background clutter

Metrics: Precision(Pr), Recall(Re), F-score

Comparison with the SOTA: The method is superior to all the tracking methods

Ablation Study: The experiment demonstrates the effectiveness of each component of the method

Visualization



Ablation

	Pr	Re	F-score
baseline	0.548	0.525	0.536
w/o distribution	0.571	0.545	0.557
w/o attention	0.596	0.562	0.579
HMAD	0.626	0.597	0.611

Comparison

	DepthTrack			RGBD1K			FPS
	Pr	Re	F-score	Pr	Re	F-score	
Siam_LTD [27]	0.342	0.418	0.376	0.543	0.318	0.398	13.0
DAL [30]	0.512	0.369	0.429	0.562	0.407	0.472	-
CLGS_D [27]	0.369	0.584	0.453	-	-	-	7.3
ATCAIS [27]	0.455	0.500	0.476	0.511	0.451	0.479	1.3
DDIMP [27]	0.469	0.503	0.485	0.557	0.534	0.545	4.7
Drefine [20]	-	-	-	0.532	0.462	0.494	-
OSTrack [38]	0.522	0.536	0.529	-	-	-	-
DeT [36]	0.506	0.560	0.532	0.438	0.419	0.428	36.8
SPT [40]	0.549	0.527	0.538	0.545	0.578	0.561	25.3
ProTrack [37]	0.583	0.573	0.578	-	-	-	-
ViPT [39]	0.596	0.592	0.594	-	-	-	14.1
HMAD	0.626	0.597	0.611	0.573	0.552	0.562	50.0

南京大学 NANJING UNIVERSITY

xuby@smail.nju.edu.cn

beijia@nju.edu.cn

MAGUS MediA recoGnition and UnderStanding