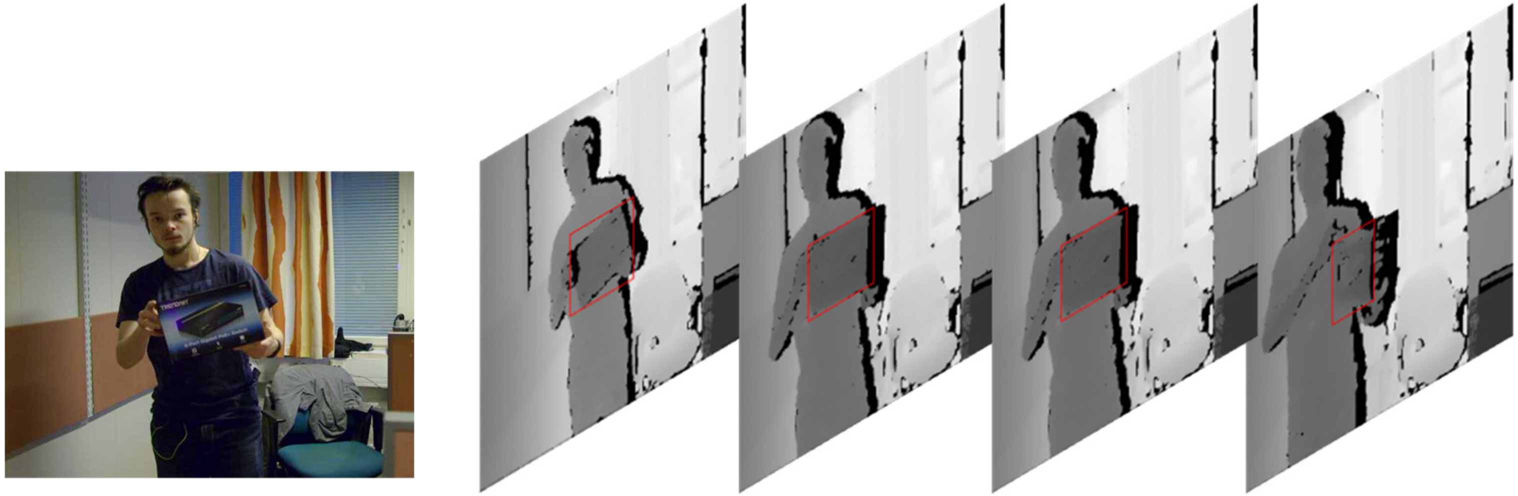
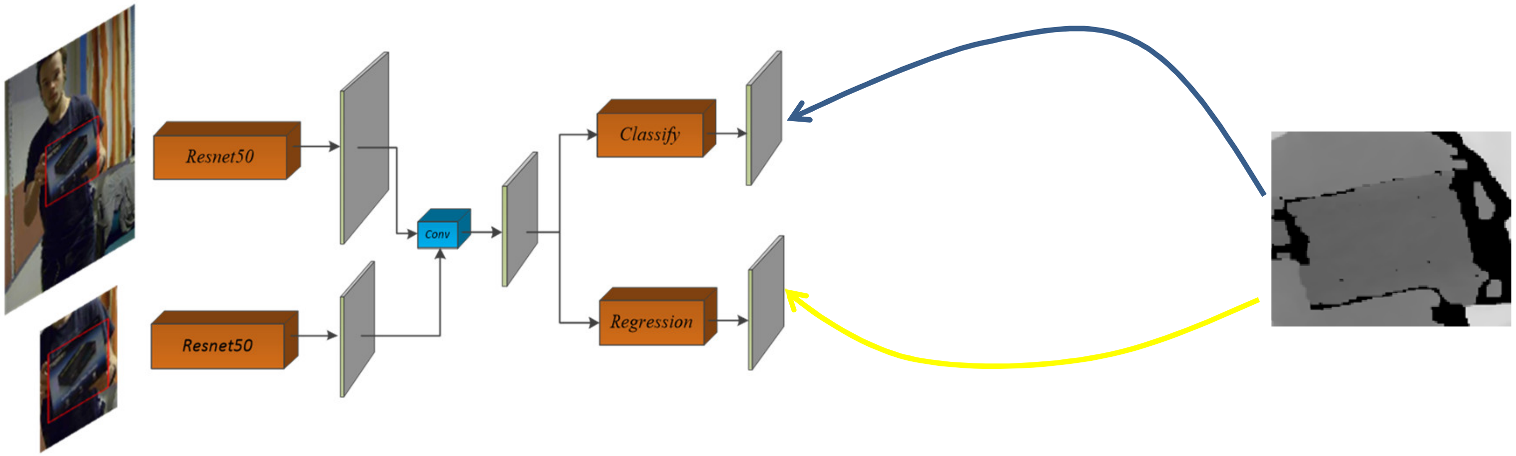
TSDM: Tracking by SiamRPN++ with a Depth-refiner and a Mask-generator



Depth (D) image provides informative cues for object tracking. However, few trackers have used it due to the lack of a suitable model.

TSDM: Tracking by SiamRPN++ with a Depth-refiner and a Mask-generator

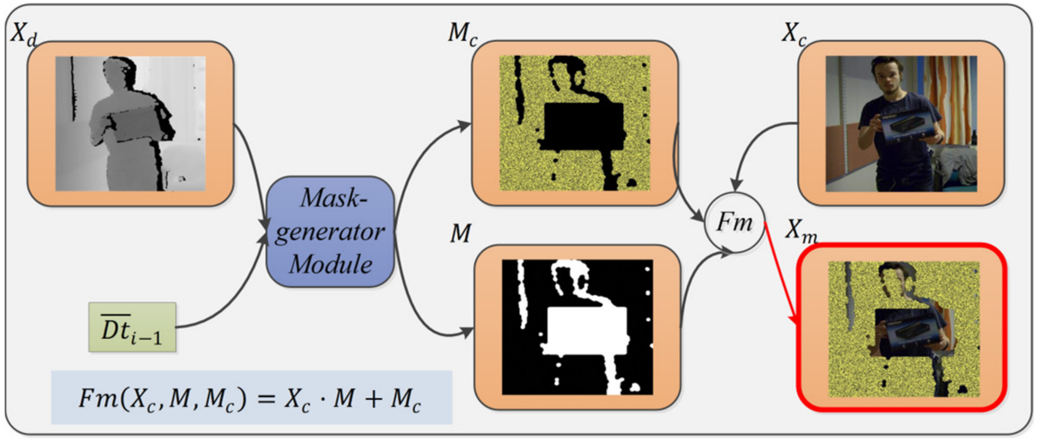


Anchor classification. ← Fore-back separation.

Anchor regression. ← Clear object edges .

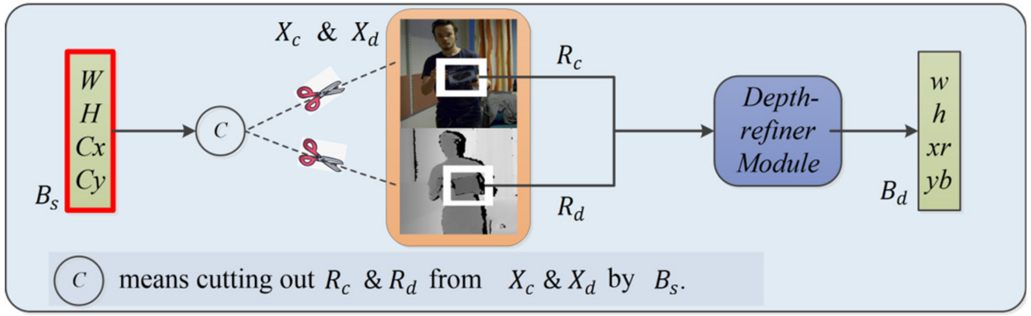
Improvement

TSDM: Tracking by SiamRPN++ with a Depth-refiner and a Mask-generator

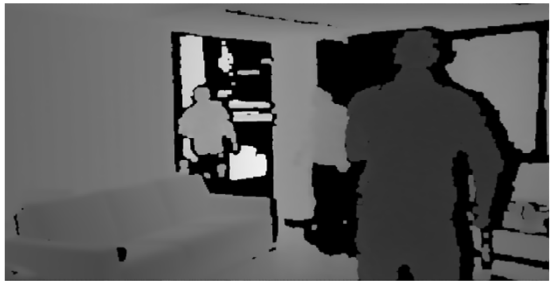


Input X_d and $\overline{Dt}_{i-1}$ into M-g to get M and M_c. Then use $F_m(\cdot)$ to get X_m.

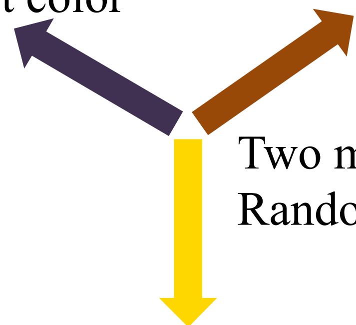
Cut out R_c and R_d from X_c and X_d by B_s respectively. Then input R_c and R_d into D-r to get the refined target bounding box B_d.



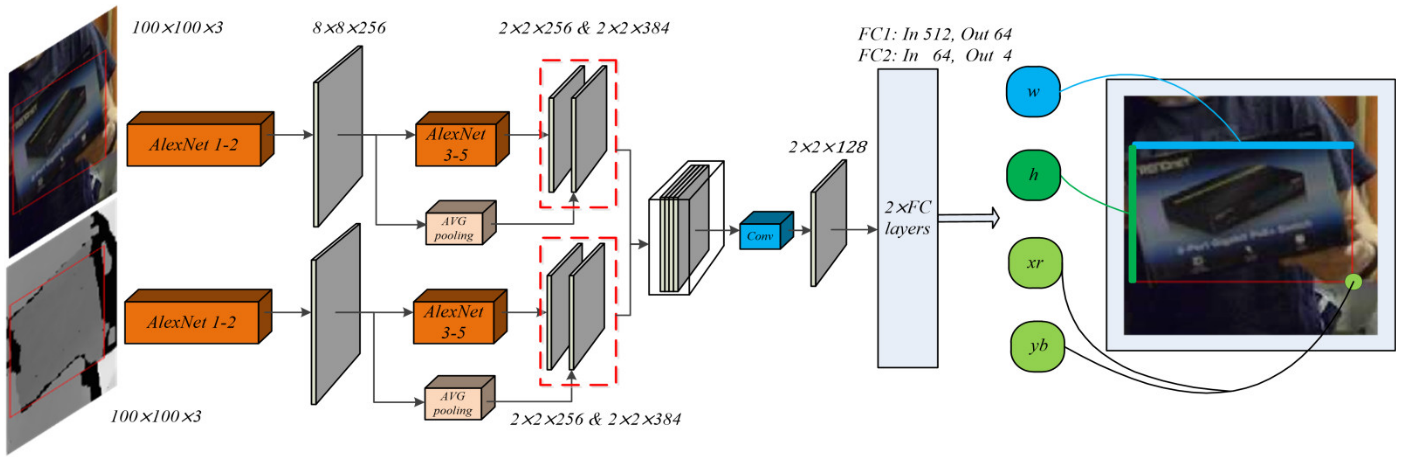
TSDM: Tracking by SiamRPN++ with a Depth-refiner and a Mask-generator



Mask-generator : put color mask to background beyond depth range.



TSDM: Tracking by SiamRPN++ with a Depth-refiner and a Mask-generator



Depth-refiner: an information fusion network

TSDM: Tracking by SiamRPN++ with a Depth-refiner and a Mask-generator

TSDM: Tracking by SiamRPN++ with a Depth-refiner and a Mask-generator

Pengyao Zhao, Quanli Liu*, Wei Wang and Qiang Guo
Key Laboratory of Intelligent Control and Optimization for Industrial Equipment of Ministry of Education
Dalian University of Technology
Dalian, China
liuql@dlut.edu.cn

Abstract*—In a generic object tracking, depth (D) information provides informative cues for foreground-background separation and target bounding box regression. However, so far, few trackers have used depth information to play the important role aforementioned due to the lack of a suitable model. In this paper, a RGB-D tracker named TSDM is proposed, which is composed of a Mask-generator (M-g), SiamRPN++ and a Depth-refiner (D-r). The M-g generates the background masks, and updates them as the target 3D position changes. The D-r optimizes the target bounding box estimated by SiamRPN++, based on the spatial depth distribution difference between the target and the surrounding background. Extensive evaluation on the Princeton Tracking Benchmark and the Visual Object Tracking challenge shows that our tracker outperforms the state-of-the-art by a large margin while achieving 23 FPS. In addition, a light-weight variant can run at 31 FPS and thus it is practical for real world applications. Code and models of TSDM are available at *https://github.com/lql-team/TSDM*.

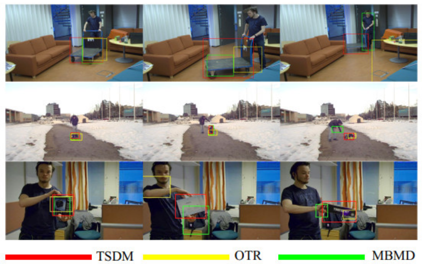
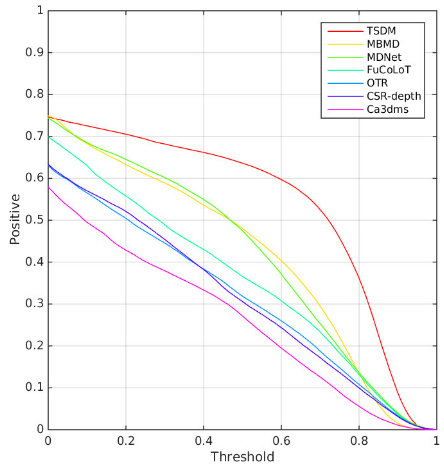


Fig. 1. Comparison examples of TSDM with two state-of-the-art trackers. The tracking targets from top to bottom are a cart, a dog and a box respectively. OTR [12], a RGB-D tracker, based on correlation filters ranked best in



Mehtod	Average IOU overlap											
	Overall	*Human*	*Animal*	*Rigid*	*Large*	*Small*	*Slow*	*Fast*	*Occ.*	*No-Occ.*	*Passive*	*Active*
TSDM	**0.792**	0.71(6)	**0.85(1)**	**0.86(1)**	0.77(2)	**0.81(1)**	**0.87(1)**	**0.76(1)**	0.69(5)	**0.94(1)**	0.84(3)	**0.78(1)**
OTR [12]	0.769	0.77(2)	0.68(3)	0.81(3)	0.76(4)	0.77(2)	0.81(2)	0.75(2)	0.71(2)	0.85(4)	**0.85(1)**	0.74(2)
ECO-TA [18]	0.754	0.77(3)	0.65(5)	0.80(4)	0.77(3)	0.74(4)	0.79(5)	0.41(8)	0.68(6)	0.85(3)	0.84(2)	0.72(4)
3D-T [2]	0.750	**0.81(1)**	0.64(6)	0.73(8)	**0.80(1)**	0.71(7)	0.75(8)	0.75(3)	**0.73(1)**	0.78(6)	0.79(7)	0.74(3)
CSR-rgbd++[11]	0.740	0.77(4)	0.65(4)	0.76(7)	0.75(5)	0.73(5)	0.80(4)	0.72(4)	0.70(3)	0.79(5)	0.79(6)	0.72(5)
Ca3dms [16]	0.737	0.66(8)	0.74(2)	0.82(2)	0.73(6)	0.74(3)	0.80(3)	0.71(6)	0.63(8)	0.88(2)	0.83(4)	0.70(6)
DM-DCF [10]	0.726	0.76(5)	0.58(8)	0.77(5)	0.72(7)	0.73(6)	0.75(7)	0.72(5)	0.69(4)	0.78(8)	0.83(5)	0.69(7)
DS-KCF [6]	0.693	0.67(7)	0.61(7)	0.76(6)	0.69(8)	0.70(8)	0.75(6)	0.67(7)	0.63(7)	0.78(7)	0.79(8)	0.66(8)