ML - Visual Object Tracking

Report

François Soulier
SCIA - 2024

Contents

1 Final algorithm

1.1 Descriptions

1.1.1 Main functions

As reminder, what we have at this step is a tracking pipeline using the Hungarian algorithm combined with Kalman filters.

For the final step of the tracking algorithm, a we use a pretrained ResNet50 to compute embedding of tracks and detections. These give us interesting visual information about the areas pointed by the bounding boxes. We then compute the cosine similarities of each detection with each track, resulting in a cosine similarity matrix. Finally, we create a combined similarity matrix, which a weighted average of the previously computed similarity matrix and the new one. We then use this matrix for the rest of the tracking pipeline, which remains the same.

1.1.2 Improvements

In order to work with more accurate detections, we can first use a YOLOv5, pretrained on the COCO dataset. Indeed, before executing the entire pipeline, we can generate new bounding boxes on the entire video. We then store these new bounding boxes in text files, with the CSV format (such as in *det.txt*). We then feed the pipeline with this new data. We can use the previous pipeline with the Hungarian algorithm using Kalman filters. The version with ResNet could not be troubleshooted (this is explained below).

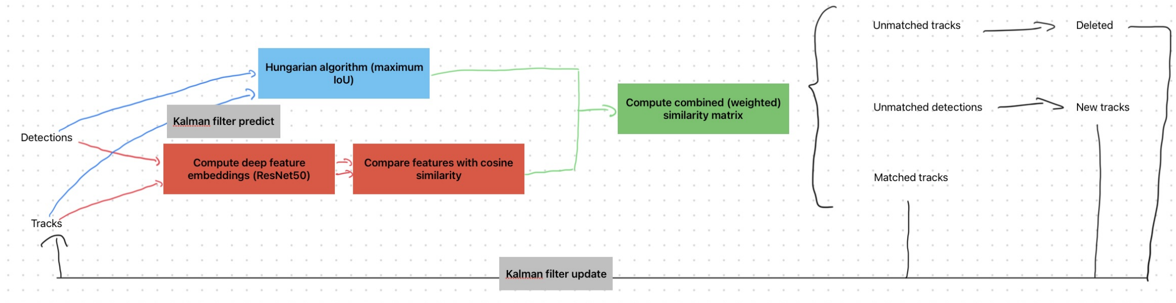
1.2 Difficulties

I have not been able make the pipeline fully work on the improved version (with YOLO and ResNet embeddings). The idea is that there are empty detection patches at some point in the video, but I have not had enough time to investigate the exact reason, and source of this problem. Overall, I have mainly encountered difficulty understanding the first version of the entire pipeline (Practical 2). I first did not understand the difference between detections and tracks, but once I understood it, I was able to implement the different pipelines smoothly.

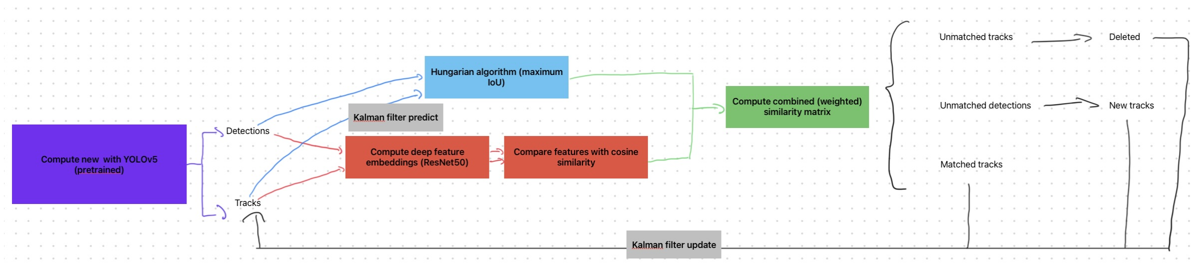
I have also ran out of time of time for metrics comparison. However, when exporting videos, we can observe (on the video images) that the detections with YOLO are much better. We can thus imagine that is improves the tracking algorithm. We can have a small insight to that regards when observing the tracking done with bounding boxes generated by YOLO.

1.3 Schemes

1.3.1 Main functions



1.3.2 Improvements



1.4 Metrics (Without improvements)

1.4.1 HOTA

HOTA: MyTracker-pedestrian	HOTA	DetA	AssA	DetRe	DetPr	AssRe	AssPr	LocA	OWTA	HOTA(0)	LocA(0)	HOTALocA(0)
ADL-Rundle-6	19.283	31.178	12.24	43.685	44.638	13.49	47.287	74.342	22.96	29.66	60.965	18.082
ADL-Rundle-8	50.136	42.473	59.466	64.39	51.185	64.879	78.407	82.262	61.849	64.3	76.786	49.373
ETH-Bahnhof	49.332	49.76	49.085	69.086	58.09	76.847	54.553	82.653	58.215	62.032	78.891	48.938
ETH-Pedcross2	43.128	34.845	53.425	36.805	79.679	59.956	77.316	85.416	44.342	52.277	81.284	42.492
ETH-Sunnyday	69.326	65.101	73.889	71.019	79.585	81.221	79.611	84.895	72.428	84.493	82.029	69.309
KITTI-13	52.596	45.486	61.025	54.807	65.052	66.841	77.838	82.404	57.796	66.794	78.225	52.249
KITTI-17	65.819	67.111	64.564	70.147	82.462	69.784	78.568	83.564	67.248	83.029	80.806	67.092
PETS09-S2L1	64.006	66.56	61.591	74.467	72.178	68.321	73.354	79.493	67.668	88.448	74.897	66.245
TUD-Campus	65.434	62.782	68.271	67.717	75.032	74.304	73.835	79.648	67.956	90.212	74.449	67.162
TUD-Stadtmitte	63.243	65.932	61.173	70.688	71.368	70.999	66.49	76.569	65.693	89.398	71.442	63.867
Venice-2	48.961	45.249	53.144	60.948	56.428	58.579	75.212	80.563	56.87	64.064	76.123	48.768
COMBINED	49.383	45.717	53.697	59.05	59.732	63.084	70.861	81.084	56.263	64.205	75.439	48.436

1.4.2 CLEAR

CLEAR: MyTracker-pedestrian	MOTA	MOTP	MODA	CLR_Re	CLR_Pr	MTR	PTR	MLR	sMOTA	CLR_TP	CLR_FN	CLR_FP	IDSW	MT	PT	ML	Frag
ADL-Rundle-6	5.0908	71.381	11.978	54.921	56.12	0	91.667	8.3333	-10.627	2751	2258	2151	345	0	22	2	297
ADL-Rundle-8	28.704	79.577	28.911	77.355	61.491	64.286	21.429	14.286	12.906	5247	1536	3286	14	18	6	4	57
ETH-Bahnhof	49.307	80.01	49.529	84.229	70.823	59.649	18.713	21.637	32.47	4561	854	1879	12	102	32	37	105
ETH-Pedcross2	39.422	83.293	39.646	42.919	92.914	14.286	28.571	57.143	32.251	2688	3575	205	14	19	38	76	13
ETH-Sunnyday	77.18	83.088	77.287	83.262	93.305	60	16.667	23.333	63.099	1547	311	111	2	18	5	7	10
KITTI-13	47.113	80.241	47.507	65.879	78.193	33.333	40.476	26.19	34.096	502	260	140	3	14	17	11	5
KITTI-17	83.309	81.133	83.602	84.334	99.139	55.556	44.444	0	67.397	576	107	5	2	5	4	0	4
PETS09-S2L1	87.265	75.725	87.489	95.331	92.399	94.737	5.2632	0	64.124	4267	209	351	10	18	1	0	62
TUD-Campus	84.68	76.752	84.68	87.465	96.914	75	25	0	64.346	314	45	10	0	6	2	0	4
TUD-Stadtmitte	91.436	71.845	91.609	95.329	96.245	100	0	0	64.596	1102	54	43	2	10	0	0	2
Venice-2	45.232	77.477	45.386	76.698	71.01	57.692	38.462	3.8462	27.957	5477	1664	2236	11	15	10	1	15
COMBINED	45.608	78.155	46.648	72.753	73.594	45	27.4	27.6	29.716	29032	10873	10417	415	225	137	138	574