

A workflow for pre-processing and semantically enriching movement data

Urbanism On Track '12: Data Challenge

Christian Gschwend and Patrick Laube, University of Zurich

This paper presents a general workflow allowing for a semi-automatic pre-processing (projecting, filtering, segmenting) as well as semantic enrichment (through data integration) of UOT movement data (Rotterdam Highrise dataset [HR11]).

GPS signals are superimposed by noise, whereas the uncertainty in the localization depends particularly on the satellite availability and the settings on the earth surface, e.g. buildings encasing a path leading to bigger positioning errors due to shadowing effects. Such urban accuracy issues may result in tracks where measurement errors exceed the actual movement signal. Additionally, positional inaccuracy is more severe with low speeds expressed by pedestrians. Hence, pre-processing is crucial in order to enable accurate movement analysis.

GPS fixes are pre-filtered by HDOP and number of available satellites (Figure 1). Aiming at a separation of true movement and inaccuracy-caused “pseudo-motion”, pre-filtered raw GPS data are segmented into stops and moves. The basic idea of a simple algorithmic approach introduced by Laube and Purves (2011) is applied using mean speed and turning angle instead of mean distance only. Here we assign indices smaller or equal to 0 to different kinds of stops (0: stops, -1: moves too short to be relevant, and -2: stops too short to split moves) and segmented moves are numbered with incrementing integers as IDs.

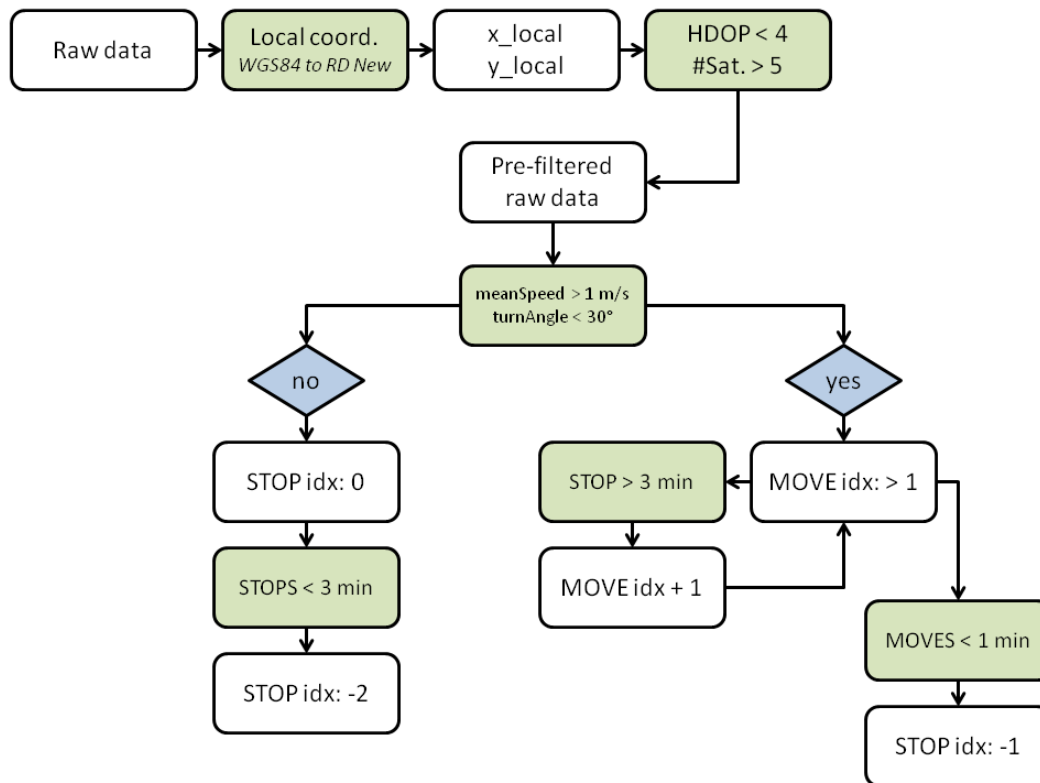


Figure 1: Process chain for detecting stops and moves.

This method is supposed to detect stops and moves, where longer stops most often correspond to places as home or work (Figure 2a). An evaluation of our segmentation method is not realized yet, but is planned, e.g. calculating density maps of stops and trying to assign these stops automatically to the underlying context (e.g. home, work, shopping, etc.).

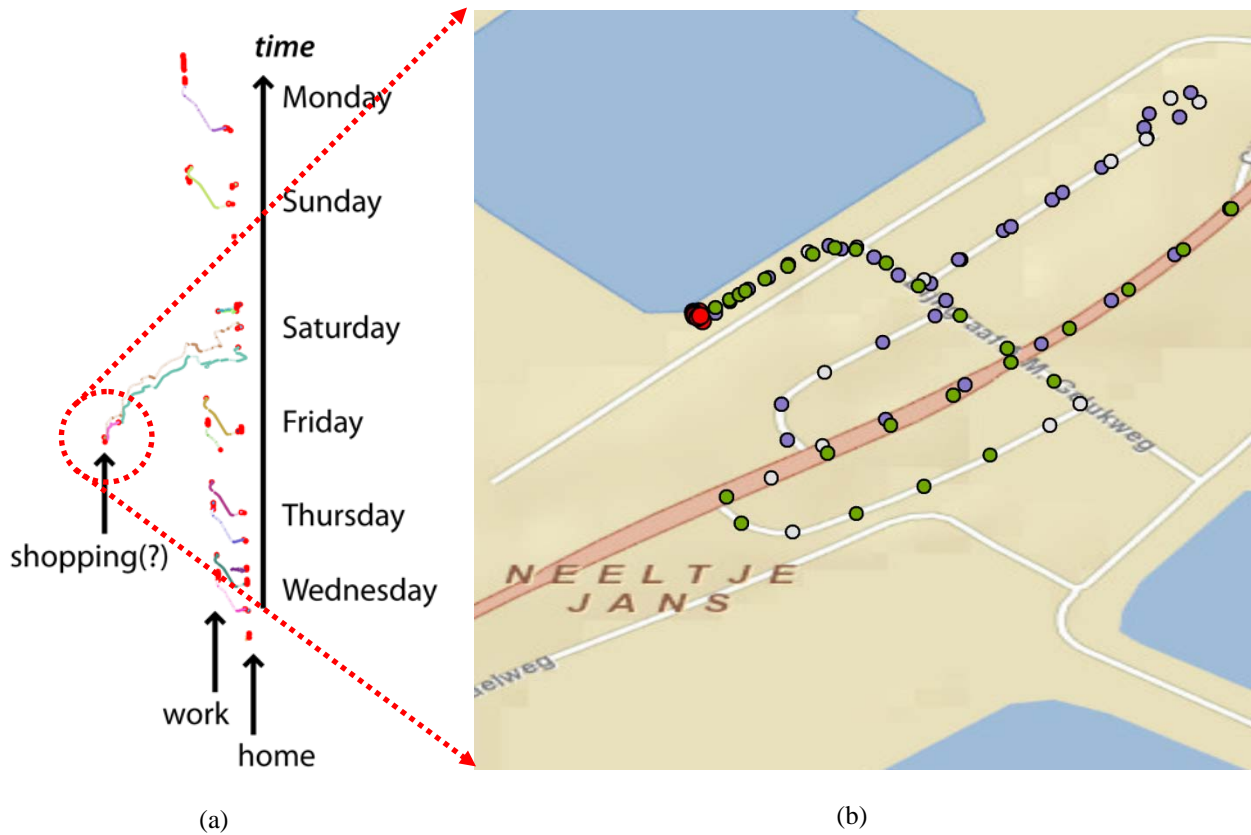


Figure 2: (a) Space-time cube for user 003_239.
 (b) Example of moves interrupted by a stop for user 003_239.
 Red: stops(0), grey: filtered fixes(-1,-2), other colors: moves(>0)

In future research, the goal of this work is to relate characteristics of movement automatically to the underlying context, in order to better understand movement in terms of environmental factors. Movement data and the calculated derivatives were joined to the semantic knowledge of the trip data details. First, the trip data details including semantic annotations are joined to the trip data by the unique *tripid*. Then, these trip semantics are linked to the raw data containing the calculated stop/move index by a unique attribute composed of *datetime* and the *user/GPS identifier* (e.g. 003_239) (Figures 3 and 4). These joins lead to semantically enriched movement data (Figure 5) and is a first step towards automatically linking movement and context information in general.

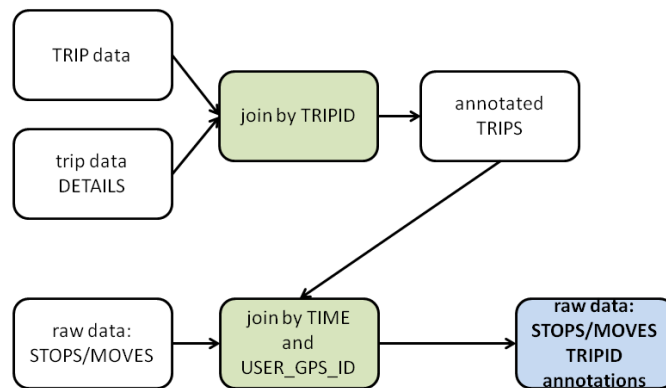


Figure 3: Attribute join from different data sources.

ID	LAT	LON	TRIPID	DATETIME
1424	51.914207	4.485859	004_228_0428_1	28-04-11 17:41:12
1425	51.914190	4.485909	004_228_0428_1	28-04-11 17:41:17
1426	51.914179	4.485994	004_228_0428_1	28-04-11 17:41:22
1427	51.913762	4.486037	004_228_0428_2	28-04-11 17:57:23
1428	51.913553	4.48557	004_228_0428_2	28-04-11 17:57:45

TRIPID	WORK	DAILY	SOCIAL
004_228_0428_1	1	0	0
004_228_0428_1	0	0	0

ID	DATETIME	LAT	LON	HDOP	NSAT	X	Y	IDX
304	28-04-11 17:41:12	51.914207	4.485859	1.02	9	92984.9	436566.3	-2
305	28-04-11 17:41:17	51.914190	4.485909	0.97	10	92984.6	436569.4	18
306	28-04-11 17:41:22	51.914179	4.485994	1.43	7	92950.1	436599.1	18
307	28-04-11 17:57:23	51.913762	4.486037	2.35	6	92938.2	436583.6	0
308	28-04-11 17:57:45	51.913553	4.48557	1.76	6	92948.8	436581.5	19

Figure 4: Schematic tables before joining.

ID	UTC_DATE	UTC_TIME	LOCAL_X	LOCAL_Y	idxStopsMoves	TRIPID	WORK	DAILY	SOCIAL
2961	4/29/2011	8:04:15 AM	93010.161046	436596.628348	18	004_228_0429_1	0	0	0
2962	4/29/2011	8:04:20 AM	92992.536201	436569.392731	18	004_228_0429_1	0	0	0
2963	4/29/2011	8:04:25 AM	92987.711192	436564.39657	-2	004_228_0429_1	0	0	0
2964	4/29/2011	8:04:30 AM	92986.565488	436565.450015	-2	004_228_0429_2	1	0	0
2965	4/29/2011	8:04:35 AM	92984.992818	436566.253953	-2	004_228_0429_2	1	0	0
2966	4/29/2011	8:04:40 AM	92984.680432	436569.417822	-2	004_228_0429_2	1	0	0
2978	4/29/2011	8:06:05 AM	92950.046707	436603.183172	-2	-9999	0	0	0
2979	4/29/2011	8:06:10 AM	92938.175278	436603.765546	-2	-9999	0	0	0
2980	4/29/2011	8:06:15 AM	92948.783347	436599.134248	-2	-9999	0	0	0
2981	4/29/2011	8:06:20 AM	92951.219523	436599.879554	-2	-9999	0	0	0
2982	4/29/2011	8:06:25 AM	92950.710193	436600.034972	-2	-9999	0	0	0
2983	4/29/2011	8:06:30 AM	92964.534111	436597.039495	-2	-9999	0	0	0
2985	4/29/2011	8:06:40 AM	92984.095177	436593.58013	18	-9999	0	0	0
2986	4/29/2011	8:06:45 AM	92988.909012	436591.129296	0	004_228_0429_2	1	0	0
3000	4/29/2011	8:07:55 AM	92989.778655	436583.4622	0	004_228_0429_2	1	0	0
3001	4/29/2011	8:08:00 AM	92992.317884	436580.809253	0	004_228_0429_2	1	0	0
3002	4/29/2011	8:08:05 AM	92994.132684	436581.774805	0	004_228_0429_2	1	0	0
3005	4/29/2011	8:08:33 AM	93024.129798	436522.976425	0	-9999	0	0	0
3006	4/29/2011	8:08:38 AM	93005.907615	436531.27474	0	-9999	0	0	0
3007	4/29/2011	8:08:43 AM	93000.322553	436529.357857	0	-9999	0	0	0
3008	4/29/2011	8:08:48 AM	92998.348615	436526.547236	0	-9999	0	0	0
3009	4/29/2011	8:08:53 AM	92999.551431	436524.882227	0	-9999	0	0	0
3018	4/29/2011	8:25:41 AM	93018.745481	436520.545203	0	-9999	0	0	0
3019	4/29/2011	8:25:47 AM	92994.108503	436514.894474	0	004_228_0429_2	1	0	0
3020	4/29/2011	8:25:52 AM	92997.769018	436509.658462	19	004_228_0429_2	1	0	0

Figure 5: Part of resulting joined table