PALMS Validation Dataset
Aug 2014

Introduction

PALMS, the Personal Activity Location Measurement System, was developed in 2010 by the Center for Wireless and Population Systems (CWPHS) at University of California, San Diego. PALMS is currently used by over 150 researchers at 90 organizations in 17 countries to discover the personal activity patterns of individual participants in free-living research studies. By outfitting participants with a GPS datalogger and a physical activity monitor, researchers can construct a detailed picture of a participant's day: travel patterns, locations, time sequences. Additional information about PALMS can be found on the PALMS wiki at http://ucsd-palms-project.wikispaces.com.

In order to validate the PALMS system, CWPHS researchers produced an annotated dataset of GPS and accelerometer data collected by research assistants over a two month period. The assistants followed prescribed transit routes that include multiple modes of transportation (walking, automotive, bus) in multiple urban settings in San Diego. The assistants kept detailed logs noting the time and starting / ending locations of each trip, stationary periods and mode of transportation. These logs were used to produce second-by-second "truth files" which were compared against the output of the PALMS algorithms. Details of the validation study can be found in the document entitled "PALMS Validation Study Design".

As part of MD2K, the PALMS GPS processing algorithms have been extracted from PALMS and made available as web services with a JSON interface. Included is a subset of the PALMS validation dataset which can be used to validate the MD2K PALMS web services. This dataset consists of the original GPX files collected by a Qstarz BT-1000 data logger, JSON files produced from the GPX files, the truth files, and the PALMS processed data (results of the PALMS calculations).

This document details the format of these files.

GPX Files

The filename reflects the date of data collection and the datalogger used (units 47 and 106). The GPX files were created by the Qstarz program (QTravel) used to download data from the devices. Details about the GPX format can be found at http://www.topografix.com/gpx.asp

GPX JSON Files

The GPX files were converted into JSON files for use by the PALMS DPU web services. The JSON format is documented on GitHub under MD2Korg / md2k-PALMS / PALMSDPU docs / GPStrack. The source code used to read the GPX file and write the JSON can be found under MD2Korg / md2k-PALMS / PALMSDPU source code / … / gps / GPSTrack.java

Truth Files

These files are supplied as both Excel spreadsheets (.XLSX) and CSV files. The filename reflects the date of data collection and corresponds to the GPX and JSON files.

Data (GPS tracks) was collected as a series of trips of various lengths and modes of transportation. Data was also recorded between trips. These are coded as "Transitions". In several cases, the GPS was turned off before the trip to force a loss of signal. Then the GPS was turned on immediately before the trip began. These are trips are coded as "cold trips". Often in cold trips, the true starting location is lost since the GPS is acquiring a position fix while the trip is in progress.

The format of the truth file is as follows:

Column Name	**Format**	**Description**
Date	MM/dd/yyyy	Date of data collection
Time	HH:mm:SS	Time (second-by-second)
TourID	Numeric	Unique identifier of Tour (trip)
ColdTourID	Number	Unique identifier when a "Cold Start"
Event	String	Event Type • Trip – person is actively traveling on a prescribed route • Pause – person is paused between trips • Transition – person in transitioning between prescribed trips – typically moving from the end of one trip to the start of another.
Mode	String	Mode of transportation: • Stand (standing still) • Sit (sitting down while stationary) • Walk • Car • Bus • Bicycle • Some Movement
InBldg	boolean	1 if inside a building
InGrge	Boolean	1 if inside a garage
InMall	boolean	1 if inside a shopping mall
ColdDevStart	boolean	1 if GPS was turned on immediately at the start of the trip

The following columns are incomplete and can be ignored:

StartRoute	String	Name of route (trip's) start
EndRoute	String	Name of route's end
LocName_S	String	Name of location at start of trip
LocAdd_S	String	Address of location at start of trip
LocName_E	String	Name of location at end of trip
LocAdd_E	String	Address of location at end of trip

PALMS Processed Data

The GPX files were processed by PALMS using the default parameters of Release 4 of the PALMS algorithms. (These are the algorithms implemented in the MD2K release.) The results are contained in a CSV and KML file. Format of the CSV can be found in the document entitled "PALMS GPS Output Variables R4".

The KML file can be viewed in Google Earth. Upon opening, the tracks of all trips are displayed. Since no accelerometer or heart rate data was included in the processing, only the *Locations*, *by Trips* and *by Stationary* folders contain information. The *Locations* folder contains all of the locations detected by PALMS. The *by Trips* folder contains the trips. Trip starting points are colored green, ending points are colored red, pause points are colored orange and movement points are colored yellow. The *by Stationary* folder contains those points were the participant was stationary (not in a trip).